How do childhood ADHD and stress relate to adult wellbeing and educational attainment?
A data science investigation using the 1970 British Cohort Study

Joanne Marie Cotton
Hughes Hall
January 2020

This dissertation is submitted to the University of Cambridge for the degree of Doctor of Philosophy

Preface

Declaration

This dissertation 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 word limit approved by the relevant Degree Committee.

Part of the research reported in chapter 4 was published jointly with my supervisor in the following article:

> *Cotton, J., & Baker, S. T. (2018). A data mining and item response mixture modeling method to retrospectively measure Diagnostic and Statistical Manual of Mental Disorders-5 attention deficit hyperactivity disorder in the 1970 British Cohort Study. International Journal of Methods in Psychiatric Research. https://doi.org/DOI: 10.1002/mpr.1753*

Acknowledgements

None of this journey would have been possible without the steadfast support and guidance of my brilliant supervisor, Dr Sara Baker. Sara encouraged and guided the development of my ideas and enthusiasm and abilities, and for that I am so grateful.

I have been incredibly fortunate to have Prof Anna Vignoles as both an advisor and supervisor (for the final stretch, whilst Sara was on leave). She is a truly gifted teacher, and her vast knowledge of methods was instrumental in raising the standard of my quantitative analysis and bringing the whole unwieldy thing to completion.

This undertaking was made feasible and bearable only by the patience, sacrifice and support of my loving husband Bob... and the companionship of my gorgeous pets: Toby, Sasha, and Kiska.

I am thankful for the funding I received to support my studies from the ESRC (award number 1652810), and from Hughes Hall. I'm also grateful to the Centre for Longitudinal Studies and the UK Data Service for providing high-quality data, and the invaluable contribution from the participants of the 1970 British Cohort Study.

Finally, I would like to thank Katie Graham, John Gill, John Cotton, Dr Dagmara Dimitriou, Dr Dervila Glynn, Prof Joe Herbert, Prof Peter Jones, Dr Golam Khandaker, Prof Edmund Sonuga-Barke, Michelle Intrepidi, Dr Helen Taylor, Prof Liz Pellicano, Dr Gabriela Roman, Dr Ricardo Sabates, Dr Duncan Astle, Dr Joe Bathelt, Dr Joni Holmes, Deusa Abreu, Catherine Bevilacqua, Dr Gill Francis, my sisters, and my parents for their inspiration and kind support either at crucial points in my PhD journey or all along the way.

Dedication

This work is dedicated to confident, happy, engaged ADHD children who develop into anxious and withdrawn adults because of their stressful experiences at school.

I hope things are better for the next generation.

Table of contents

Index of Tables

Index of Figures

Abbreviations

ADHD – Attention Deficit Hyperactivity Disorder
ALSPAC – Avon Longitudinal Study of Parents and Children
APA – American Psychiatric Association
BCS70 – 1970 British Cohort Study
CEM – Coarsened Exact Matching
CLS – Centre for Longitudinal Studies
DSM – Diagnostic and Statistical Manual
EDL – Educational attainment Level
EF – Executive Function
ESRC – Economic and Social Research Council
FIML – Full Information Maximum Likelihood
ICD – International Classification of Diseases
IRT – Item Response Theory
MAR, MCAR, MNAR – Missing at Random, Missing Completely at Random, Missing Not at Random
MCS – Millennium Cohort Study (2001)
MLR – Maximum Likelihood Robust
MTA – Multimodal Treatment study of ADHD
NCDS – National Child Development Study (1958)
OLS – Ordinary Least Squares (regression)
ONS – Office of National Statistics
RCT – Randomised Controlled Trial
RQ – Research Question
SES – Socioeconomic Status
SWB – Subjective Wellbeing
ZIMM – Zero-Inflated Mixture Model

Abstract

How do childhood ADHD and stress relate to adult wellbeing and educational attainment?
A data science investigation using the 1970 British Cohort Study
Joanne Marie Cotton

Background: Attention Deficit Hyperactivity Disorder (ADHD) is a childhood and adult disorder characterised by nonnormative inattentive, impulsive, and hyperactive behaviour. Over time the condition has become increasingly medicalised, and whilst it is estimated to affect 5-7% of schoolchildren internationally (Sayal et al., 2018), only 1.6% are diagnosed with ADHD in the UK (NHS Digital, 2018). Reviews report that childhood ADHD leads to poor adult outcomes in all areas of life (e.g. Costello & Maughan, 2015; Erskine et al., 2016). Although about 50% of ADHD children function well as adults, knowledge is limited about psychosocial factors in outcomes, (Costello & Maughan, 2015) such as those related to stress.

State regulation theory, (Sanders, 1983; Sergeant, 2000) was the basis for an investigation using data from the age 0, 5, 10, 34, and 42 sweeps of the 1970 British Cohort Study (BCS70; Centre for Longitudinal Studies: UCL/IoE, 2019). Stress and protective factors were operationalised as stressful life events, chronic stressors, self-esteem, and locus of control. The following questions were examined[1]:

1) What robust measures of DSM-5 ADHD can be retrospectively measured and validated?
2) What is the relationship between childhood ADHD and stress?
3) What is the effect of childhood ADHD on adult a) subjective wellbeing, and b) educational attainment, the latter as a proxy for SES[2] and objective wellbeing?

Method: Innovative data science methods were applied, including:

1) A data mining framework (Kurgan & Musilek, 2006) to derive new constructs in old data;
2) Robust linear and logistic regression models (e.g. MLR, FIML; Muthen & Muthen, 2017);
3) Zero-inflated mixture modelling (Wall et al., 2015) to estimate an ADHD severity score;
4) Machine learning (vselect; Lindsey & Sheather, 2010) to aid selection of an optimal set of covariates for quasi-experimental matching; and
5) Coarsened Exact Matching (CEM; Iacus et al., 2014) to derive a weighted matched sample of ADHD children and similar controls.

Key findings: A DSM-5 ADHD subgroup and subtypes were retrospectively derived and validated using age 10 BCS70 data (N=11,426; n_{ADHD}=594, 5.2% prevalence, 30% girls, 46% inattentive subtype). Overall prevalence aligned with epidemiology estimates, but the

---

[1] Research questions are paraphrased, and question 3 here corresponds to research questions 3 and 4 in the thesis text; i.e. there were separate questions for each outcome.
[2] SES: socioeconomic status

relatively high percentages of ADHD girls and inattentive cases enabled rare new insights for these groups. The distribution of the ADHD severity score (*N*=11,426, *M*=0.06, *SD*=0.91) supported dimensionality of the construct.

Stressful life events, chronic stressors, self-esteem and locus of control significantly predicted DSM-5 ADHD symptomatology and explained 19.5% of the ADHD severity score at age 10 (*N*=11,426), supporting State Regulation Theory at the psychosocial construct level.

Quasi-experimental methods were employed to create a pruned longitudinal sample of ADHD and control cohort members matched on evidence-based confounds (*N*=6,207). Regression models on this sample did not support a significant effect of childhood ADHD on adult outcomes, contrary to prevailing evidence from mostly clinical samples matched on fewer confounds. Matching confounds used were sex, father's education, depressed mother, mother smoked during pregnancy, childhood wheezing, and low standard home. Replication and refinement are needed, but the finding suggests future experimental studies should consider stratifying samples on these factors, and that ADHD per se may not drive poor outcomes.

In the matched sample (*N*=6,207), age 10 maths scores (boys and girls), externalising problems, and engagement in leisure activity (girls only), were significant factors predicting a continuous composite measure of adult subjective wellbeing. Parent education, age 10 maths, reading (boys and girls), locus of control, and authoritarian child-rearing views (girls only), were significant childhood factors predicting a dichotomous academic qualification measure of adult educational attainment, as a proxy for SES/objective wellbeing. All effect sizes were small[3].

In a longitudinal ADHD subsample (*n*=369), age 10 chronic stressors, externalising problems, and reading significantly predicted adult subjective wellbeing, explaining 7.1% of variance (boys and girls). Father's education and age 10 reading significantly predicted adult educational attainment. The effects of chronic stressors and reading, and the higher proportion of girls and inattentive ADHD cases in the sample provide novel insights which should be translatable into teacher training and practice.

Findings are applicable internationally, subject to demographic generalisability parameters.

---

[3] Per guidelines in Ferguson, (2009) and Olivier et al., (2017).

Chapter 1 Introduction

1 Terminology in the title

1.1 ADHD

Attention Deficit Hyperactivity Disorder (ADHD) is a childhood and adult disorder of inattention, impulsivity, and hyperactivity that interferes with functioning in multiple settings. It has three presentations: primarily inattentive, primarily hyperactive/impulsive, and combined (American Psychiatric Association, 2013), and affects approximately 5-7% of children worldwide (Polanczyk et al., 2007; Sayal et al., 2018; Willcutt, 2012).

ADHD is classified as a medical disorder, and the most widely researched version of that disorder is defined in the Diagnostic and Statistical Manual of Mental Disorders, currently on version 5 (DSM-5; American Psychiatric Association, 2013). Based on the 2019 school census, there are ~10.3 million children in school in the UK (Department for Education, 2019), and 5-7% of them amounts to 515,000 – 721,000 children. However, many of the children who meet the DSM-5 criteria for ADHD in the UK are not diagnosed or treated. There is a tendency in the UK to use the more stringent ICD-10 diagnostic criteria for Hyperkinetic Disorder, which only identifies the most severe cases of DSM-5 ADHD (NHS Digital, 2018). Thus, the rate of ADHD diagnosis and treatment in the UK is about 1.6% of school-aged children (NHS Digital, 2018). Whether or not ADHD should be more or less medically diagnosed and treated is not my interest here. I am interested in childhood factors, particularly for children with ADHD, that may be open to influence through education, and may lead to positive adult outcomes.

We don't often talk to children with ADHD (Brady, 2014), but when we do we find their subjective experience of school is traumatic. They report struggling to get to school every morning, frustration about forgetting things they need, getting thrown out of class on a regular basis, feeling angry and 'fed-up' about getting in trouble with teachers, not being able to concentrate or finish their work, getting picked on every day, not having any friends, feeling ashamed that they are different and/or need medication, and avoiding school completely, sometimes once a week or more (ADDISS, 2006). They also report feeling worried and unhappy about not being able to perform or behave in the ways expected of them (Singh, 2012), being incompetent, and not getting along with teachers (Rogers & Tannock, 2018). School is not intended to be a hostile place, yet for the 5-7% of children with ADHD symptomatology, this qualitative evidence indicates that it may be.

ADHD is a critical topic of study in education because it is the most prevalent difficulty for school-age children, and every teacher and educational professional who has worked with children has encountered the inattention, hyperactivity, and impulsivity symptoms (American Psychiatric Association, 2013) associated with ADHD. The interaction between an ADHD child and their education influences the course of their academic experience and ultimately their relationship with themselves and society. In spite of this importance, and volumes of published research over decades, it is still not clear how education systems can best support optimal outcomes for these children, in the longer term.

1.2 Stress

Stress is a central construct in this thesis (and accordingly in the title) because the theoretical basis for my research is State Regulation Theory, a cognitive-energetic model of ADHD relating stress and human performance (Sanders, 1983, 1998; Sergeant, 2000). The term 'stress' has a long history and can be broad and ambiguous. Two complementary definitions are provided for clarity about what is meant by the term when used in this thesis:

> *"Stress arises when an individual has a subjective feeling of being 'under threat'; they assess the situation, and based on the assessment feel they are unable to cope with the perceived threat. Consequently, performance on tasks with a high cognitive load is degraded." (Salas, Driskell, & Hughes, 1996, p. 11)*

> *"Stress will arise whenever the effort mechanism is either seriously overloaded over time or falls altogether short in accomplishing the necessary energetical adjustments." (Sanders, 1983, p. 79)*

State Regulation theory is often used as a basis for research in cognitive tasks measuring reaction time and accuracy (for a discussion, see Kuntsi & Klein, 2011), and more generally for linking biological data to behaviour. However, no biological data was available in the secondary data analysed, so here the theory was operationalised at the psychosocial level with measures of stressors as stimuli, i.e. events and ongoing situations that are widely thought to lead to stress, or as defined above, *a feeling of being unable to cope*, and *overloaded effort mechanism*. In the context of this thesis, a person's response to this feeling can include hyperactivity, inattention, excessive worry, low mood, withdrawal, or aggressive behaviour. These are defining characteristics of ADHD, anxiety, depression (internalising), and conduct (externalising) problems, which were evaluated in this thesis as constructs separate from stress.

My general hypothesis was that exposure to stressors should lead to more severe ADHD symptoms, which in turn should lead to greater impairment in functioning, and factors that reduce stress should have the opposite effect. This hypothesis aligns with recent trends in

ADHD research to study symptom management through stress-reducing activities like physical exercise and mindfulness meditation (Diamond & Lee, 2011; SAGE journals and APSARD, 2019). The process I used to select the theory, as well as a fuller description of the model and related hypothesis are described in chapter 3.

1.3 Adult outcomes of wellbeing and educational attainment

I chose to focus on how ADHD and stress relate to two long-term life outcomes. I selected wellbeing and educational attainment because they are not highly correlated with each other, but each is correlated with many other outcomes, thus they provide a broad, person-centred view. Wellbeing is a measure of how well life is going, and in the present thesis I used subjective measures based on self-assessments. Educational attainment is widely studied as an outcome in its own right, but is also evaluated here as a proxy measure of socioeconomic status and indicator of objective wellbeing (Office for National Statistics, 2019). The measure used was based on the highest level of academic qualifications achieved, at age 42 or 34.

Studies of ADHD outcomes often report on a long list of detailed 'event-based' outcomes, such as specific diagnosis of psychiatric disorders, criminal convictions, traffic accidents, and more (Cherkasova et al., 2013; Erskine et al., 2016). These are less ambiguous, particularly compared to wellbeing, but do not provide an overall view of how life is going; a person can experience negative events, yet overall still function relatively well. Evaluating only two outcomes also has the advantage of allowing more detailed analysis of each one, whilst minimising risk of false positive findings attributable to multiple comparisons.

1.4 Data science

The term 'data science' is used in the title to describe the methodological approach. Data science is a relatively recent term which has come about with the explosion of data that is now widely available and the technology that makes it accessible. It draws on multiple competencies to facilitate discovery of new knowledge and insights from large datasets (Schutt & O'Neil, 2019). Data science is used to describe analysis that combines maths/statistics, subject matter expertise, data mining, and data visualisation in creative new ways (Schutt & O'Neil, 2019).

There are academic and non-academic researchers: economists, epidemiologists, biostatisticians, and financial, marketing and political analysts, and more, who use statistics, data mining, and data visualisation in large datasets to study constructs about which they have specialist knowledge. This is not new; these analysts have been 'doing data science' for quite some time. Thus, I use the term here not to imply I am doing something new, but rather to use current terminology for a piece of work that combines knowledge from different domains in a way that has become more prevalent, and discussed as a field of study in its own right. Despite using these methods, it is important to note that my work is theory-driven and not based on data mining per se.

I endeavoured to explore and identify creative ways to answer questions about ADHD and test state regulation theory using available data from BCS70 with the empirical studies in this thesis. Data mining techniques were used to retrospectively derive measures of constructs that were not prospectively measured, including ADHD, stressful life events, chronic stressors, and leisure activity. Advanced statistics were used to build optimal models to address violation of parametric assumptions and complex missingness and create quasi-experimental conditions (i.e. balanced treatment and control groups) in observational data. Data visualisation was used to illuminate important relationships within the data, such as the zero-inflation of ADHD in a non-clinical sample, item characteristic curves for each of the ADHD symptoms, and the overlap of constructs (e.g. ADHD severity, wellbeing, and educational attainment) between subgroups defined as ADHD and non-ADHD. The combination of methods and specialist knowledge about ADHD applied in this work created novel insights into the relationships between ADHD, stress, wellbeing, and education.

1.5 The 1970 British Cohort Study

The 1970 British Cohort Study (BCS70) is the data source used for this thesis. It is an ongoing study of 17,198 children born in the UK in one week of April 1970. The study has captured thousands of data points on these individuals over the course of their lives, and provides a rich long-term view of biological, psychological, and social antecedents in childhood to life outcomes in adulthood. The process I used to select the BCS70 and a full description of the study is covered in chapter 2.

2 Developing a research objective

2.1 Prior to PhD

My PhD study follows on from MPhil research, which was summarized in a published conference paper: “An exploratory case study of three children with ADHD and social difficulties: Child and parent responses to an educational intervention designed to facilitate self-regulation and deep learning” (Cotton et al., 2015). I used qualitative thematic analysis of interviews and videos of the children and parents and found themes of ‘just want to play’, ‘don’t see problems’, and ‘stress and frustration’. These findings influenced the development of my research questions. At the start of reading for my PhD, I explored the idea of developing a targeted ADHD educational approach focused on social skills, building on my MPhil work and drawing on theory from play and stress research. However, after my initial literature review, I felt there were significant gaps in theoretical and evidential support, particularly with regard to longer-term outcomes, and I wanted to learn more before engaging with children and families. Ultimately, I decided that analysis of a large secondary longitudinal dataset would best support my interest in long-term outcomes. The methodology is also well suited to my previous experience working as an analyst in industry and government. I still intend to develop and test a targeted educational approach for children with ADHD and hope the design will be richer because of my findings and experience from analysing secondary data.

2.2 Contextual influences

This research was undertaken within the Faculty of Education at the University of Cambridge, primarily under the guidance of the psychology, education, and learning studies group (PELS). Throughout the course of my studies, I also participated in events and training, connected to other researchers, and gathered input and inspiration for my work by interacting with the Centre for Research on Play in Education, Development and Learning (PEDAL), the Research for Equitable Access and Learning (REAL) centre, and Cambridge Neuroscience. The wide array of subject areas (psychology, psychiatry, epidemiology, neuroscience, economics) covered by these groups or centres demonstrates how multi-disciplinary education research can be. The multi-disciplinary approach greatly enriched the development and implementation of my project.

2.3 Inspiration from the literature

I noted from early literature reviews that at the group level, ADHD has been associated with poor childhood academic and social functioning, and poor long-term life outcomes in health, relationships, educational and vocational attainment, accidents, and criminal activity (Costello & Maughan, 2015; Kuriyan et al., 2013; Loe & Feldman, 2007; Matza et al., 2005). However, in Costello & Maughan's (2015) review of long-term outcomes for childhood ADHD, they found that about half were functioning well in their 20s and 30s. Their definition of functioning well was "free from significant difficulties in the areas of work, health, relationships, and crime" (Costello & Maughan, 2015, p. 324). The authors noted that research in psychiatric disorders often focuses on the risk of negative outcomes, and they opted instead to explore paths to more positive, or optimal outcomes. They concluded that predictive factors for positive outcomes were "a mixture of personal characteristics and environmental supports" (Costello & Maughan, 2015, p. 324). They cited evidence to suggest that better outcomes were associated with less severe symptoms, higher IQ, socio-economic status (SES), educational attainment, and absence of comorbidity, especially conduct problems (Costello & Maughan, 2015). They noted the lack of evidence for the role of other psychosocial factors in long-term outcomes for ADHD, and encouraged the field to direct attention to this topic (Costello & Maughan, 2015).

Psychosocial is a shorthand term combining psychological and social. In the context of mental health, social factors of interest are typically constructs like stressful life events and chronic stressors, including loss, health problems, socio-economic disadvantage, and lack of social support (Brown, Harris, et al., 1973; Stansfeld & Rasul, 2006; Williamson et al., 1995). Individual psychological factors are thought to influence the way social factors are perceived and thus ultimately affect an individual. Examples of psychological factors are personality/temperament characteristics, and attitudes or beliefs, like self-esteem and locus of control (Stansfeld & Rasul, 2006). Attitudes and beliefs are especially interesting for the development of targeted educational approaches or interventions, because they may be more open to influence (Durlak et al., 2011), compared to perhaps less malleable characteristics like IQ, socio-economic status (SES), and temperament.

From neuroscience research we know that the brain continues to develop until about the age of 25 or later (Lebel et al., 2008; Sowell et al., 2003), and for ADHD that development may take longer (Shaw et al., 2013). Also in Western cultures there is a tendency for a protracted period of emerging adulthood, with settled adulthood (e.g. independence from parents, completion of education, stability of residence) not happening until around age 30 (Arnett, 2000). Thus, I chose to search for a large representative longitudinal secondary dataset that measured behaviour and psychosocial factors in childhood, and life outcomes after the age of 30.

Based on my previous research, cross-discipline engagement, and the inspiration of the Costello & Maughan review I developed a broad research objective, to:

> *...evaluate how psychosocial factors for those with ADHD in childhood relate to positively framed outcomes in settled adulthood.*

The use of a secondary dataset necessarily imposed limitations on the scope and nature of the research questions. Because of this, I chose to present the next two sections in an unconventional order, i.e. the selection and description of the dataset is first, followed by the literature review that builds up to the research questions, informed by the data source available.

3 Structure of the thesis

There are nine chapters in this thesis. Chapter 2 describes the process of selecting a secondary dataset, and attributes of BCS70. Chapter 3 reports on a review of literature on ADHD, theories, and long-term outcomes. Based on this a rationale was developed for selection of the outcome focus and translation of the research objective into research questions. Chapter 4 answers the first research question. It starts with a summary of a pilot study to retrospectively identify ADHD in BCS70, which is followed by a copy of a final published study. Chapter 5 reports a pilot study which tested the use of quasi-experimental methods to balance ADHD and control samples and isolate the unique effects of ADHD on adult outcomes. Chapter 6 is a description of improvements to measures and methods based on learning from chapter 5. Chapter 7 evaluates the relationship between childhood stress and ADHD, and between childhood ADHD, stress, and adult subjective wellbeing, i.e. research questions 2 and 3. Chapter 8 evaluates the relationship between childhood ADHD, stress, and adult educational attainment, i.e. research question 4. Finally, chapter 9 discusses the findings, compares them to existing literature, and evaluates strengths, limitations, and implications for future research.

Chapter 2 Design frame, dataset selection, and dataset description

1 Design frame

The design frame for this collection of studies is secondary analysis of longitudinal birth cohort data. Longitudinal data was selected because it allowed evaluation of early life ADHD symptoms and a rich set of psychosocial factors, as well as outcomes in settled adulthood. Secondary data was used because it is the only practical way to analyse longitudinal data (across childhood and settled adulthood) within the timeframe of a PhD. However, even if it were not the only practical option, secondary data available for re-use is generally collected and maintained by large organizations with extensive resources, which can offer advantages of high-quality design, sampling, follow-up, and data management (Cohen et al., 2011; Vartanian, 2011).

Use of secondary data in research requires a specific set of methods, some similar and some quite different to primary research. With secondary research, one of the first steps is selection of a dataset that can best meet the research objectives, taking into consideration practical factors like ease and cost of access to the data (Vignoles, 2014). Once a dataset is selected, the methods used and data collected by the original researchers are studied, and an iterative process begins, interpolating between the secondary researcher's desired method and research questions, and the practical limitations imposed by the original methods and data available (Hennekens & Buring, 1987; Vartanian, 2011; Vignoles, 2014). The following paragraphs describe the process used to select a dataset for the present study and a description of the selected dataset, including the original study's sampling, attrition, missing data, and ethics procedures.

2 Dataset selection

There are few longitudinal birth cohort datasets available which cover a timeframe (early childhood to settled adulthood), optimal for analysis of long-term ADHD outcomes. My preference was to use data from the United Kingdom, because a longer-term objective of my research is to inform UK education policy for children with ADHD. Thus, I wanted to use data that was as representative as possible of the UK population, to allow for national generalizability. I spent some time evaluating the Avon Longitudinal Study and Parents and Children, or ALSPAC, on children born in 1991 (University of Bristol, 2015), but decided against using it because I concluded the participants were still not old enough at age 25[4] to evaluate outcomes in settled adulthood. Also, the sample was drawn from the Avon region of the UK only, and suffered from substantial non-random attrition, in part due to the intensive nature of the assessments. Thus, there were only two available datasets that met all my criteria: the 1970 and 1958 British birth cohort studies. In the interest of using the most current data, the 1970 study (BCS70) was selected, which has extensive childhood behaviour data, and adult outcomes up to age 46[5].

Even though the BCS70 is the most recent of the two candidate data sets, many social, economic, and educational changes have come to pass since this cohort were children. In particular, the concept of ADHD in its current form did not exist then. Given that a measure of ADHD is central to the research, I conduced a pre-review of questionnaire data available in the BCS70 and concluded it should be feasible to identify cohort members with ADHD symptoms in childhood and use this sub-group to examine variation in long-term outcomes. An advantage of this approach is that the subgroup is in a sense more 'pure', because it not affected by biases that lead to misreporting of ADHD, which is more likely in more recent and clinical samples. I also found a previous working paper using BCS70 that identified ADHD-related symptoms at age 10 and reviewed a long list of outcomes at age 30 (Brassett-Grundy & Butler, 2004). I noted there were a number of ways I could build on their work using more recent techniques, measures and data, and this is discussed further in the literature review on long-term outcomes (chapter 3).

3 Description of the 1970 British Cohort Study (BCS70)

This section provides background information about the BCS70 study, including the purpose, sweep history, sampling, and missing data and attrition. It concludes with an ethics statement covering both the proposed PhD research, and the BCS70 itself.

BCS70 is an active longitudinal study of persons born in the UK during the week of 5-11 April 1970. The initial sample consisted of 17,198 children, and so far ten full research sweeps have been completed when the participants were aged 0, 5, 10, 16, 26, 30, 34, 38, 42, and 46. (Centre for Longitudinal Studies: UCL/IoE, 2019; Elliott & Shepherd, 2006). In sweeps up to and including age 16, immigrants identified with birth dates in the reference week were added to the study, increasing the total number of participants in any sweep to over 18,000 (Centre for

[4] At the time of study selection in 2016.
[5] The age 46 sweep data did not become available until the October 2019, which was too late for it to be included in my analyses. The most recent data included were from the age 42 sweep.

Longitudinal Studies: UCL/IoE, 2019; Elliott & Shepherd, 2006). Table 1 below summarizes participation over the course of the study, along with pertinent notes on each sweep.

Sweep (age)	Year	No. of participants[6]	Notes
0	1970	17,198	Home interviews conducted by midwife present at birth, augmented with admin medical data
5	1975	13,135	Four parts: home interview, maternal self-completion, skills test, developmental history
10	1980	14,875	Six parts: parent interview, parent self-completion questionnaire, medical exam, participant self-completion questionnaire, educational assessments, teacher/head-teacher self-completion questionnaire
16	1986	11,622	Eight parts: parent interview, parent self-completion questionnaire, medical exam, participant self-completion questionnaire, educational assessments, teacher/head-teacher self-completion questionnaire, diaries Participation from schools was much lower than expected due to a teachers' strike
26	1996	9,003	One part: 16-page postal self-completion questionnaire Participation was relatively low due to inadequate funding, postal survey, 10 years since last contact, and consent now required from cohort members rather than parents Study ownership moved to CLS/IoE
30	2000	11,261	Two parts: face-to-face interview and self-completion questionnaire
34	2004	9,665	Three parts: face-to-face interview, self-completion questionnaire, and basic skills assessment Plus, participants who were parents also completed an additional interview and questionnaire about the child(ren); the child(ren) were assessed for skills, and did a self-completion questionnaire, if appropriate
38	2008	8,874	Telephone survey
42	2012	9,841	Four parts: face-to-face interview, self-completion questionnaire, skills assessments, plus written consents to link NHS, DWP, HMRC data
46	2016	8,581	Five parts: face to face interview, self-completion questionnaire, cognitive assessments, basic health exam (by nurse), dietary diary. Included collection of biological data.

[6] Reports of the number of participants vary slightly across different sources; the CLS website data was used here.

Table 1. BCS70 participation numbers and data collection notes by sweep

There were also five sub-studies completed with smaller samples at ages 22 and 42 months, 7, 8-9 (twins only), and age 21.

(Centre for Longitudinal Studies: UCL/IoE, 2019)

BCS70 was originally designed to collect medical and social data on mothers and analyse them in relation to illness and mortality in infants, and also for comparison to the 1946 and 1958 British Cohort Studies (Centre for Longitudinal Studies: UCL/IoE, 2019; Elliott & Shepherd, 2006). The data collection remit expanded over time to include extensive psychological, educational, and economic data, and has been used worldwide by researchers across the social sciences. To date[7], 1,071 studies and working papers using BCS70 have been published and documented in the CLS online bibliography (Centre for Longitudinal Studies: UCL/IoE, 2019).

As described in Table 1 above, the data collection methods for BCS70 included face-to-face and telephone interviews, questionnaires, medical and physical examinations, psychological and educational assessments, and diaries (Coleman, 2015, p. 10). Some of the items were repeated (longitudinal) across multiple sweeps, but most items were chosen for relevance at the participants' life stage, so were cross-sectional.

Over time, BCS70 has had different names, managing organizations, and funding sources. It was originally called the 'British Births Survey', and directed by the National Birthday Trust and Royal College of Obstetricians and Gynaecologists (Coleman, 2015; Elliott & Shepherd, 2006). After several interim moves and varying sources and levels of funding, ownership settled in 1998 with the Centre for Longitudinal Studies (CLS) at the Institute of Education (IoE), with the Economic and Social Research Council (ESRC) as a key funder (Centre for Longitudinal Studies: UCL/IoE, 2019). BCS70 is unusual in its large sample size, longevity, and breadth of data coverage, and these characteristics make it ideal to support research on theories and hypotheses of predictors in childhood for long-term life outcomes in in the UK, such as the subject of this thesis.

3.1 Sampling

The sampling method for BCS70 of using all children born in a single week is not technically a sampling method, but rather a 'whole universe' approach. This same method was used in the 1946 and 1958 UK birth cohorts, and so gives an advantage of comparability between the three studies. The original sample in BCS achieved 96% coverage of all eligible births in Great

[7] 24 May 2020

Britain, and was at that time reported to have minimal bias in terms of representativeness of the UK population (Bynner & Joshi, 2007; Chamberlain et al., 2013). However, over time BCS70 suffered from significant non-response and missingness of individual data items, which created systematic biases in the subsequent sweeps. Given the importance and complexity of missingness in the BCS70, a general background on classifications of missingness is provided next, followed by the specific characteristics known about the missingness in the study up to the 2012 sweep.

3.2 Missing data and attrition

Missing data can be classified as Missing Completely at Random (MCAR), Missing not at Random (MNAR), or Missing at Random (MAR) (Baraldi & Enders, 2010; Rubin, 1976; Sainani, 2015). If data are MCAR, there is no relationship between the missing data and the characteristics of the participants, and the observations with missing data can be deleted whilst the assumption of a random sample is maintained (Rubin, 1976). This reduces power, which may be a problem if the sample is not large enough. Most data are in reality Missing not at Random (MNAR), i.e. there is some systematic relationship between the missingness and the participants, or a combination of MCAR and MNAR (Baraldi & Enders, 2010; Rubin, 1976; Sainani, 2015). However, a third classification, Missing at Random (MAR), is usually used in statistical analysis. The term 'random' in MAR is misleading; this assumption actually allows for missingness to be non-random, as long as there are observed covariates in the data source to support reasonable estimation of the missing values (Baraldi & Enders, 2010; Rubin, 1976; Sainani, 2015). MAR is untestable, because testing would require knowledge of unobserved variables, so MAR must be supported by analysis of missing data patterns, literature, and theory (Baraldi & Enders, 2010; Rubin, 1976; Sainani, 2015).

There are two main types of missing data in a longitudinal study like BCS: attrition (cohort member non-response to one or more sweeps), and missing answers to individual items. Attrition can be a permanent cessation of response, which is referred to as 'monotone', or intermittent response, referred to as 'non-monotone' (Mostafa & Wiggins, 2014, p. 5). Both attrition and missing items can cause problems with analysis by reducing sample sizes and statistical power. Additionally, if the pattern of missing data is not completely at random, which is almost always the case, representativeness of the sample changes across sweeps and an undesirable systematic bias is introduced (Rubin, 1976). When this happens, the observed (or unobserved) characteristics of individuals in the sample correlate with the missing responses (Mostafa & Wiggins, 2014, p. 5), and care must be taken to account for this in reporting. Therefore statistical techniques which only analyse complete cases, although still widely used, are not recommended for BCS70 (Allison, 2008; Baraldi & Enders, 2010; Mostafa & Wiggins, 2014; Rubin, 1976; Sainani, 2015).

Only 20%, or about 3,400 of the original sample, participated in all nine BCS70 sweeps (to age 42). This group has much less power and is not fully representative of the individual characteristics in the original cohort group.

An assessment of response to BCS70 sweeps 2-7 (ages 5-34) found that:

> *"Response was lower for cohort members who were men, having a mother who was younger at the birth, a mother who did not attempt to breastfeed, a lower birth weight baby, in a family with 2 or more children, born of non-married parents, a manual father and living in London. Many of these findings are indicators of comparative disadvantage ..."*
>
> *(Ketende, McDonald, & Dex, 2010, p.26)*

Literature has suggested a relationship between comparative disadvantage and ADHD. For example, research using ALSPAC data reported that children with low SES were more likely to have a diagnosis of ADHD at age 7 (Russell et al., 2015). The greatest influence came specifically from financial difficulties, mediated by parent involvement at age 6 and adversity between ages 2 and 4 (Russell, Ford, & Russell, 2015, p. 1). This type of finding underlines the importance of attempting to mitigate and/or account for the loss of low SES participants, as this group may be of particular importance in a study of ADHD symptomatology.

Full-Information Maximum Likelihood (FIML) or multiple imputation methods, as opposed to sample weighting, have been recommended to address the issues of attrition in BCS (Mostafa, 2014). Both methods can account for non-response and missing items, different structures of data (e.g. continuous, categorical, multilevel) and maintain realistic variance in the imputed data, which is lost, introducing further bias, if simpler weighting methods are used (Mostafa & Wiggins, 2014). FIML is preferable to MI because whilst it is equal to MI in robustness (when MAR assumptions are met), it is simpler and more reproducible (Allison, 2018; Williams, 2015b). Thus, FIML was used to handle missing BCS70 data in this thesis, where an assumption of MAR was reasonable.

3.3 Ethics

An ethics checklist for projects overseen by the Faculty of Education at the University of Cambridge was completed and filed with the higher-degrees office. The checklist was derived from British Educational Research Association (BERA) guidelines (BERA, 2011). Only a small number of items on this checklist were relevant, as this project used anonymized secondary data that was made available by the UK Data Service to eligible researchers. The checklist was reviewed, and feedback provided by my supervisor and advisor, to ensure that all relevant matters were considered. These matters included secure storage of data, data and code management, and no attempts made to identify individuals.

The ethical considerations and procedures for BCS70 are currently managed by the Centre for Longitudinal Studies (CLS), and full details can be found in their Ethical Review and Consent report (Shepherd & Gilbert, 2019) and Code of Practice document (Centre for Longitudinal Studies: UCL/IoE, 2014). The Code addresses the requirements of the BERA guidelines, and adheres to the key principles of the Economic and Social Research Council (ESRC) Research Ethics Framework (ESRC, 2016). In brief, the Code sets out to ensure that:

- Cohort member (CM) rights under the Data Protection Act (1998) and Freedom of Information Act (2000) are not affected by the study;
- CM well-being is protected, and strict confidentiality is maintained;
- All data is gathered with informed consent from the appropriate party;
- CMs have a right to refuse participation at any time, and
- Due consideration is taken to ensure participation is not overly time consuming or stressful.

(Centre for Longitudinal Studies: UCL/IoE, 2014)

The Code stipulates that "Anonymised information may be supplied to data archives and others for research purposes only, and on the understanding that no attempt must be made to use it to try to trace, contact, or identify Cohort Members." (Centre for Longitudinal Studies: IoE/UCL, 2014, p. 2) BCS data is accessed via the UK Data Service (University of Essex et al., 2012), and researchers must sign an undertaking form acknowledging understanding of the code of practice, which I have done for the present study.

Additionally, I intend to send a copy of my thesis to the Centre for Longitudinal Studies library, and I am committed to wide dissemination of my findings for the benefit of individuals with ADHD.

3.4 Summary

In summary, the BCS70 study collected a broad set of medical, psychological, educational, social and economic data over ten sweeps in 46 years and is still ongoing. Problems with attrition and missing data are typical for a study of this size and duration, and there are statistical approaches, such as FIML, which can be used to address these problems and minimise the risk of bias in reported results. The ethics procedures used in BCS70 sweeps were appropriate for their time, have been adapted to meet new standards as they have arisen, and currently meet BERA and ESRC standards.

Chapter 3 Developing research questions based on the literature and BCS70

Chapter 3 starts with a review of ADHD as a construct. First the DSM-5 definition is provided, followed by other characteristics commonly observed in clinical practice and research, and a brief discussion of the history and future of ADHD. Next, three dominant theories that have been proposed to explain ADHD are discussed, selection of State Regulation theory as a basis for my research is explained, and a hypothetical model proposed. Third, literature on long-term outcomes is reviewed, taking into consideration the research objective defined in chapter 1, the BCS70 data, and State Regulation theory. The fourth major section discusses learning from the ADHD long-term outcomes literature review, and selection of two constructs, subjective wellbeing and educational attainment, to be evaluated as outcomes in this thesis. The fifth section discusses operationalisation of two sets of constructs to facilitate testing evidence for State Regulation theory: stressors, and protective factors against stress. At the end of this chapter the four thesis research questions are stated.

1 ADHD the construct

1.1 DSM-5 ADHD

ADHD is understood to be a behavioural disorder. A clinical diagnosis is made by a clinical professional using questionnaires, interviews and observation, usually with input from parents and teachers; there is no diagnostic biological test. In research, questionnaire results are often used as a proxy for diagnosis. Both diagnosis and questionnaires used in research are usually based on the criteria defined in the DSM-5 (American Psychiatric Association, 2013).

> *The ADHD criteria include nine hyperactive/impulsive symptoms, nine inattentive symptoms, and six additional conditions. Six from either or both lists of nine must be met plus all six conditions to make a diagnosis. A subtype is determined based on which sets of criteria are met (hyperactive, inattentive or both/combined).*
>
> *Examples[8] of the nine DSM-5 hyperactive criteria include: fidgets/squirms, acts as if driven by a motor, leaves seat unexpectedly, interrupts or intrudes.*
>
> *Examples of inattentive criteria are easily distracted, forgetful, disorganised, fails to finish tasks.*
>
> *The six conditions are: several symptoms observed before age 12, persist for at least six months, occur in two or more settings, evidence symptoms interfere with*

[8] The full list of DSM-5 ADHD criteria are included in the appendices for chapter 3.

functioning, do not happen only during psychosis, and not better explained by another mental disorder.

Severity can be indicated by the diagnostician as mild, moderate, or severe.

As mentioned in the introduction, there is a second definition sometimes used in diagnosis and research. The ICD-10 definition of hyperkinetic disorder (HKD, coded F90) is similar to ADHD, but requires that both hyperactive and inattentive symptoms are present (like the combined type in DSM-5), and that symptoms are severe and present before age 6 (World Health Organization, 1994). Thus HKD corresponds to only the most severe and early-onset cases of ADHD (Lee et al., 2008).

A recent summary of meta-analytic studies reported that the community prevalence (not necessarily diagnosed) of ADHD ranges from 2-7%, with an average of 5%, whilst administrative prevalence based on prescription rates varies from 0.06-2.5% (Sayal et al., 2018). In community samples, the ratio of boys to girls is between about 2:1 and 3:1, and primarily inattentive is the most common subtype (Sayal et al., 2018; Willcutt, 2012). Clinical samples are biased towards boys and combined type ADHD (Willcutt, 2012). Although there have been media reports that ADHD has increased in prevalence over time, meta-analysis does not support a significant change over the last three decades (Thomas et al., 2015). As mentioned in the introduction chapter, it is estimated that some impairment related to ADHD symptoms continues into adulthood for about 50% of those affected in childhood (Costello & Maughan, 2015).

1.2 Additional characteristics

There are several additional characteristics of ADHD that have been widely observed and reported in research, but do not form part of the diagnostic criteria. They include intra-individual variability, hyperfocus, social difficulties, and heterogeneity amongst cases.

1.2.1 Intra-individual variability

Studies have observed ADHD participants to have greater intra-individual variability than controls in motor activity, mood, and attention (Kuntsi & Klein, 2011), and this has been one of the only consistent findings across ADHD research (Castellanos & Tannock, 2002; Jonna Kuntsi & Klein, 2011). The most widely replicated research has found that whilst mean reaction times do not necessarily differ for ADHD, reaction time variation (i.e. standard deviation) does (Kuntsi et al., 2001; Kuntsi et al., 2010; Kuntsi & Klein, 2011; Metin et al., 2014; Sjöwall et al., 2013; Uebel et al., 2010). Also, non-ADHD siblings of ADHD participants have been found to differ from typically-developing controls in reaction time variability (RTV), though not by as much as those with ADHD, which supports genetic influences (Uebel et al., 2010), and a continuous rather than categorical construct. Replicability and consistency of findings is rare in ADHD research, and thus RTV has been regularly discussed as a key factor in theories of cause and effect (Kuntsi & Klein, 2011; Sergeant, 2000). Intra-individual variability suggests a regulation difference or difficulty rather than a 'deficit', which is integral to the DSM-based nomenclature.

1.2.2 Hyperfocus

ADHD children who are unable to pay attention to assigned tasks have been observed to have an unusual ability to focus intently on specific activities that interest them, and difficulty switching attention away from them (Brown, 2013; Greenberg et al., 2007; Hupfeld et al., 2019; Lovecky, 1999; Mayes & Calhoun, 2007; Thompson & Thompson, 1998). This is called hyperfocus, and a common example is hours spent playing video games. It has been noted particularly with gifted ADHD children, and viewed as a coping mechanism (Lovecky, 1999). Hyperfocus also supports the idea of a regulation problem rather than deficit; i.e. sometimes there is a surplus of attention, sometimes a deficit.

1.2.3 Social difficulties

It has been estimated that as many as 80% of children with ADHD are rejected by their peers (Hoza et al., 2005; Merrell & Wolfe, 1998). Here is a poignant example: "…in a play group study that involved placing children with ADHD in groups with unfamiliar non-ADHD peers, the non-ADHD participants began complaining about the behaviour of their ADHD peers within minutes." (William E. Pelham & Bender, 1982, p. 656). Peer problems are indicated to be an important predictor of poor adjustment and long-term life outcomes (Hoza, 2007; McQuade et al., 2014), and accordingly this is an important area of focus for ADHD research and management (Chronis et al., 2006; Cotton et al., 2015; de Boo & Prins, 2007; Gol & Jarus, 2005; Pfiffner et al., 2013; Pfiffner & McBurnett, 1997; Ole Jakob Storebø et al., 2012; Wilkes-Gillan et al., 2016).

1.2.4 Heterogeneity, equifinality and multifinality

ADHD children are a more heterogeneous group inter-individually than what is suggested by the three DSM-5 subtypes (Luo et al., 2019). Many different patterns of symptoms can combine to indicate an ADHD diagnosis (for an illustration see chapter 4, Table 7). Additionally, as many as 70% of children with ADHD are diagnosed with another (comorbid) disorder. The most common comorbidities are learning disability, oppositional defiant disorder, conduct disorder, anxiety, and depression (Brown, 2013; Jensen et al., 2001).

However, it is important to note that ADHD affects children across the full spectrum of IQ (Brown, 2013), even those who are considered gifted [9] (Antshel et al., 2007).

Numerous explanations have been explored for possible root causes and trajectories leading to ADHD. There is evidence for correlational relationships at the group level between ADHD and polygenic risk scores, observable brain characteristics, environmental toxins, pre/perinatal, dietary, and psychosocial factors, but none have been established as causal on their own (Sciberras et al., 2017; Thapar et al., 2013). Even for strong risk factors, such as low birth weight, they are only found in a proportion of ADHD cases (Sciberras et al., 2017; Thapar et al., 2013). There is significant evidence from twin, adoption, genetic, RCT, and longitudinal studies to conclude that ADHD is consistently highly heritable (0.70 or higher), but ultimately manifests from an interaction between genes and environmental factors (Thapar et al., 2013). Accordingly, Thapar et al. (2013, p. 12) asserts that it is incorrect and unhelpful to think of or communicate about ADHD in solely genetic or environmental terms.

Additionally ADHD has been associated with numerous somatic health problems, including metabolic, atopic, auto-immune conditions, and more (Chen et al., 2014; Spencer et al., 2014; World Federation of ADHD, 2015, 2017, 2019). Experts have pondered whether ADHD might be more "like a fever", i.e. a general symptom presenting as a result of many different starting points and pathways (World Federation of ADHD, 2015).

The concept of different starting points and pathways that lead to a single observed phenomenon (i.e. a cluster of behaviours in the case of ADHD) is more concisely encapsulated by the terms equifinality and multifinality. Both of these concepts are characteristics of open systems, and are widely attributed to a classic description of general system theory (von Bertalanffy, 1968). Brief definitions are listed below:

- An open system[10] interacts with its environment via feedback loops and the interaction has an effect on the outcome
    - This is in contrast to a closed system, in which the same input will always produce the same outcome
- Equifinality – in an open system, multiple inputs and multiple processes produce the same outcome
- Multifinality – in an open system, similar inputs can lead to dissimilar outcomes

(Cicchetti & Rogosch, 1996; von Bertalanffy, 1968).

---

[9] Gifted children with ADHD are sometimes referred to as 'twice exceptional' (Budding & Chidekel, 2012)

[10] Examples of open systems include: all living biological organisms, most social organisations, artificial intelligence networks, market-based economies, democratic government, and evolution.

Extensive evidence shows that equifinality and multifinality are consistent features of ADHD, thus any theoretical explanation of causal mechanisms should depict an open system.

1.3 Brief discussion of the history and future of ADHD

The concept of children who are more inattentive and/or hyperactive than the norm has probably always existed in some form. A report by a Scottish medical doctor in the late 18th century described "The incapacity of attending with a necessary degree of constancy to any one object…" and noted that when this occurs in children, it affects their ability to receive an education (Crichton, 1798 p. 203, as cited in Lange, Reichl, Lange, Tucha, & Tucha, 2010). More recent labels for ADHD-like behaviour have included minimal brain dysfunction, hyperkinetic reaction of childhood, and attention deficit disorder (ADD), with or without hyperactivity, in the DSM versions I (1952), II (1968), and III (1980), respectively. Attention Deficit Hyperactivity Disorder, or ADHD, has been the officially recognized term since the DSM-III-R, which was published in 1987 (American Psychiatric Association, 2016).

It is not clear or perhaps desirable that the construct of ADHD as it is currently defined will survive indefinitely. A recent analysis of a selection of diagnoses from the DSM-5 (not including ADHD), reported that symptoms between disorders overlapped considerably and had internal inconsistencies, particularly with regard to the role of stress or trauma (Allsopp et al., 2019). The authors concluded that whilst this ambiguity may be useful in clinical settings to aid in diagnosis, it precludes discrete categorisation and obstructs the linking of research on clinical constructs to biological or social causal pathways (Allsopp et al., 2019).

Accordingly, there are at least three significant initiatives currently underway to strengthen the link between ADHD specifically and/or psychiatric constructs more broadly and causal pathways. First, specifically in ADHD research, work has been ongoing for nearly two decades to identify subtypes related to genetic phenotypes, other biological correlates, and/or well-established psychological constructs that could guide targeted treatment plans (Castellanos & Tannock, 2002; Karalunas et al., 2014; Karalunas & Nigg, 2019; Willcutt et al., 2012). Secondly and more broadly, the Research Domain Criteria (RDoC) initiative has been underway since 2010 to reclassify psychiatric disorders into categories that correlate with observable biological characteristics (National Institute of Mental Health, 2019). For example, a recent and exciting study of children struggling in school (*N*=530) used machine learning algorithms to identify four cognitive and behavioural clusters that aligned with biological characteristics: 1) broad-cognitive difficulties, 2) age-typical cognitive abilities[11], 3) working memory difficulties, and 4) phonological difficulties (Astle et al., 2019). Thirdly, the

[11] See protocol for cognitive abilities measured: http://calm.mrc-cbu.cam.ac.uk/protocol/

Hierarchical Taxonomy of Psychopathology (HiToP) has been created to redefine psychiatric constructs using spectra, which align with clinical practice, and use continuous measures rather than categorical ones (Kotov et al., 2017). It is unclear at this time how these classification systems will be integrated into research and practice going forward. I searched Web of Science (Clairivate Analytics, 2019) for publications in 2018-2019, and found that ADHD and DSM[12] were included in the title or topic metadata of 430 papers, ADHD and RDoC in 42, and ADHD and HiToP in two. Based on this, DSM-defined ADHD still appears to be the most widely used classification.

2 Causal theories

As highlighted by the discussion above on the future of ADHD, lack of a well-supported and widely agreed explanatory systems theory is a problem area in the field. There are causal theories, however, and three of the most influential ones (Johnson et al., 2009) are considered here. They are: Executive Dysfunction (Barkley, 1997), Dynamic Developmental (Sagvolden et al., 2005), and State Regulation (Sanders, 1983, 1998; Sergeant, 2000).

2.1 Executive Function

The Executive Function theory (Barkley, 1997) proposed that ADHD symptoms are a problem with executive functions (EF), and inferred that these are in turn caused by problems with underlying neural structure and function (Johnson et al., 2009). Executive function theory has been featured or implied in a large number of studies over the last two decades (e.g. Brown, 2013; Happé, Booth, Charlton, & Hughes, 2006; Kempton et al., 1999; Semrud-Clikeman, Walkowiak, Wilkinson, & Butcher, 2010), starting around 1997, when it was first published (Barkley, 1997). Executive function is a term used to describe a collection of cognitive functions associated with top-down, effortful control. Core EFs include inhibitory control, working memory, and cognitive flexibility; higher level EFs include planning, problem solving, and reasoning (Diamond, 2013). A rigorous meta-analytic study examined 83 studies of EF and ADHD, involving a total of 3,734 ADHD participants (Willcutt et al., 2005). The evidence overall showed consistent and significant group differences in EF between ADHD and non-ADHD groups; however, they found that "fewer than half of children with ADHD exhibit significant impairment on any specific EF tasks" (Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005, p. 1342). The authors concluded that whilst there may be an epiphenomenal association between EF and ADHD, it is unlikely there is a cause and effect relationship (Willcutt et al., 2005).

[12] Searches included both abbreviated and unabbreviated terms

2.2 Dynamic Developmental

The second theory of the three is Dynamic Developmental (Sagvolden et al., 2005), which is derived from the principles of operant conditioning. This theory proposes that the process of conditioning appropriate behaviour fails in ADHD children, because there is a shorter amount of time after a behaviour event when reinforcement is effective. The suggested underlying cause is insufficient functioning of the meso-cortical dopamine branch (Sagvolden et al., 2005, p. 397). This explanation would support use of stimulant medication to increase dopamine levels, and behavioural training that maximizes reinforcement speed and consistency as ideal interventions for ADHD. Computer games were suggested as particularly promising because the speed and consistency would be better than that of a human 'reinforcer' (Sagvolden et al., 2005). Dynamic Developmental theory is considered a possible explanation for hyperactive and combined sub-types of ADHD, but not inattentive (Johnson et al., 2009; Sagvolden et al., 2005), which is the most prevalent presentation in community samples (Willcutt, 2012). The theory contains a thorough summary and well-argued interpretation of evidence. However, it is not able to explain the prevalent inattentive sub-type, stimulant medication non-response in approximately one-third of patients (Elliott et al., 2017; Greenhill et al., 1999), or the lack of treatment effects found in probably blinded assessments of cognitive training and behavioural interventions (Sonuga-Barke et al., 2013).

2.3 State Regulation

State Regulation theory is a cognitive-energetic model, which explains information processing efficiency through interactions between evaluation, effort, arousal, and activation (Sanders, 1983). The model was not originally articulated to explain ADHD, but has been widely adapted for ADHD research (Gozal & Molfese, 2005; Johnson et al., 2009; Kuntsi et al., 2010; Metin et al., 2014; Sergeant, 2000, 2004). State Regulation theory evolved from Yerkes-Dodson law (1908); an inverted u-shaped relationship between arousal and performance. See Figure 1 below:

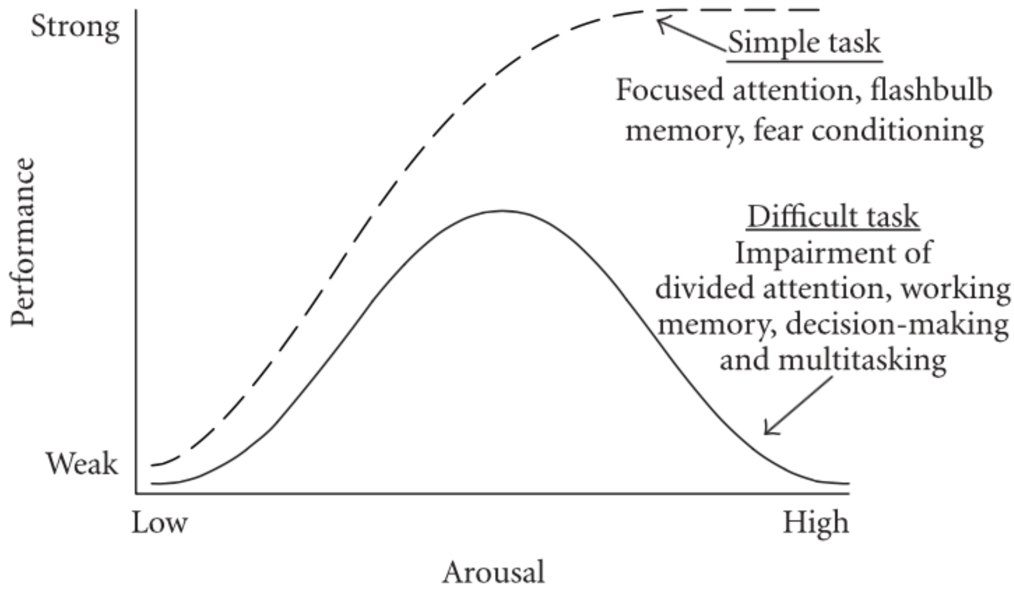


Figure 1. Reprinted graph representing Yerkes-Dodson law (1908) (Diamond, Campbell, Park, Halonen, & Zoladz, 2007, p. 3).

N.B. This was the original version; the version more often reprinted does not include the higher curve labelled 'simple task' (Diamond et al., 2007).

It has long been argued that stress and arousal interact to produce performance (Salas et al., 1996; Welford, 1973), in line with the Yerkes-Dodson law. Sanders (1983) proposed that this stress-arousal-performance relationship was over-simplified because it was unidimensional. He argued that that stress cannot be reliably described as a stimulus or a response, but rather a multi-stage interaction between the two, with feedback loops (Sanders, 1983), i.e. an open system. The model builds on and links to classic information processing models in attention research (e.g. Broadbent, 1958; Posner, 1980), as well Kahneman's (1973) theory of effort. Sanders (1983) defines four stages of processing, as follows:

- Stimulus pre-processing
- Feature extraction
- Response choice
- Motor adjustment

The model also incorporates arousal, effort, activation, evaluation, and feedback loops, as shown in Figure 2 below, which is reproduced from Sanders' (1983) paper.

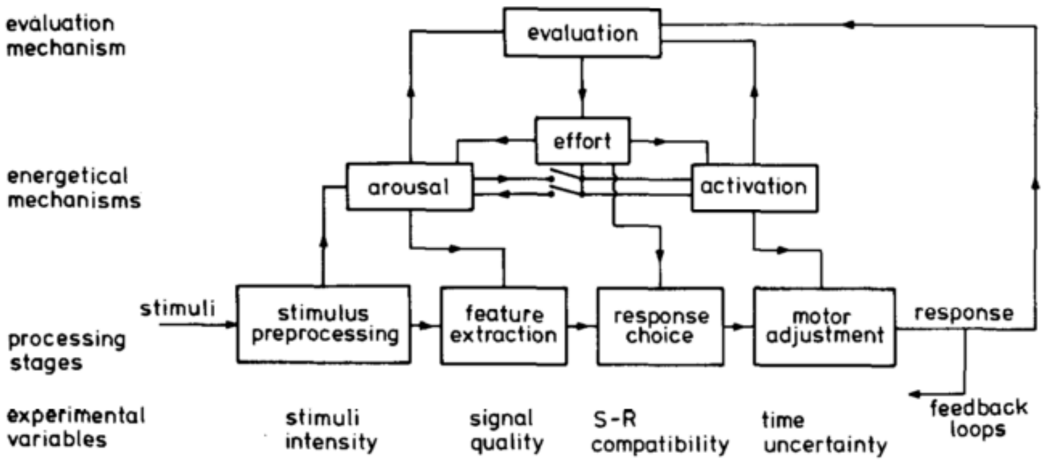


Fig. 1. A cognitive-energetical linear stage model of human information processing and stress. The cognitive level consists of computational processing stages derived my means of the additive factor method (Sternberg 1969). There are three energetical supply mechanisms, two of which are basal (arousal and activation) and coupled to respectively input and output processing stages. The basal mechanisms are coordinated and supervised by effort, which is also directly linked to the central stage of response choice. Apart from direct energetical supply to this stage, effort serves the function of keeping the basal mechanisms at an optimal value. Information about the state of the basal mechanisms is mediated by an evaluation mechanism. For further details, see text.

Figure 2. Reprinted Sanders (1983) model of stress and performance (p. 79)

The accompanying assertion of the model is (as also noted in chapter 1):

> *"Stress will arise whenever the effort mechanism is either seriously overloaded over time or falls altogether short in accomplishing the necessary energetical adjustments." (Sanders, 1983, p. 79)*

State Regulation theory assumes that automatic response does not require effort but controlled/top-down responding does (Sanders, 1983). Stimulus pre-processing is assumed to be automatic, whilst the other three stages require effort, and the amount of effort required probably increases by stage (Sanders, 1983, p. 74). 'Top-down/controlled responding' in today's terminology corresponds to executive functions. In this model stress occurs because the system cannot produce a level of effort over time or in a specific situation to keep the system in equilibrium; i.e. the appropriate amount or type of activation, arousal, or evaluation cannot be achieved or maintained (Sanders, 1983). This is proposed to explain five patterns of stress: over and under activation, over and under arousal, and insufficient reasoning/decision making (Sanders, 1983).

State Regulation theory was selected as the theoretical basis for this thesis. However, using BCS70 data, it was not possible to engage with State Regulation theory at the level of detail shown in Figure 2 and described above. Instead, it was assumed that the underlying process happens in a 'black box', and relationships were modelled at the level of psychosocial constructs. ADHD symptoms (or related constructs, for example internalising/anxiety/depression or externalising/conduct problems) were assumed to be emotional and/or behavioural responses to one or more of the five patterns of stress.

2.3.1 Stress, anxiety, and ADHD

It is worth noting here that the concepts of anxiety and stress can be closely linked or overlap significantly in literature. To demonstrate by example, acute stress disorder was classed as an anxiety disorder in DSM-IV and as a trauma/stress disorder in DSM-5 (Substance Abuse and Mental Health Services Administration (US), 2016). In both DSM versions, acute stress included several of the same symptoms, including: sleep problems, irritability, hypervigilance, difficulty concentrating, and an exaggerated startle response (American Psychiatric Association, 2013a; Substance Abuse and Mental Health Services Administration (US), 2016). Of these, sleep problems, irritability, and difficulty concentrating are also symptoms of generalised anxiety disorder (American Psychiatric Association, 2013b). Decades ago a review attempted to address confusion between stress and anxiety by proposing that generally stress should be seen as a feeling triggered by an external event (either a single event or ongoing), whilst anxiety can be either a trait (a tendency attributable to an individual), or a state, which is an emotional response to stress, and these three constructs, plus coping mechanisms, interact multi-directionally (Endler & Parker, 1990). The treatment of stress and anxiety as constructs in this thesis is in accord with the Endler & Parker (1990) review.

There is also overlap between the DSM-defined symptoms of stress and anxiety with symptoms of ADHD (e.g. restlessness, difficulty concentrating). As described previously (section 1.3), this overlap is problematic in numerous DSM definitions and hinders research aiming to link psychological constructs to causal pathways (Allsopp et al., 2019).

2.3.2 Evidence supporting State Regulation theory

State regulation theory does not have significant contradictory evidence against it like executive dysfunction and dynamic developmental do. The model in Figure 2 depicts an open system (with feedback loops) and supports the equifinality and multifinality observed in ADHD. It could explain all sub-types of ADHD and is compatible with key observed features including intra-individual variability, hyperfocus, and comorbidity. Support for State Regulation theory is also demonstrated in recent literature trends towards publishing and readership of papers on interventions that should reduce stress and/or improve the ability to cope with stress. Examples include: increasing play time with peers and parents (Cordier et al., 2009; Halperin et al., 2013; Panksepp, 2007; Whitebread, 2017; Wilkes-Gillan et al., 2016), mindfulness meditation training (Cairncross & Miller, 2016; SAGE journals and APSARD, 2019; Tang et al., 2007), yoga/physical exercise (Chimiklis et al., 2018; Diamond & Lee, 2011; Frazier & Wilson, 2018; Smith et al., 2013), and coaching on development of 'if-then' situation-behaviour strategies (Gawrilow et al., 2011, 2012; Gawrilow & Gollwitzer, 2007; Hung & Gawrilow, 2016).

2.3.2.1 Support from research on executive functions

Executive functions (EFs) are a central topic in psychology and education because they have been shown to predict academic readiness and achievement as well as economic, social and health outcomes (Diamond & Lee, 2011). In their seminal review, Diamond & Lee (2011) evaluated the results of interventions to improve EFs in children. The review assessed studies using computerized training, non-computerized game-based training, fitness/exercise programs, mindfulness, and partial and entire pedagogies targeting EF development (Diamond & Lee, 2011). The authors concluded that children with low EF abilities benefit most from interventions, and many approaches are probably viable, particularly if they demand high levels and multiple components of EF, continually increase the challenge, and consistently devote significant time to the activity (Diamond & Lee, 2011). Their parting conclusions about successful EF interventions were:

- "They do not expect young children to sit still for long;
- …programs tend to reduce stress in the classroom, cultivate joy, pride, and self-confidence, and foster social bonding; and
- "…stress, loneliness, and lack of physical fitness impair prefrontal cortex function and EFs."

(Diamond & Lee, 2011, p. 963)

In a subsequent, more detailed review, these points were reiterated, and it was proposed that impairment to EF may be an early warning system that something is wrong; an analogy was drawn between a child exhibiting poor EF and a "canary in the coal mine" (Diamond, 2013, p. 153). Diamond (2013) also suggests that stress can impair EF in a way that makes it look like a person has ADHD when they do not. Thus, even if problems with executive functions is an epiphenomenon with ADHD, the study of EF may shed important light on optimizing ADHD outcomes.

It is well-established that many cases of ADHD remit (at least clinical ADHD) in adolescence and adulthood (Costello & Maughan, 2015; Faraone et al., 2005), and also that siblings of children with ADHD show tendencies towards the disorder without meeting criteria for diagnosis (Uebel et al., 2010). Accordingly, there is wide agreement in the literature that ADHD is a best represented as a continuous trait rather than a categorical one. Therefore, perhaps a nuanced interpretation of Diamond's (2013) statement is this: a person with a predisposition to ADHD may move into higher levels of the continuum, and experience impairment to EF, because of stress.

2.3.2.2 Stress and programming

Lupien, McEwen, Gunnar, & Heim, (2009) discussed human and animal evidence for 'programming': "when an environmental factor that acts during a sensitive period affects the structure and function of tissues, leading to effects that persist throughout life." (Lupien et al., 2009, p. 434). The hypothalamus-pituitary-adrenal (HPA) axis is activated by stress. The authors reported that "chronic or repeated exposure to stress has enduring effects on the brain" (Lupien et al., 2009, p. 440). More specifically, children with early, pre and post-natal experience of stress are more sensitized to stress in the future (Lupien et al., 2009).

Perhaps children who are biologically sensitized to stress are more likely to have problems with EFs, and/or develop ADHD? This idea is compatible with the concept of equifinality observed in EF and ADHD research, through correlations with many different but all arguably stressful early-life circumstances, including low birth weight, pre-natal smoking, exposure to environmental toxins, and extreme deprivation (Nigg, 2006; Sonuga-Barke et al., 2017; Thapar et al., 2013).

Taking into consideration the evidence discussed above, I concluded that State Regulation theory has the greatest number of strengths and fewest weaknesses of the main ADHD causal theories, and thus, although I did not have data available to engage with the biological level of the model, it was the basis of my research at a conceptual level.

2.3.3 Conceptual model for operationalising State Regulation theory

Building on evidence that ADHD arises from interactions between genetic influences, programming effects on brain development and stress sensitivity, and an operationalisation of state regulation theory as psychosocial stressors and protective factors against stress, I proposed a hypothesis of relationships between high-level concepts in Figure 3:

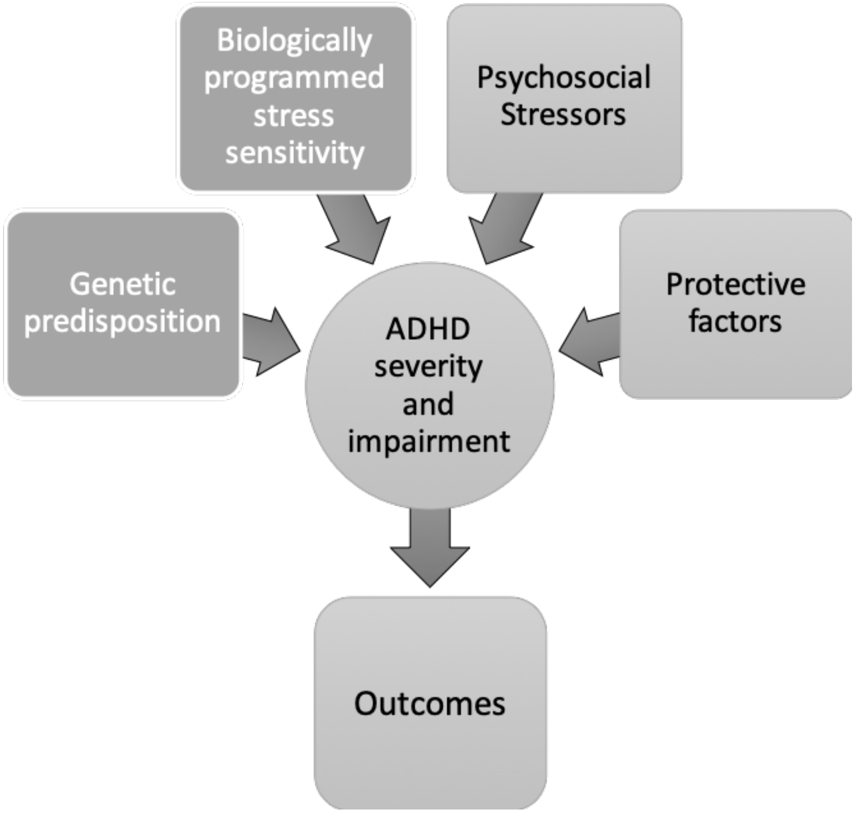


Figure 3. Hypothesized conceptual model of stress, ADHD, and outcomes

The grey boxes with white text are out of scope for this thesis, because biological data was not available in the BCS70 up to the age 42 sweep.

The conceptual model in Figure 3 could account for significant heterogeneity, including partial heritability estimates, comorbidity (other health problems could be a result of, or cause of additional stress), and importantly, the intra-individual variability (IIV) of performance on cognitive tasks. As highlighted previously, IIV is one of the only characteristics that appears consistently across cases of ADHD (Castellanos & Tannock, 2002; Kuntsi & Klein, 2011).

The model implies there are indirect effects, i.e. that stressors and protective factors effect outcomes indirectly, through ADHD severity and impairment. Indirect effects portray equifinality and an open system, i.e. that outcomes may vary, depending on factors other than a proposed primary predictor (Alwin & Hauser, 1975). Indirect effects are also referred to as third-variable effects, or more commonly, mediators and moderators (Baron & Kenny, 1986). Mediators are variables that have a strong relationship with both a predictor and an outcome, and they are a necessary part of a causal chain (Baron & Kenny, 1986). Moderators are not part of the causal chain, but have an effect on the strength or direction of a relationship between a predictor and an outcome (Baron & Kenny, 1986). I intended to test indirect mediator and/or moderator effects in my thesis, but ultimately did not. Indirect relationship modelling is complex, and the pre-requisite data preparation and modelling of simpler direct effects consumed all the space allotted to the thesis, so the work was deferred to a future phase of work. Further explanation is included in section 5.1 of chapter 9.

3 Long-term outcomes in ADHD

At the end of chapter 1 my stated research objective was to evaluate how psychosocial factors for individuals with ADHD in childhood related to positively framed outcomes in settled adulthood. With that in mind, I selected and described the BCS70 dataset in chapter 2, and so far in chapter 3 have discussed the construct of ADHD, causal theories, and selected state regulation as a theoretical basis for my research. The next step is to review literature on long-term outcomes in ADHD, to inform the development of research questions that can be answered using BCS70.

3.1 Literature search

Life outcomes cover a broad range of areas, and it is not possible or desirable to explore every outcome that has been related to ADHD within the resource constraints of a PhD thesis. Considerable literature is dedicated to the study of the relationships between ADHD and specific outcomes such as crime, obesity, substance use, and reduced symptoms as a response to treatment with medication and/or non-pharmacological interventions. All the above require engagement with specialist literature which is outside the scope of my study. Therefore, all studies focused solely on those outcomes were excluded from my search.

Additionally, based on literature reviewed so far, I identified the following criteria for my search of literature on ADHD outcomes:

- Outcomes reported in settled adulthood (ideally over age 30)
- ADHD identified in childhood (ideally a prospective measure)
- Reporting of statistics that allow for comparison of findings (effect sizes, or components of effect size)
- Large samples, good representativeness, or review of multiple studies
- Focus not specifically on persistence of ADHD symptoms into adulthood (data not available in BCS70)
- Published in the year 2000 or later (reviews capture earlier findings)

To ensure completeness and rigor of my review, I consulted the Cambridge library literature search guide (Cambridge Libraries, 2019), and the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-P). I did not conduct a full PRISMA-P review, but did include some aspects of the protocol, e.g. the review process and diagram.

Following library guidance for psychology and education research (Cambridge Libraries, 2019), I used four database/search engines: EBSCOHost (EBSCO Industries Inc, 2019), Web of Science (Clairivate Analytics, 2019), PubMed (National Center for Biotechnology Information and U.S. National Library of Medicine, 2019), and Scopus (Elsevier, 2019). The search terms, exclusions, and counts of returned search results are reported in Table 2.

Database(s)	Search Terms	Exclusions (Not)	Refined by	Results
EBSCOHost: ERIC, BEI, PsycINFO	(ADHD, Attention Deficit Hyperactivity Disorder, Attention-Deficit/Hyperactivity Disorder) + (outcomes, adult, long term, long-term, longitudinal, panel)	Methylphenidate, atomoxetine, pharmacological, crime, criminal, prison, animal, lisdexamfetamine, treatment, cbt, intervention, mph, psychostimulant, medication	Search title, peer reviewed journals, published 2000-2019, English, human subjects, full text available	230
Web of Science	Same as above	Same as above	Title search, 2000-2019	297
PubMed	Same as above	Same as above	Same as above	262
Scopus	Same as above	Same as above	Same as above	250
Total				1,039

Table 2. Search criteria used in literature review of long-term outcomes for ADHD

Result lists were exported to csv or xls files and combined into a single table in an MS Access database. Duplicate titles were identified using an SQL query (group by unique titles). Next, titles were reviewed by eye for duplicates to identify those not found by the query because of non-substantive differences like typographical errors, spaces, or use of different case (e.g. upper vs. proper). 506 unique studies remained to be reviewed further. Titles were reviewed again to identify themes for further exclusion criteria. The themes identified were:

- Erroneous match – study not related to ADHD. Most of these were picked up because 'ADD' was used as a search term.
- Age - study participants did not cover the desired time span, i.e. they did not have ADHD measured in childhood and outcomes measured in adulthood. Many of these were intervention studies which used the phrase 'long-term outcomes' to describe a relatively short period (e.g. 18 months, 2 years, etc.) compared to the longer period of interest in the present thesis.
- Sub-population – studies focused on a special sub-population of ADHD, for example boys with conduct disorder, or babies with low birth weight.
- Other psychological or psychiatric focus – studies of specific constructs related to ADHD, such as working memory, ASD, or substance abuse disorder.
- Medication – studies of the effects of medication on outcomes, most of which were relatively short-term.
- Crime – studies of ADHD in relation specifically to criminal activity.
- Neuroscience – studies relating data about neural/brain structure characteristics and ADHD outcomes – data not available in BCS70.
- Genetics – studies of genes related to ADHD outcomes – data not available in BCS70.

- FTNA – full text not available, or not available in English.
- Methods deemed to be low quality.

The process was repeated in two iterations to reduce the likelihood of human error and to extend the list of reasons as needed. Each of the 506 was marked with an exclusion reason or flagged 'keep'. 37 studies survived this stage of review (Figure 4).

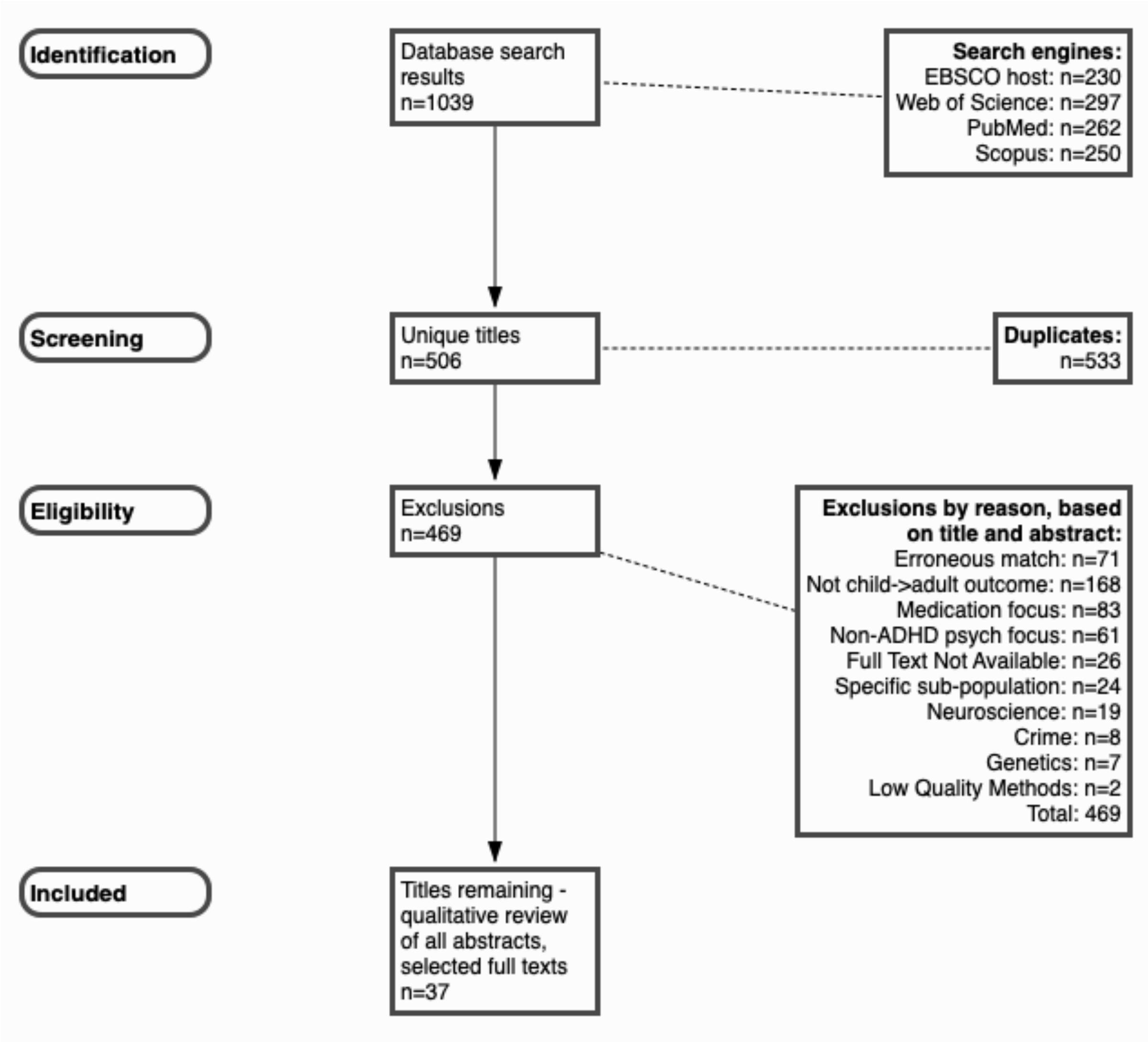


Figure 4. Diagram of systematic literature review process, based on PRISMA-P (Moher et al., 2015; Shamseer et al., 2015)

Next, abstracts and/or full-texts of the 37 studies were reviewed, and a further 30 were excluded because key data provided was not available in BCS70 for comparison (e.g. lifetime earnings), the studies were methods reviews or commentaries, or were already included (double-counted) as part of one of the two reviews/meta-analyses in my final set of seven (Table 3).

No.	Study	N (ADHD)	Findings	Notes
1.	Brassett-Grundy & Butler, 2004	721	ADHD as defined had significant independent contribution to low education level, benefits, unemployment, low income, poor housing, single parenthood, no partner, crime, accidents, substance use, life satisfaction, and mental health, for both men and women.	Sample identified using several non-ADHD items (per DSM). No adjustment for multiple tests, did not report effect sizes. Controlled for large number of confounds. ADHD N at 10 and 30 = 721. Comparison N was remaining sample, varied with non-missingness on each outcome.
2.	Cherkasova et al. 2013	11 studies	Outcomes for ADHD were worse in education, occupation, mental health, criminality, driving, relationships, and sexual behaviour.	Selective review. Reported effect sizes (Cohen's d). ADHD severity, EFs, parenting practices, and persistent adult ADHD were reported as influential predictors of outcomes.
3.	Costello & Maughan, 2015	N/A	Outcomes mediated by symptom severity, comorbidity (conduct disorder), education, SES, and IQ.	Qualitative review. Found few studies able to answer questions about optimal outcomes. Call for more evidence on psychosocial factors. Review – N not reported. Concluded 50% of adults not impaired.
4.	Erskine et al., 2016	98 studies	Significant effects of ADHD on numerous outcome areas, including mental health, education, and crime. Particularly large effects: ODD, CD, Bipolar, suspension from school, service use in education.	PRISMA review – age at outcome measure varied widely. Included several other studies I found as candidates for inclusion in my review.
5.	Hechtman et al., 2016	476	ADHD group worse in educational attainment, receipt of benefits, emotional lability/neuroticism, mental health (persisters only), risky sexual behaviour. Not worse for criminality or substance use.	MTA follow-up. Symptom persistence key factor in worse outcomes. Retention at 16-yr follow-up was 82%, attrition was biased towards disadvantage, including parent mental health problems, and higher ADHD severity.

6.	Roy et al., 2017	579	Effects of childhood factors on adult outcomes comparable for ADHD group and controls (n = 258). Educational attainment particularly important factor in other outcomes.	Further analysis of MTA follow-up data. Key childhood predictors of outcomes overall were household income, ADHD severity, and IQ. Key predictors of educational attainment were IQ, parent education, parenting style, parent marital problems, and ADHD severity.
7.	Shaw et al., 2012	351 studies	ADHD group worse in education, substance use, employment, self-esteem, driving accidents, antisocial behaviour, social functioning, and obesity.	Meta-analysis Study funded by pharma Meds improved outcomes but did not normalise

Table 3. Summary of seven selected studies on long-term outcomes in ADHD

3.2 Discussion of selected ADHD outcome studies

I selected three of the seven studies to review in further detail: Brassett-Grundy & Butler, 2004; Erskine et al., 2016; and Hechtman et al., 2016. The other studies were not discussed further for the following reasons: Cherkasova et al., (2013) because 10/11 studies in the review overlapped with the Erskine review, Costello & Maughan (2013) because it was a qualitative review, Roy et al., (2017) because it focused on potential mediators and moderators of ADHD on outcomes, and Shaw et al., (2012) because it focused on the effects of medication, and only evaluated statistical significance, not effect sizes.

3.2.1 ADHD in BCS70

First, I reviewed a working paper that reported on ADHD measured in the BCS70 at age 10 and outcomes measured at age 30 (Brassett-Grundy & Butler, 2004), as it is the most pertinent paper to this thesis. Their measure of ADHD was derived based on Conners (Conners, 1969) and Rutter (Rutter, 1967) child behaviour scales as administered in the age 10 sweep (Brassett-Grundy & Butler, 2004; Butler et al., 1997), and they evaluated 24 social, economic, and health outcomes at age 30 (Brassett-Grundy & Butler, 2004). All outcomes were coded as binary and framed in a negative way. Examples included: indicators of no qualifications, low qualifications, low income, worklessness, single parenthood, criminal offending, substance use problems, and psychiatric problems (Brassett-Grundy & Butler, 2004).

They controlled for confounding factors using 38 data items: 14 from age 0 and 24 from age 5, which included, for example: birth weight, breastfed, ethnicity, birth region, and parent social class and education level (a full list is provided in the chapter 3 appendix; Brassett-Grundy &

Butler, 2004). They evaluated mutually exclusive categories/levels of each data item as a separate dummy (0/1) variable; for example, the breastfed variable had three levels: breastfed, never breastfed, and unknown, so they created three dummies, one for each level (Brassett-Grundy & Butler, 2004, p. 19). This dummy variable approach meant that the actual number of variables controlled for expanded from 38 to 136.

They ran 47 probit regressions which comprised two separate sets: 23 for males and 24 females[13] (Brassett-Grundy & Butler, 2004). In these regressions, controlling for 136 binary covariates, they found that belonging to their ADHD subgroup was a significant predictor of 18/23 poor outcomes for men, and 15/24 for women (Brassett-Grundy & Butler, 2004). They did not report standardised effect sizes, but z-values, sample means, and sample sizes were provided, which do support computation of effect sizes post-hoc.

The study had some limitations. For example, they used listwise deletion of missing data and did not assess or account for bias from missingness. Several of the items used to identify ADHD are not part of current DSM ADHD criteria. Also they noted the control variables were selected based on literature (Brassett-Grundy & Butler, 2004) but were not specific about literature sources. Their approach entailed running 47 separate tests, and the significance (p-value) threshold was not adjusted for multiple tests (Brassett-Grundy & Butler, 2004), so there was a considerable risk of one or more 'false positives', i.e. significant effects that are misleading (Benjamini & Hochberg, 1995). Additionally, some of the control variables were highly correlated with each other, but the report did not discuss collinearity. In particular, items 11 (Rutter score group) and 24 (parent rating of hyperactivity) from the age 5 list (see chapter 3 appendix) would have been strong predictors of the main predictor variable, the ADHD indicator. Inclusion of highly related independent variables in the multivariate regression was likely to have made the regression coefficients behave in unpredictable ways (Dormann et al., 2013; Farrar & Glauber, 1967). Finally, the study was published in 2004; ADHD research has moved on considerably since then, and more recent data is now available on the BCS70 cohort members.

3.2.2 98-study systematic review

The most comprehensive of the other selected studies was a rigorous PRISMA (Moher et al., 2015) review of ADHD outcomes studies published internationally between 1980 and 2015 (Erskine et al., 2016). They screened 13,268 records, reviewed 278 in detail, included 98 in their quantitative meta-analysis, and evaluated 52 outcomes categorised into mental

[13] There was one less outcome for males: early pregnancy.

disorders, substance use disorders, suicide, sexual health, academic achievement, criminality, driving behaviours, employment, and service use (Erskine et al., 2016). For inclusion, the authors required that outcomes were measured a minimum of two years after ADHD was identified, and ADHD cases needed to be identified prior to age 18 and based on diagnosis or diagnostic criteria from two informants (Erskine et al., 2016). Pooled Odds Ratios were reported, and 37 of the 52 were found to be significant, indicating worse outcomes at the group level for ADHD vs. controls (Erskine et al., 2016). Particularly notable was the greater odds for ADHD to be associated with low educational attainment/no tertiary education (OR = 6.47, 95% CI = 4.58 - 9.14; Erskine et al., 2016).

Erskine et al. (2016) was strong in terms of the broad range of literature covered and the rigorous PRISMA-P approach. Most of the reviewed studies did not include outcomes from settled adulthood (over age 30), and there was substantial heterogeneity (I^2)[14] in many of the pooled findings.

There were two specific studies from the Erskine et al., (2016) review of particular interest, because they did measure outcomes after the age of 30. The two are discussed below individually, even though their results make up part of the pooled odds ratios reported by the review.

3.2.3 33-year follow-up

Klein et al., (2012) conducted an outcome follow-up 33 years after initial diagnosis, at a mean age of 41. Their sample (N = 135) was all boys with combined-type ADHD, diagnosed between 1970 and 1978 between the ages of 6 and 12, with no evidence of conduct disorder or other psychiatric problems, and an IQ of at least 85. The control group (n = 136) was matched based on ethnicity (all Caucasian), gender, age, parents' occupation class, and on the condition they had 'unremarkable' behaviour (Klein et al., 2012). As adults, the ADHD group had significantly worse educational attainment, lower job/social class, more unemployment, lower salaries, more divorces, more adult ADHD and antisocial personality disorder, nicotine addiction, incarceration, deaths, and psychiatric hospitalisations (Klein et al., 2012). They did not provide effect sizes, but p-values for the outcomes listed above were all < 0.01 (Klein et al., 2012). Some may not have been significant if adjustments had been made for multiple (11) tests[15], but no adjustment was noted. There were no significant group differences for social functioning, alcohol-related disorders, or mood and anxiety disorders (Klein et al., 2012).

---

[14] I^2 is a measure of consistency between studies included in a meta-analysis which is proposed to be less susceptible to bias from the number of studies included than other measures (Higgins et al., 2003)
[15] E.g. FWER .05/11 = .004

Whilst this study has the advantage of longevity and the validity associated with a clinical diagnosis, the sample is likely to suffer from some bias. ADHD and psychiatric diagnosis of any kind was much rarer in children at that time (1970-78) compared to today (Centers for Disease Control and Prevention, 2018b; Centre for Longitudinal Studies: UCL/IoE, 2019), so it is likely that these children's behaviour problems were severe. Also, they were all boys, combined type, and the matching criteria for the control group may not have captured truly similar children on the basis of gender, age, parent social class, and unremarkable behaviour, i.e. there may have been other unmeasured confounding variables affecting the outcomes reported.

3.2.4 Dunedin study

An age 38 follow-up (*N*~956) of the Dunedin cohort was also included in the Erskine et al., (2016) review, and discussed further here because of its unique relevance to my research objective. The Dunedin study follows 1,037 individuals born in Dunedin, New Zealand in 1972 and 1973. 61 children (6%) were identified with DSM-III ADHD based on data from sweeps at ages 11, 13, and 15. Thus this is an unusual prospective study of carefully assessed ADHD in an otherwise non-clinical sample with a follow-up over age 30. The group of 61 with ADHD in childhood was 78.7% male, 59.0% also had a conduct disorder diagnosis, and they had significantly lower IQ and worse cognitive performance than comparisons at the group level, matched on sex. At age 38 the childhood ADHD group had fewer university degrees, lower income, more debt, lower credit scores, a higher number of months on benefits, and more criminal convictions than the non-ADHD comparison group. Half of the childhood ADHD group still met the criteria for ADHD (3% of the sample) at age 38.

The cohort members in the Dunedin study are close in age to the BCS70 cohort, and also a non-clinical sample, so would appear to provide an excellent comparison point. However, the study was conducted in a single geographic area, and although recruited to be representative of that area, may not generalise more widely. Also, these children were actually diagnosed, and psychiatric diagnosis of children was relatively rare (like in the Klein et al., 2012 study) at the time they were assessed (Moffitt et al., 2015). Hence, they were likely to be severe cases, i.e. similar to the composition of a more recent clinical sample. This inference is supported by the composition of the group: 79% were boys, and 59% diagnosed with comorbid conduct disorder (Moffitt et al., 2015). The overall study size was well-powered but the ADHD group at n=61 was relatively small, and only sex was controlled for as a confound (Moffitt et al., 2015). So, there is a risk that statistically significant effects could be misleading due to lack of power (Button et al., 2013), and/or omitted variable bias. Finally, numerous hypotheses (outcomes) were tested, and adjustments for multiple corrections were not noted, although several of the findings were significant at $p < 0.001$ (Moffitt et al., 2015).

3.2.5 Multimodal Treatment study of children with ADHD (MTA)

The MTA (MTA, 1999) is one of the largest and longest prospective studies of ADHD, treatments and outcomes to date, with a baseline sample of N=579, age 7-9.9 years, 80% boys. There have been eight naturalistic follow-ups between two and 16 years after baseline (Hechtman et al., 2016; MTA, 1999; Roy et al., 2017). After the most recent, 16-year follow-up (Mean age = 24.7) functional outcomes from three points in adulthood were evaluated for 476 of the original ADHD sample, and 241 from the comparison group[16], who were classmates matched on age and sex (Hechtman et al., 2016). Outcomes evaluated included ADHD persistence, education, occupation, emotional state, legal contact, substance use, sexual behaviour and death. Adjustments were made for multiple tests. The specific outcomes that were significantly affected by ADHD were educational attainment (tertiary), number of times fired or quit a job, current income, receipt of public assistance, and risky sexual behaviour, and the effects were greatest for those with persistent ADHD (Hechtman et al., 2016). The authors also noted that although not statistically significant, there were 10 deaths in the ADHD group compared to one in the comparison group (Hechtman et al., 2016).

Strengths of the study were the large carefully diagnosed prospective sample, and thorough follow-ups. With a mean age of 24.7, the participants had not yet reached settled adulthood, which is a limitation only with regards to the broad research objective of this thesis. The sample was all combined-type ADHD, 80% boys, and 30% of the children had been treated with medication prior to the start of the study (MTA, 1999). Thus, this group contained commonly found clinical sample biases towards boys and more severe cases[17]. The comparison group matched on age and sex was recruited from the same school classrooms as participants, which may have controlled for some confounding, but specific cognitive and family characteristics were not specifically controlled for. It was noted that participants from families with greater socioeconomic disadvantage and more severe teacher-rated ADHD symptoms were less likely to participate in the longitudinal follow-ups (Hechtman et al., 2016), but there was no further discussion of missingness/attrition.

---

[16] They intentionally did not use the word 'control group' because the study did not meet with standards for a Randomised Controlled Trial. An RCT design was not possible due to ethical reasons, i.e. medication could not be given or withheld if the family did not agree to it (MTA, 1999).
[17] Combined-type cases are more severe, because 12 symptoms must be met instead of 6, which is the requirement for the other two types.

3.3 Summary

Overall the five studies reviewed showed a general trend for worse adult outcomes in ADHD vs. controls or comparison groups, though not for all outcomes. Many of the study samples had biases towards boys and more severe cases (combined-type ADHD), and pooled odds ratios reported in the systematic review contained considerable heterogeneity (per I^2 statistics).

All of the studies reported long lists of specific event-based outcomes, for example number of arrests, psychiatric diagnoses, traffic accidents, and early pregnancies. These have the advantage of being objective measures, or in the case of psychiatric diagnoses, not based on self-report. However, testing many outcomes is exploratory, and accompanied by an increased risk of false positive findings. Evaluating separate events fails to take into account how a person's life is going overall. For example, a person may have been diagnosed with anxiety at some point, had a traffic accident, or an early pregnancy, but they function relatively well. Also, some of these outcomes are far from unusual. In the Dunedin study, 49% of adults met the criteria for an anxiety problem at some point between age 18 and 32 (Moffitt et al., 2010). This event-based and predominantly negative outcome reporting approach motivated me to investigate more person-based and positive outcomes and examine a broader view of overall functioning.

3.4 Effect size comparison across studies

In search of indicators of overall functioning, I again reviewed the outcomes reported in the five studies discussed in the previous section and chose three that were included in all the studies. I identified depression/malaise, which should negatively correlate to the broader construct of subjective psychological wellbeing (Mensah & Hobcraft, 2008; Schoon & Kneale, 2013), income/class of job, and educational attainment. However, on further investigation, I found that income/class of job was inconsistently reported across the five studies under consideration to the extent that meaningful comparisons were not possible. Educational attainment was more consistently reported, directly relevant to the research context of this thesis (psychology and education) and should correlate to income/class of job. Consequently, I compared effect sizes across the five studies for ADHD on depression and educational attainment.

Statistics provided (Table 4) were converted to Odds Ratios (ORs) if not already provided, because these are the most appropriate effect size for dichotomous outcomes (Durlak, 2009). Qualitative interpretation of magnitude was based on thresholds for rare events (<10% probability) for depression (Small=1.22, Medium=1.86, Large = 3.00), and non-rare events for low education (Small=1.32, Medium=2.38, Large=4.70; Olivier et al., 2017).

	S1 (2004) Boys/Girls	S2 (E2016)	S3 (H2016)	S4 (K2012)	S5 (M2015)
Depression	1.32/1.19	2.31	1.43	1.55	1.20
Size±	S/NS	M	S	S	NS
Low Education	1.41/1.20	6.47	2.50	7.04	3.67
Size	S/NS	L	M	L	M

Table 4. Odds ratio effect sizes for comparable adult ADHD outcomes across five studies

Low education = no tertiary (no qualifications beyond high school in the US or beyond A-levels in the UK)
Study abbreviations: S1(2004)= (Brassett-Grundy & Butler, 2004), S2 (E2016)= (Erskine et al., 2016), S3 (H2016)= (Hechtman et al., 2016), S4 (K2012) = (Klein et al., 2012), S5 (M2015) = Moffitt et al., 2015
Where z and d were provided they were converted to OR using (Lenhard & Lenhard, 2016). For Klein et al., (2012), 2x2 tables were used to calculate ORs.
± Effect size interpretation: NS = not significant, S = small, M= medium, L = large

For depression, evidence from S1 (BCS70 ADHD study; Brassett-Grundy & Butler, 2004) indicated a small effect for boys (*OR* = 1.32), and not practically significant for girls (*OR* = 1.19), though close to the threshold for small. S2 (the 98-study review), indicated a medium effect (*OR* = 2.31), and the other three studies indicated small or non-significant effects.

For the low education outcome (no tertiary), S1 (BCS70) again indicated small/not significant effects of ADHD for boys/girls respectively. S2 (98-study review) and S4 (33-year follow-up, boys only) indicated large effects, whilst S3 (MTA) and S5 (Dunedin) indicated medium.

The systematic review overall included a good mix of clinical and non-clinical samples (Erskine et al., 2016), but based on my own review of the data[18], most of the studies with depression (9/13) and education (8/10) outcomes reported were based on clinical samples, and for education, more than half of the studies were boys only (6/10). Also the I^2 values indicated substantial heterogeneity (Higgins et al., 2003) in between-source study estimates (depression I^2 = 68.49%, education I^2 = 45.16%; Erskine et al., 2016).

Thus, my conclusion based on these data is that ADHD is likely to have a small effect on outcomes of depression/malaise[19] in settled adulthood. For tertiary education, ADHD may

[18] Spreadsheet summary of study data retrieved from supplementary information online (Erskine, 2016)
[19] …and these are expected to be related to wellbeing

have a medium-large effect for boys in clinical samples, and a small or non-significant effect when girls and less severe cases of ADHD are included.

3.5 Learning from strengths and weaknesses in existing literature to inform my study design

The review of other long-term outcomes studies revealed a number of study strengths that could be built upon, and limitations that could be addressed in my own study. My use of the BCS70 provided many of the strengths found in the Brassett-Grundy & Butler (2004) study, including the large non-clinical sample and rich array of behavioural and psychosocial data to control for confounding. There have been three further data sweeps for BCS70 since age 30, which allowed for outcomes to be evaluated later in life. The additional data collection points provided more chances to capture outcomes for a greater number of the original cohort members. There were also opportunities to refine some of the methods used in Brassett-Grundy & Butler (2004). First, the measurement of (categorical) ADHD could be more closely aligned with the current definition of ADHD per the DSM-5, a scaled score calculated, and subtypes derived. Second, it is possible to use more robust methods to account for missing data. Finally, controlled confounds could be linked more directly to literature, and, quasi-experimental methods used to improve balance on key covariates between ADHD and controls. All of the above were incorporated into my investigation of ADHD in BCS70.

Based on learning from all the studies reviewed, I chose to focus on wellbeing as an outcome. Wellbeing is framed in a positive way, provides a whole person, rather than a fragmented event-based view, and reduces the risk of false positive findings from multiple hypothesis testing.

4 Wellbeing the construct

Wellbeing is a widely studied and discussed construct, but there is not wide agreement on its definition. A full discussion of the history and all conceptualisations of wellbeing was not attempted here. I do frame why wellbeing is important, key aspects of the concept, and how it has been measured and studied in BCS70.

4.1 Background

Broadly, wellbeing is feeling good and functioning well (Huppert, 2009; What Works Centre for Wellbeing, 2019). The concept has become incredibly important in government policy and evaluation globally, especially in the last decade. In the UK, numerous government, NGO[20],

[20] Non-governmental organisation

and charity centres and programmes have been created to measure wellbeing and develop and promote policies to improve it (appg: wellbeing economics, 2019; Centre for Economic Performance, 2019; ESRC, 2019; Health Improvement Analytical Team, 2014; Office for National Statistics, 2019; What Works Wellbeing, 2019).

Why has wellbeing become such an important topic? A key reason is the positive relationship that has been found between wellbeing, health, and health behaviours (Health Improvement Analytical Team, 2014; Huppert, 2009). Better health and health behaviours lead to increased participation in work and reduced demand on health services (Health Improvement Analytical Team, 2014; Huppert, 2009). Wellbeing is also associated with a wide range of other positive outcomes, including education, employment, and relationships (Dolan et al., 2011; Helliwell et al., 2012). These are the same outcome areas that are often negatively affected in ADHD, so it is not difficult to argue that wellbeing should be an essential goal of managing ADHD.

4.2 How is wellbeing measured?

Measurement of wellbeing is driven by three main theoretical concepts:

- Hedonic: wellbeing is a state of happiness, positive affect, and by extension, absence of negative affect;
- Eudaimonic: wellbeing is about a life that is virtuous, contributes to society, and has meaning;
- Self-evaluation: if a person claims life satisfaction, i.e. that their life is going well, then they have wellbeing.

(for discussions, see Dolan, Layard, & Metcalfe, 2011; Ryan & Deci, 2001)

The What Works for Wellbeing organisation and the ONS have defined their measures based on academic recommendations linked to the three theoretical concepts listed above (Dolan et al., 2011; Helliwell et al., 2012), and by conducting surveys on what aspects of life are most important to people in the UK (What Works Centre for Wellbeing, 2019). The ONS reports on 41 different measures across 10 domains of life, including: "the natural environment, personal wellbeing, our relationships, health, what we do, where we live, personal finance, the economy, education and skills, and governance" (Office for National Statistics, 2019; What Works Centre for Wellbeing, 2019). The domains are classified as subjective (personal wellbeing) and objective (the other nine) (Health Improvement Analytical Team, 2014; Office for National Statistics, 2019).

Whilst this approach is certainly comprehensive, it is also complex and prone to ambiguity. In a recent review of 99 different measures of wellbeing, the authors observed that it is often not clear which subcomponents predict, define, or come about as a result of wellbeing (Linton et al., 2016). However one clear recommendation has been made that wellbeing should ideally be measured using both subjective and objective domains, because they do not necessarily tie to each other (Dolan et al., 2011).

4.3 Studies of wellbeing in BCS70

Based on a search of the Centre for Longitudinal Studies' research bibliography (University College London, 2019), I identified seven studies using BCS70 that evaluated wellbeing in adulthood. They tested effects for an array of wellbeing predictors, including childhood and/or adolescent socioeconomic status, social and emotional skills, experience of divorce, physical exercise, cognitive skills, emotional health, and behaviour (Clark et al., 2017; Layard et al., 2014; Mensah & Hobcraft, 2008; Sacker & Cable, 2006; Schoon & Kneale, 2013; Sigle-Rushton et al., 2005; Wood et al., 2017).

To measure subjective wellbeing, four of the seven studies used self-rated life satisfaction (Goodman et al., 2015; Layard et al., 2014; Schoon & Kneale, 2013), three used the Rutter Malaise shortened inventory score (Goodman et al., 2015; Sacker & Cable, 2006; Sigle-Rushton et al., 2005), and two used the Warwick Edinburgh Mental Health and Wellbeing Scale (WEMWBS) (Goodman et al., 2015; Wood et al., 2017). Five used a single measure, and two used a combination of two or three. Most of them also evaluated one or more measures of what the ONS has classified as objective wellbeing, e.g. indicators of employment, relationship status, and/or education, in addition to a subjective measure, in line with the Dolan et al. (2011) recommendation noted above.

To explore possible measures of wellbeing for use here, I created a mapping of the ONS wellbeing domains to available data items in BCS70 at age 42 (Table 5).

No.	ONS domain	BCS70 age 42 data
	Natural environment	NA*
	Personal/subjective wellbeing	WEMWBS, malaise inventory, life satisfaction
	Our relationships	Lives with a partner
	Our health	General health indicator, disability indicator, alcohol problems indicator
	What we do	Working indicator, social class of job
	Where we live	Satisfaction with accommodation
	Personal finance	Income for last 12 months (not usable, 80% missing)
	The economy	NA
	Education and skills	Highest level of qualification

Governance	NA

Table 5. Mapping of 10 ONS-defined areas of wellbeing and BCS70 data available at age 42

(Brown & Hancock, 2014; Office for National Statistics, 2019)

Based on the ONS recommendations and other studies of wellbeing in BCS70, I decided to develop a pilot study (see chapter 5) and test possible measures of wellbeing. The pilot, correlational analysis, evaluation of missingness, and further literature review led to the selection of two outcome measures that would provide the desired completeness, stability, and breadth:

1. Subjective wellbeing: a composite factor score based on self-rated life satisfaction, the Rutter Malaise Inventory, and the WEMWBS (validated subjective wellbeing scale)
2. Educational attainment as a proxy measure of socioeconomic status/position and indicator of objective wellbeing

Further explanation of the outcome measure selection process which took learning from the pilot study into account is discussed in chapter 6 (Chapter 62.2.2).

5 Operationalising state regulation theory in BCS70

Before I could fully articulate research questions, a further set of constructs needed to be defined: those used to operationalise State Regulation theory. Looking back to the diagram of my working hypothesis in section 2.3.2, there were two concepts of psychosocial stress proposed to have an impact on ADHD severity and outcomes. They were stressors and protective factors against stress. In this section I discuss literature on measures of stressors and protective factors, and how these were operationalised in the BCS70.

5.1 Stressors and protective factors

5.1.1 Stressful life events

Events that bring about major change and/or distress in a person's life have long been associated with the development of clinical psychiatric conditions, particularly depression (Brown, Harris, et al., 1973; Brown, Sklair, et al., 1973). These are usually measured using a self-assessment questionnaire or interview, and ideally include a subjective rating to indicate how intensely the event is perceived by the individual (Brown, Sklair, et al., 1973). Example life events include: death of a close friend or family member, moving house, changing schools, divorce, and more, and most lists of items used today can be sourced back to the Social Readjustment Rating Scale (Holmes & Rahe, 1967).

I searched the Centre for Longitudinal Studies bibliography site for papers using BCS70 with 'stress' in the title or abstract. The search returned 47 results, and I reviewed titles and abstracts. Only one study referred to stressful life events as a construct (Langton et al., 2011), and the rest either examined specific stressors, such as a bereavement or family structure changes, or distress in adulthood. This is not surprising, because although numerous childhood events were recorded in the BCS70 data which are commonly categorised as stressful life events, they were not captured or documented as part of a wider construct.

Next I searched beyond BCS70 using Web of Science (WoS; Clairivate Analytics, 2019). Search terms were: "cohort", "stressful life events", and "child" and filters were added to only search for articles and book chapters from the last 10 years (2009 – 2019). 136 results were returned. Titles were reviewed and 84 were excluded because stress was measured in adulthood, for a specific population, or a non-western country, leaving 52. Abstracts of the 52 were reviewed and 14 more were excluded due to small samples, specific population, or non-western country (some of these were missed in initial title review). I noted a large proportion of the studies that had measured stressful life events in children in British cohorts, and because the BCS70 is British, I decided to exclude studies conducted outside of Britain. This left 23 studies to review, and 10 of them used the Avon Longitudinal Study of Parents and Children (ALSPAC) data. ALSPAC is similar to BCS70 in terms of size, longevity, and breadth. Thus, for guidance on identifying stressful life events in childhood in BCS70, I focused on the one available BCS70 study and the 10 ALSPAC studies.

The BCS70 study examined the relationship between family income level and emotional problems in teenagers (Langton et al., 2011). They tested the relationship using data from four cohorts: The National Child Development Study (1958), BCS70, and two British Child and Adolescent Mental Health Surveys (1994, and 2004). A count variable of 10 stressful life events as reported by the parents was used as a control variable in their regressions; they noted the events included "parental separation, court appearances, bereavement, serious illnesses, and accidents" (Langton et al., 2011, p. 1083) and referenced (Goodyer et al., 1990), but did not report the full list of events. The referenced study (Goodyer et al., 1990) used a life events list and specialised interview technique to determine impact of the event, both developed by Brown & Harris, (1978). So presumably the Langton et al. (2011) study also used the Brown & Harris (1978) events. They found that having 3 or more stressful life events had a significant influence on the association between housing tenure (as an indicator of low income) and teenage emotional problems (Langton et al., 2011).

Six of the 10 ALSPAC studies found in my review included useful information about measuring stressful life events. The other four were focused on adolescent stress, parental stress, and biological processes. The key points from each of the six are summarised in Table 6.

Reference	Description	Source of life events
(Araya et al., 2008)	16 events on questionnaire completed by mother. Events occurred between ages 5 and 7. A count of events was used as the measure.	Brown & Harris (1978) and Barnett et al (1983), the latter of which included events specifically relevant to pregnancy
(Joinson et al., 2016)	Used 42-item list developed for ALSPAC, included in detail with the supplementary information	Brown & Harris (1978) and Barnett et al (1983)
(Enoch, Steer, Newman, Gibson, & Goldman, 2010	Used a list of 15-18 life events, full list not provided, focus on prenatal events experienced by mother	Brown & Harris (1978) and Barnett et al (1983)
(Slopen et al., 2013)	Five severe adverse life events: foster care, physically assaulted, sexual abuse, separated from mother, separated from father. Used count as measure.	Not reported
(Flouri et al., 2019)	Set of 43 events defined for ALSPAC, measured between ages 1 and 9 and 9 and 11. Detail provided.	Brown & Harris (1978), Barnet et al 1983, and Honor et al 1994
(MacKinnon et al., 2018)	Used 42-item list developed for ALSPAC	Brown & Harris (1978), Barnet et al 1983

Table 6. Review of stressful life events used in ALSPAC studies

From this collection of ALSPAC studies, associations were found between stressful life events and hyperactivity, conduct problems, emotional problems, enuresis, and inflammation (Araya et al., 2008; Enoch et al., 2010; Flouri et al., 2019; Joinson et al., 2016; MacKinnon et al., 2018; Slopen et al., 2013).

Nearly all of the studies reviewed referenced the Brown & Harris (1978) life events as a key source. However, the set of events used varied and was sometimes not fully reported (i.e. only examples given). Consequently, a validated approach was not found that could be used here for measuring non-specific childhood stress in BCS70. When I reviewed the most comprehensive set of items from the ALSPAC studies (43 items) against the BCS70 data, I found that several of them were not available in the BCS70. However, extensive data was collected on stressful life events in BCS70 that could be mapped to existing lists (e.g. Brown & Harris, 1978), so a stressful life events measure was created based on these items.

5.1.2 Chronic stressors

In the course of reviewing BCS70 and ALSPAC studies on stressful life events, I noted a repeated reference to a particular article on childhood stress, and traced the article. Compas, (1987) argued that whilst stressful life events have been correlated to later biopsychosocial problems in children and adolescents, there was perhaps more evidence to support effects of chronic stressors. Compas (1987, p. 277) gave examples of chronic stressors including ongoing hardship in the child's environment, disability or degenerative disease, or relationship conflicts, and also emphasised the importance of considering the protective effects of individual level social supports and coping skills.

These arguments from the Compas (1987) paper were subsequently developed further under the topic of resilience (e.g. Compas, Connor-Smith, Saltzman, Thomsen, & Wadsworth, 2001; Lavoie, Pereira, & Talwar, 2016; Oldehinkel, Ormel, Verhulst, & Nederhof, 2014; Rutter, 1987; Zimmer-Gembeck & Skinner, 2016). This list is a tiny selection of an extensive literature on the topic. A full review of resilience was not within the scope of my work here, because an at-risk status is a prerequisite to measurement and discussion of resilience (Masten, 2001). My objective was to measure/operationalise stressors and protective factors against stress in childhood within the available BCS70 data, without limiting the study to children who were considered at risk. However, to take into consideration the resilience literature-indicated influence of chronic stressors and protective factors, I re-examined the BCS70 childhood questionnaires for relevant items. I found numerous items relating to both. Operationalisation of protective factors is discussed briefly next, and measurement details for the related constructs are provided in chapter 6.

5.1.3 Protective factors

Individual, family, and community factors have been indicated as protective against stressors and/or risk of poor outcomes (Rutter, 1987). A search for widely-cited articles since 2010 that referred to the Compas (1987) and Rutter (1987) papers revealed a review of resilience literature with 600+ citations[21], which provided several candidate themes and/or constructs for operationalising protective factors, including: self-concept, pro-social engagement, optimism, parenting practices, and safe and positive recreation opportunities in the community (Zolkoski & Bullock, 2012). I reviewed the BCS70 age 10 data questionnaire data again with these themes and constructs in mind and identified three that could be used to

---

[21] Per Google Scholar, January 2020.

operationalise protective factors against stress. They included two indicators of self-concept: self-esteem and locus of control (Hattie, 1992; Judge & Hurst, 2007; Sherer et al., 1982), and engagement in leisure activities. Each of the three constructs is discussed next, and details about measurement are provided in chapter 6.

5.1.3.1 Locus of control

Locus of control is a psychosocial construct that ranges from (low) external to (high) internal. It is a measure of an individual's belief about whether events happen because of their own skill/effort/actions (internal), or because of luck, or complex forces outside their control (external) (Rotter, 1966). Locus of control was measured at age 10 in BCS70 using the CARALOC scale (Butler et al., 1997; Gammage, 1974). The CARALOC measure has been used in several other BCS70 studies and has been found to be associated with higher intelligence (von Stumm et al., 2009), as well as positive outcomes in adulthood, such as mental health, health behaviours educational attainment, and socioeconomic status (Feinstein, 2000; Goodman et al., 2015; Murasko, 2007; Percy & Iwaniec, 2008).

5.1.3.2 Self-esteem

Self-esteem is an individual's view of their own worth, or mental and physical characteristics; and ranges between negative and positive on a continuum (Cottle, 1965; Lawrence, 1981). Self-esteem was measured at age 10 in the BCS70 using the Lawrence Self-Esteem Questionnaire, or LAWSEQ (Butler et al., 1997; Lawrence, 1981). LAWSEQ has been used in other BCS70 studies, and the childhood measure has been found to correlate to behaviour problems, anxiety and depression in childhood (Ferro & Boyle, 2015; Prevoo & ter Weel, 2015a), and life satisfaction, wealth, and health behaviours in adulthood (Goodman et al., 2015; Joshi et al., 2016). Additionally, self-esteem has been found to have a positive association with educational attainment (Flouri, 2006; Hart, 1985).

5.1.3.3 Engagement in leisure activities

Leisure activities can be defined as those undertaken in 'free' time, i.e. they are not obligatory, like work or school (Newman et al., 2014). Within the maternal self-completion questionnaire of the BCS70 age 10 sweep, mothers (or parents) indicated how often (never, sometimes, often) their child participated in a list of 14 common childhood leisure activities in their spare time. Participation in leisure activity is associated with feelings of enjoyment, social support, and competence, which are protective against stress (Caldwell & Smith, 1988; Coleman & Iso-Ahola, 1993; Denovan & Macaskill, 2017; Iwasaki & Mannell, 2000; Iwasaki & Schneider, 2003; Newman et al., 2014).

6 Research questions

In the previous sections, wellbeing and educational attainment were selected as outcomes, and decisions taken to operationalise psychosocial stress and protective factors with measures of stressful life events, chronic stressors, self-esteem, locus of control, and engagement in leisure activity. With state regulation theory as a basis, a hypothesis about the relationships between psychosocial stress, ADHD severity and long-term outcomes, and a full set of key constructs, the following research questions were articulated:

RQ1: How can data science methods be used to retrospectively identify and validate robust categorical and continuous measures of DSM-5 ADHD in the BCS70?

RQ2: How do chronic stressors, life event stressors, locus of control, self-esteem, and engagement in leisure activity relate to ADHD and ADHD severity, all as measured at age 10? Does the relationship provide evidence to support state regulation theory?

RQ3: What is the effect of childhood ADHD on adult wellbeing using different methods and covariate sets to estimate treatment effects? Do the results support state regulation theory?

RQ4: What is the effect of childhood ADHD on adult educational attainment using different methods and covariate sets to estimate treatment effects? Do the results support state regulation theory?

7 Summary

This chapter covered selective reviews of literature on the key constructs evaluated in this thesis. The review discussed ADHD, long-term outcomes, wellbeing, stressors, and protective factors against stress. Based on these, four research questions were defined that are answered with the analyses presented in chapters 4, 7, and 8.

Chapter 4 Developing robust measures of ADHD in BCS70

Chapter 4 answers the first research question. It is divided into two parts: part one reports a pilot study to measure ADHD, and part two is the accepted version of a published paper that resulted from supervisor feedback and reflection on the pilot.

1 Part 1: Pilot study to measure ADHD and ADHD severity in BCS70

The purpose of this study was to test and validate a new method for identifying an ADHD subgroup in BCS70, using a mapping of BCS70 questionnaire items to current DSM-5 criteria. Here I justify using the DSM-5 criteria, latent trait theory, and statistical procedures to verify the subgroup. Next, a substantial explanation is provided on the statistical procedure that was used for validation, Item Response Theory (IRT). The remainder of the pilot study is reported in the format of method, results and discussion.

1.1 Introduction

1.1.1 Why is it not necessary to have a diagnosis?

ADHD has become increasingly medicalized over the last two decades, and is classified officially as a psychiatric disorder, so one might infer that a diagnosis is necessary to identify ADHD individuals. This approach was not possible with BCS70, since today's definition of ADHD did not exist then, and the related diagnosis that did exist, Hyperkinetic Disorder (HKD), was very rare. The Medical Examination Form was at least partially completed for 13,869 of the BCS70 cohort members at age 10, based on observations of medical professionals. Eleven of those children had an ICD-9 code (applicable at the time, in 1980) starting with '314', which was the diagnostic indicator for Hyperkinetic Syndrome of Childhood. This is 0.08%, which is a fraction of the currently reported rate of 1.6% in the UK, and dramatically less than international estimates that 5-7% of children meet the criteria for DSM ADHD symptomatology (Polanczyk et al., 2007; Willcutt, 2012).

Also, regardless of the limitations in BCS70, a professional diagnosis is not without its own biases. ADHD has historically been and is still today a disorder that often goes undiagnosed, and the diagnosis process itself may interact with trajectories and outcomes. As referenced in the Introduction chapter, there is wide variation in diagnosis rates by country, and even by region within country (Erskine et al., 2013; Hinshaw et al., 2011; Polanczyk et al., 2007). The estimated diagnosed prevalence ranges from around 1% of children in some countries (Polanczyk et al., 2007), to nearly 15%, in some parts of the United States (Centers for Disease

Control and Prevention, 2014). The variation in those ranges is influenced by attitudes and beliefs of parents, teachers, and physicians, and factors like the structure of the healthcare payments and education systems (Hinshaw et al., 2011). There is also concern in the medical research community about diagnoses being made inconsistently between practitioners within similar cultures and systems, and statistical methods have been explored to provide quantifiable validation (Lindhiem, 2013). Diagnostic criteria and practices have changed over time and even today they are applied inconsistently. Therefore, even if official diagnoses were available in the BCS70, they would not necessarily be an ideal method for identifying all or most of the children who met today's DSM-5 criteria for ADHD. Since there was no diagnosis in BCS70, there was almost certainly also no treatment with stimulant medication. Both the lack of medical diagnosis and treatment could have implications for the generalisability of the sample identified, which are discussed next.

1.1.1.1 Possible implications of no medical diagnosis

Diagnosis necessarily entails the assignment of a formal label by a relevant authority. In the case of ADHD, the authority is usually a medical doctor. It has been argued that a diagnosis/formal label is likely to lead to stigma by changing perceptions, expectations, and behaviours of those who know about it, such as doctors, parents, teachers, peers, and the child themselves (Mehan et al., 1986; Scheff, 1999; Shifrer, 2013).

Diagnosis may also lead to embarrassment or shame, as evidenced by the following quotes from ADHD children:

> *"I'd never want to tell a girlfriend I had ADHD" Adrian, 12*
>
> *"I just wish I didn't have it. I'd do anything not to have it. It ruins your life" Nick, 10*
>
> (ADDISS, 2006)
>
> *"No one knows [about ADHD diagnosis] except my teacher… I don't want anyone to know I have ADHD. They'll spread it all around school and then everyone will laugh at me." Brendan, 11*
>
> (Singh, 2012)

Changes in perceptions, expectations, and behaviours of the child and those around them and the experience of shame could affect developmental trajectories and outcomes. Therefore, the children in my BCS70 sample are not directly comparable to children in other samples who have a formal diagnosis.

1.1.1.2 Possible implications of no treatment with medication

Since the BCS70 cohort were not diagnosed with ADHD, it can be assumed they were also not treated with stimulant medication. Stimulants have physiological, psychological, and social effects above and beyond those of a formal label/diagnosis. The evidence is mixed, indicating both positive and negative effects. There is extensive literature on this topic, which I will not attempt to cover in depth here, but will provide a brief summary, to provide a context for limits on comparability between this sample and samples of children who are medicated.

Research over decades has consistently shown that stimulant medication reduces teacher-rated ADHD symptoms in the short-term with medium to large effect sizes, and side-effects are tolerable for approximately 80% of patients (Cortese et al., 2018; Storebo et al., 2012; Storebø et al., 2015). There is some evidence that medication improves academic performance in math (Kortekaas-Rijlaarsdam et al., 2018). There are reports that medication improves other areas of academic performance and social functioning in the longer-term, but it does not hold up under robust methodological scrutiny (Shaw et al., 2012; Storebo et al., 2012).

Although there is substantial evidence that serious cardiovascular events associated with stimulant treatment for ADHD are relatively rare (Greenhill et al., 1999; Habel et al., 2011), some side-effects are common, including difficulty sleeping, loss of appetite, nausea, vomiting, dry mouth, anxiety, tic disorder, increased sweating, and irritability (Greenhill et al., 1999; Novartis, 2019; Schachter et al., 2001; Storebø et al., 2015). These side effects are usually contextualised as 'relatively minor'. However, it's not unreasonable to expect that effects like these could contribute to other problems. For example, long-term follow-ups of the MTA sample (N = 436, for a description of the MTA study see Chapter 33.2.5) indicated that children who took stimulant medication between the ages of 7 and 9.9 were more likely to suffer with clinical anxiety or depression 6-8 years later (Molina et al., 2009).

The general and specific evidence from systematic reviews and the large MTA study suggests that medication is likely to have some influence on life outcomes, and that children who are not medicated (e.g. those in the BCS70 sample) may not be directly comparable to children who are medicated.

1.1.2 Previous study of ADHD outcomes in BCS70

ADHD prevalence and long-term outcomes in BCS70 have not been studied extensively, probably because of the changes in ADHD's definition and cultural attitudes towards it since the BCS70 were children. However, some work has been done. As mentioned in Chapter 3, a working paper was published on prevalence of ADHD in BCS70 and long-term outcomes observed at age 30 for the cohort members (Brassett-Grundy & Butler, 2004). The study identified an ADHD symptomatology sub-group using adapted applications of diagnostic cut-off scores on items from Conners Hyperactivity Scales (Conners, 1969) rated by both parents and teachers, and Rutter Behaviour Scales (Rutter, 1967) rated by parents at the age 10 sweep. Their method identified 1,101 cohort members with ADHD symptoms, out of a total of 14,797 they counted as participants in the age 10 sweep[22], or 7.44% (Brassett-Grundy & Butler, 2004). This percentage is slightly high compared to the reported worldwide prevalence of 5 – 7% (Polanczyk et al., 2007; Sayal et al., 2018; Willcutt, 2012), but close enough to provide some validation for their method at a macro level. The group was comprised of 689 boys and 412 girls, giving a ratio of 1.7 boys to 1 girl, which is also fairly close to epidemiology estimates for non-clinical samples (ranging from 1.9-3.2 : 1, see Willcutt, 2012). The subgroup size of 1,101 they identified was encouraging, because it appeared to be reasonable in proportion to the overall sample and was large enough to allow for use of robust statistical procedures.

I could have replicated the method used by Brassett-Grundy & Butler (2008) here. However, only a subset of items from the Conners and Rutter scales were included in the BCS70 questionnaires, so the validity and reliability of the scales is compromised. Also, in the intervening years, the scales have changed, and some of the items are significantly different from today's DSM-5 criteria. For example, a behaviour item used on the age 10 educational questionnaire from BCS70 was: "Is noticeably clumsy in formal or informal games" (University of Bristol & National Birthday Trust, 1980), which was adapted from the Conners teacher rating scale item "Coordination Poor" (Conners, 1969; Werry et al., 1975). This item, along with several others, bears no resemblance to any of the current DSM-5 ADHD symptoms. This is an example of "'fading relevance', reflecting earlier, and now obsolete, research concerns" (Bynner & Joshi, 2007, p. 174). Hence, using ratings of items like this, completely out of scope of the current DSM-5 definition, fails to take into consideration the extensive research, learning and refining of the definition of ADHD that has taken place in the intervening decades (American Psychiatric Association, 2016). Therefore, an alternative approach was taken here, using a direct mapping of BCS items to the DSM-5.

1.1.3 Measuring ADHD as a continuous latent construct

A latent construct (or trait) is not directly observable but can be measured indirectly through observed constructs. Examples of latent constructs are common in psychology and education research, including intelligence, mathematical ability, executive functions, personality, and locus of control. They are measured by proxy, often using more than one indirect

[22] Number of sweep participants reported varies slightly between sources

measurement, and often with a set of items on a questionnaire (Bollen, 2002). ADHD is a latent construct, because it cannot be observed directly (i.e. with a medical test), but it is identified using a list of related symptoms measured through observation, questionnaire, and/or interview.

The BCS70 childhood behaviour data collected cannot be treated as continuous[23]. It can be analysed as either ordered categorical (e.g. likert scale) or binary (only two possible values, true or false). The most appropriate statistical method for measuring the level of a latent trait with these types of data is Item Response Theory (IRT) (Baker, 2001; Bollen, 2002; Reeve & Fayers, 2005; van der Eijk, 2016b; van der Eijk & Rose, 2015). IRT is not currently widely reported outside of psychometrics research, so, as mentioned at the beginning of this chapter, an explanation of the method is provided here.

IRT is often used to develop banks of questions for standardized tests, for example to measure (latent) math or language ability. IRT is similar to factor analysis; it is a measurement procedure, as opposed to a path procedure, which is used to test relationships between independent and dependent variables (Field, 2009; van der Eijk, 2016b). There are two types of factor analysis: exploratory and confirmatory. Exploratory factor analysis is used to identify an unknown number of dimensions (or latent constructs) present in a set of questionnaire items, using correlation procedures. Confirmatory factor analysis instead assumes a fixed number of latent constructs, and the procedure tests the assumption (Field, 2009; Stata Corp LP, 2013). IRT is more like confirmatory factor analysis, but it is usually implemented as unidimensional, i.e. assumes a single underlying latent trait or factor.

Classical Test Theory (CTT) is useful for understanding the context of IRT and may be more familiar to the reader. CTT usually assumes all items on a test have an equal (i.e. un-weighted) status, and scoring is done with averages or simple sums. Reliability and validity in CTT are calculated for an entire test instrument (Reeve & Fayers, 2005), and an example of a CTT instrument is the Strengths and Difficulties Questionnaire (SDQ; youthinmind, 2012). Cronbach's alpha (α) is the measure most often used to calibrate, and strictly speaking is only valid in the sample or population used to validate the test (Reeve & Fayers, 2005), though it is often interpreted to be generalisable. IRT, in contrast, assumes that all items can have different levels of importance, and thus weight them in terms of difficulty and ability to discriminate between individuals with low and high levels of a latent trait. A unique function is derived for each item in an IRT model, called an Item Characteristic Curve (ICC; Baker, 2001; Embretson & Reise, 2000; Reeve & Fayers, 2005). A contrasting assumption to CTT is that the

---

[23] Explained further in this chapter, section 2.4.2.

ICC can be expected to generalise to other samples, if the sample used to build the model is large enough. The IRT procedure uses these ICC functions to calculate a score for each individual (denoted theta, or θ) which has a corresponding estimate of the probability (P(θ)) that a person with θ level of a latent trait would select a specific item response pattern (Reeve & Fayers, 2005). To illustrate, DSM-5 diagnostic criteria are met for ADHD if six out of nine DSM-5 hyperactivity criteria are present, but response patterns can and do vary (Table 7).

	DH1	DH2	DH3	DH4	DH5	DH6	DH7	DH8	DH9
Child 1	1	1	1	1	1	1	0	0	0
Child 2	0	1	1	1	1	1	1	0	0
Child 3	0	0	1	1	1	1	1	1	0

Table 7. Examples of different possible response patterns that could meet DSM-5 diagnostic criteria

The level of a latent trait measured using IRT could, in theory, range from negative to positive infinity, but for practical purposes IRT assumes a midpoint of 0, standard deviation of 1, and a range of theta (θ) from roughly -4 to +4 (Baker, 2001, p. 6); in this way it is similar to a standardized z-score. IRT uses the response patterns (see Table 7 above) to construct the ICC curve and predict theta. This is a function of the probability of a response choice (0 or 1) and levels of the latent trait, or values of theta (roughly -4 to +4). A typical ICC would show the probability of an affirmative answer being near 0 for the lowest levels of ADHD, and near 1 for high levels of ADHD (Baker, 2001). The ICC curve is central to the method, as are the associated discrimination and difficulty parameters. To illustrate these concepts, two example ICC curves (from the present study) are shown in Figure 5 below. They are the two items in my data that discriminate the most (di8[24] - easily distracted) and least (dh6 – talks excessively) between individuals high and low on my latent ADHD trait scale. The value of the slope at its steepest point is the discrimination parameter. The steeper the slope, the more that small changes in an individual's latent ability affect the latent trait score (Baker, 2001). Note the strong 's' shape for di8, which is the more discriminating item. If the slope were 0 (a horizontal line), the item would provide no information about the underlying latent trait (Reeve & Fayers, 2005, p. 58). The point on the function where the slope is steepest is marked on each graph; dh8's is at a theta value of 0.205, and dh6's at 2.97. This is the point for each item on the theta scale where values of 0 and 1 are equally likely (i.e. 50% probability). The

[24] The variables (e.g. di8) are named in the convention "*dtn*". They all begin with "d" for "DSM-5"; t = type (h for hyperactive, i for inattentive), n = number (corresponds to criteria item numbers shown - see Table 11).

higher this value, the more difficult the item (Baker, 2001; Reeve & Fayers, 2005) Thus, dh6 is more difficult than di8.

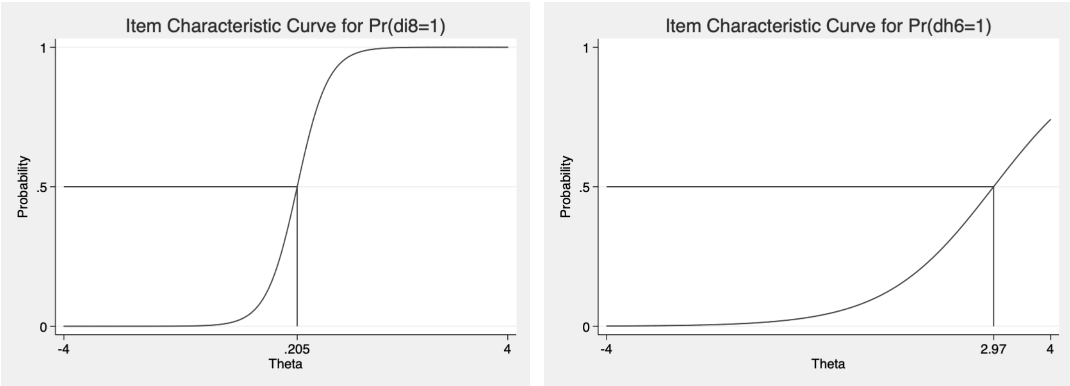


Figure 5.   Examples of Item Characteristic Curves

In IRT, item and test information functions (IIF and TIF) are used to calculate reliability. Reliability can vary over the ability scale; i.e., an item or test can be most reliable at low, medium, or high levels of ability (Embretson & Reise, 2000; Reeve & Fayers, 2005). In the present research, ideally, I will find a group of items that is highly reliable for individuals at the higher end of the ADHD symptomatology latent trait scale.

This introductory section has discussed justification of mapping DSM-5 criteria to BCS70 data items and use of Item Response Theory to score ADHD as a continuous trait. This leads to some basic research questions for a pilot study, which were used as a learning process to develop a final method for measuring ADHD in BCS70.

PRQ1[±]: Is mapping DSM-5 items to BCS70 data and applying the criteria to identify an ADHD subgroup a viable approach?

PRQ1a: How does the identified sample size compare to external sources of ADHD prevalence estimates?

PRQ1b: How many of the identified ADHD subgroup responded to the BCS70 age 42 sweep (for evaluation of long-term outcomes)?

PRQ2: Can an Item Response Theory (IRT) model with a good fit be built to score cohort members on a continuous latent construct of ADHD?

[±] *PRQ = pilot research question*

1.2 Method

1.2.1 Participants

Participants were cohort members in 1970 British Cohort Study 1970. A full description of BCS70 and ethics statement was provided in chapter 2.

1.2.2 Selection of sweep

As described in chapter 2, there have been ten BCS70 data collection sweeps, at cohort member ages 0, 5, 10, 16, 26, 30, 34, 38, 42, and 46. So which sweep (or sweeps) should be used to identify childhood ADHD? Sweeps during childhood were 0, 5, 10, and 16. ADHD is diagnosed on average at age 7 (Centers for Disease Control and Prevention, 2014), so sweeps at 0 and 5 are not viable. ADHD symptoms often decline in adolescence, and there is a reduced likelihood that a child who once met diagnostic criteria still will by age 16 (Faraone et al., 2005). Also, the BCS70 age 16 dataset suffers from extensive missing teacher/school data, because there was a teachers' strike that year (1986; Centre for Longitudinal Studies: UCL/IoE, 2019). Finally, other studies which make comparisons between BCS70 behavioural data in childhood and long term outcomes tend to use the age 10 sweep, because one of its strengths is the behavioural data (Brassett-Grundy & Butler, 2004; Butler et al., 1997; Goodman et al., 2015; von Stumm et al., 2009). Although I did some preliminary data analysis and considered combining age 5 and/or 16 data with the age 10 data, I concluded for the above reasons that it would have a detrimental effect on the integrity and size of the sample. The age 10 maternal and educational questionnaires contained the most relevant data and using both of these allowed me to accommodate the DSM-5 requirement that symptoms should occur in two settings (i.e. home and school). Thus, data from both mothers and teachers from the age 10 sweep was used to identify an ADHD symptomatology subsample in the present study (N=14,875).

1.2.3 Materials

BCS data was downloaded from the UK Data Service, registered to this author under usage number 93379, titled 'Pilot of PhD analysis on ADHD', expiring 24/9/2017. BCS data was available from this service at no cost for non-commercial use and accessed through Shibboleth using University of Cambridge Raven login credentials. Detailed user guides, data dictionaries, and Stata datasets for each sweep were included in the downloaded data. Also, copies of scanned, annotated questionnaires administered at ages 5, 10 and 16 were downloaded from the Centre for Longitudinal Studies website (Centre for Longitudinal Studies: UCL/IoE, 2019).

The pilot study analysis was conducted using Stata ME version 14.2 for Mac.

1.2.4 Identification of ADHD subgroup

1.2.4.1 Mapping DSM-5 criteria to BCS70 data

All potentially relevant data items from the age 10 maternal and educational questionnaires were recorded in a spreadsheet and mapped to DSM-5 ADHD criteria based on semantic similarity. The wording of items on the BCS70 questionnaires was not exactly the same as the DSM–5 criteria, so judgment of face validity (Howitt & Cramer, 2008) was used to link items from one list to the other. It is possible that parents and/or teachers presented with items exactly as they are worded in the DSM-5 may have answered differently. The mapping was done in three iterations and reviewed by my supervisor to help ensure a robust interpretation. For several symptoms, more than one item was identified as a map to the BCS70 data. In these cases, DSM-5 criteria were considered met if any of the mapped items were answered with an 'often' answer, by either rater, mother or teacher. Using an affirmative answer from either rater follows the 'or rule' used in DSM field trials (Lahey et al., 1994; Willcutt, 2012). I mapped a total of 37 items from the BCS age 10 sweep to DSM-5 criteria; 13 from the maternal questionnaire and 24 from the educational (teacher) questionnaire. I subsequently dropped two of the teacher items because in a review of descriptive statistics I found they produced an 'often' answer for nearly half of the children, i.e. they were not appropriate indicators of a rare construct like disorder-level ADHD.

Some of the DSM-5 items could not be mapped at all, so the diagnostic rules could not be applied exactly as they were intended. There are nine criteria in each of the two DSM-5 symptoms lists; inattentive and hyperactive. The DSM-5 specifies that six or more of either or both lists of nine should be met (behaviour described is 'often' observed), to meet diagnostic criteria. I was able to map BCS70 items to all nine of the hyperactive DSM-5 items, but only five of the inattentive items. Since the DSM-5 specifies that six out of nine (or two-thirds) of the criteria should be met for diagnosis, I assumed a cut-off of two-thirds for both lists; six out of nine for the hyperactive list, and four out of five (3.33 rounded up for conservative estimates), for the reduced list of inattentive symptoms.

1.2.4.2 Coding level of response

The BCS items on the original questionnaires were presented to parents and teachers using a visual analog scale (VAS). The scale consisted of a horizontal line, and the rater was asked to make a vertical mark on the line to indicate the degree to which the statement applied. An example item is shown in Figure 6:

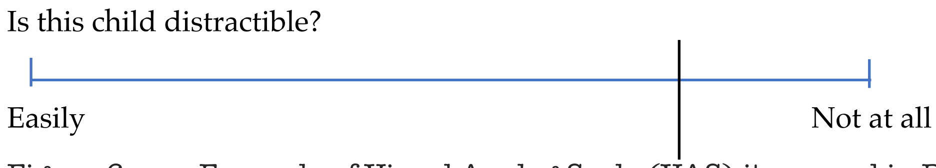


Figure 6. Example of Visual Analog Scale (VAS) item used in BCS70 age 10 sweep

The responses were systematically (i.e., by a coder using a ruler) coded either between 0 and 100 (maternal questionnaire) or between 1 and 47 (educational questionnaire)[25]. The resultant data values look like continuous data, and thus make it tempting to use the associated and simpler parametric statistical procedures. However, VAS data does not have the mathematical properties of continuous data, such as equal distance between points, which would allow operations like addition, or calculating averages/means, to have meaningful results (Svensson, 2001). Also, simulation studies have shown that a 100-point VAS can only reliably identify at most nine or ten distinct categories, and often it is more like three or four (van der Eijk, 2016a; Wewers & Lowe, 1990).

As discussed, most of the items were originally adapted from the Rutter (1967) or Conners (1969) scales, which used three and four levels of response, respectively. The Brassett-Grundy & Butler (2004) study recoded the VAS data to three levels, using the lowest common denominator between these two scales. However, the DSM-5 criteria are worded in a binary way; a child either has a symptom often or does not. Also, other studies that have confirmed or screened for ADHD or similar diagnoses (Lindhiem, 2013; Lindhiem et al., 2015), chose to recode their data to binary, based on the two levels implied by the DSM-5. Therefore, even though nuances of the variability in the data were lost, the data in the present study was recoded as binary.

In order to map to binary, I evaluated the three levels used in the Rutter scales. They were:

- Does not apply
- Applies sometimes
- Certainly applies

(Centre for Longitudinal Studies: UCL/IoE, n.d.; Rutter, 1967).

'Certainly applies' maps best to 'often', the wording used in the DSM-5 criteria. Based on this, and the approach used in (Brassett-Grundy & Butler, 2004) to divide the VAS scale into equal thirds, the top one-third of the VAS scale was equated to an answer of 'yes, often', and mapped to a value of 1, and the bottom two-thirds of the VAS was mapped to 0 ('no, not

[25] I checked this with the team at the Centre for Longitudinal Studies (CLS), who support the BCS70 data, because I was unsure if it might be an error. They confirmed that indeed two scales, 0-100 and 1-47, were used.

often'). Missing data were mapped as missing. Any relevant items worded to indicate the opposite of an ADHD symptom (e.g. child concentrates well) were reverse coded, i.e. the lower one-third of the VAS was mapped to 1, and top two-thirds to 0.

A few mapped questionnaire items did not use the VAS scale, and these were coded differently. A section on the educational questionnaire asked the teacher "When the child is expected to be working, roughly what percentage of the time (i.e. within the period) would you describe the child's behaviour as: Concentrating on the task at hand … " etc. (University of Bristol & National Birthday Trust, 1980, p. 6, question A26). The answer was entered as a percentage, and there were seven items meant to add up to 100%. Three of these items contained data that mapped to DSM-5:

- "Talking to other children,
- Moving around the classroom, and
- Fidgeting and indulging in other minor distracting activities"

(University of Bristol & National Birthday Trust, 1980, p. 6).

A precedent could not be found for mapping this type of data to a binary value. Without a tested model to follow, I took a conservative approach. I created a box plot for each variable (j080, j081, and j082), and coded all the values shown as outliers to 1 (behaviour displayed often), inferring that outliers represented unusual behaviour compared to the rest of the cohort, and all other values to 0 (Figure 7, Figure 8, Figure 9).

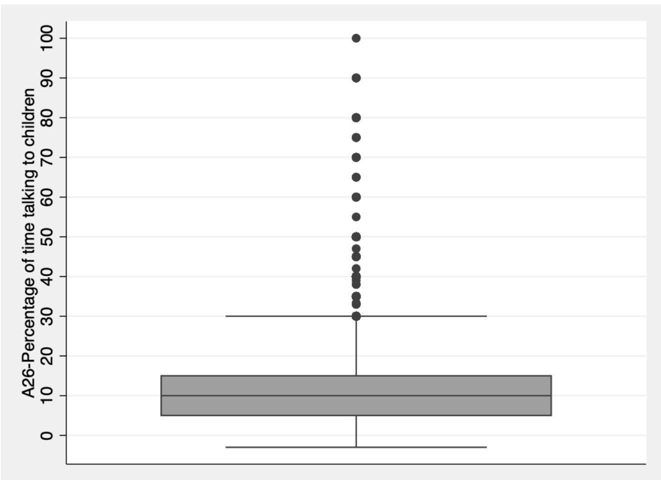


Figure 7. j080 – Percentage of time talking to other children (outliers >=30)

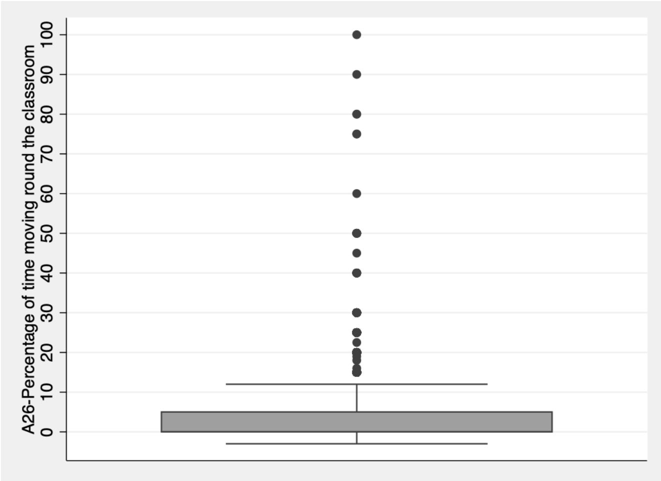


Figure 8. j081 - Percentage of time moving around the classroom (outliers >=12)

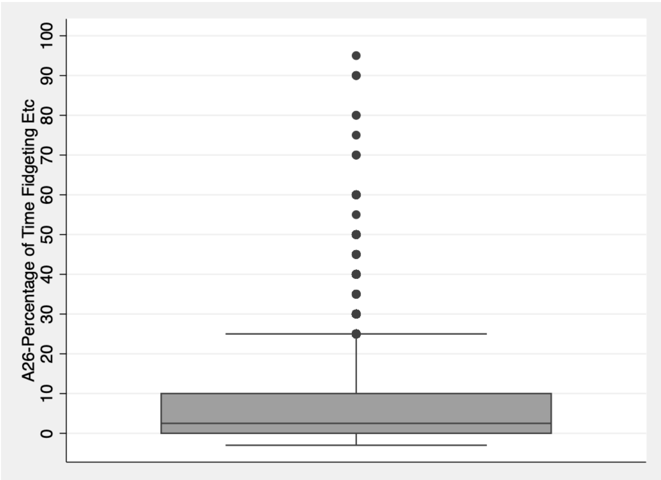


Figure 9. j082 – Percentage of time fidgeting (outliers >=26)

Once all the BCS items were recoded into binary data (either 0, 1, or missing), I mapped them to new variables for each of the 14 DSM-5 criteria. As mentioned previously, some DSM-5 criteria mapped to multiple BCS items, some only one, and some none. So, if a DSM-5 criterion mapped to multiple BCS items, I updated the DSM-5 variable to 1 if any of the relevant BCS variables had a value of 1. If all mapped BCS items were 0, the DSM-5 item was coded to 0, and the remaining (all missing) were coded missing.

Once mapped, the items were summed to a simple score: total ADHD symptom count = hyperactive count + inattentive count. The score had the distribution characteristics shown in Table 8 and Figure 10:

Range	0-14
Mean	2.97

Standard Deviation	3.07
Skewness	0.96
Kurtosis	3.03

Table 8. Descriptive statistics of simple ADHD symptom count

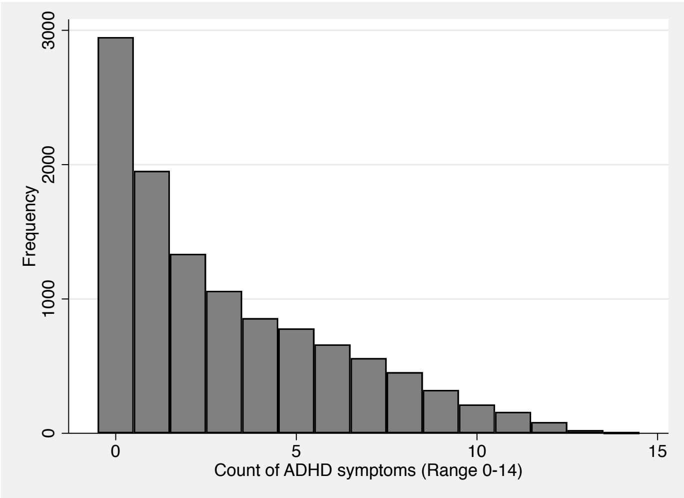


Figure 10. Histogram of simple ADHD symptom count

The shape of the histogram in Figure 10, with most of the values close to zero, and tailing off dramatically as scores increase, is what one would expect of a disorder like ADHD, which is rare in the population (Finkelman et al., 2011; Wall et al., 2015).

1.2.4.3 Contextual criteria to identify ADHD subgroup

Once the simple score described above was calculated, the DSM-5 contextual criteria were evaluated (Table 9).

No.	DSM contextual criteria*	Marked as 'met' (1) when
1	Several inattentive or hyperactive-impulsive symptoms were present before age 12 years	Always – all criteria were evaluated at age 10, so all symptoms were present before age 12
2	Several symptoms are present in two or more settings, (such as at home, school or work; with friends or relatives; in other activities)	If both mother and teacher indicated one or more symptoms were present
3	There is clear evidence that the symptoms interfere with, or reduce the quality of, social, school, or work functioning	Child was indicated in the 'moderate' or 'severe' behavior problems group based on their (mother) Rutter items score (derived BCS item)

4	The symptoms are not better explained by another mental disorder (such as a mood disorder, anxiety disorder, dissociative disorder, or a personality disorder)	Cohort members were excluded if they had been diagnosed with another psychiatric disorder, as per the medical questionnaire, identified using ICD codes[26]
5	The symptoms do not happen only during the course of schizophrenia or another psychotic disorder	See item 4
6	Symptoms present for at least six months	Could not be mapped

Table 9. Method for evaluating DSM-5 contextual criteria

* *contextual criteria from:* (American Psychiatric Association, 2013)

Finally, individuals were marked with an ADHD subgroup indicator if:

- Six out of nine hyperactivity symptoms met AND/OR four out of five inattention symptoms met
- AND mother-rated Rutter behaviour score indicated moderate or severe behaviour problems
- AND both mother and teacher indicated one or more symptoms
- AND no other DSM-5-specified psychiatric diagnoses were reported

The composition of symptoms (hyperactive, inattentive, or both) was used to create an ADHD subtype category. Descriptive statistics of the ADHD subgroup are provided in Table 10.

Number of children in subgroup	889
% of total N = 14,875	5.97%
Boys	598 (67.27%)
Girls	291 (33.73%)
Ratio of boys to girls	2.05 : 1
Combined subtype	282 (31.72%)
Hyperactive subtype	134 (15.07%)
Inattentive subtype	473 (53.21%)

Table 10. Characteristics of ADHD subgroup

All percentages were compatible with recent epidemiology estimates of ADHD in population-based samples (Willcutt, 2012), supporting some degree of construct validity. The subgroup of 889 was about 20% smaller than the 1,101 found in the other studies that used Rutter and

[26] Only two children had a relevant psychiatric diagnosis per ICD-9 codes

Conners score cut-offs to identify ADHD in BCS70 (Brassett-Grundy & Butler, 2004; Brassett-Harknett & Butler, 2007), indicating the new method was more conservative. This could be due to the items in their scales that are no longer associated with DSM-5 ADHD, and my additional use of 'necessary' contextual conditions.

1.2.5 Estimation of a continuous ADHD measure

As discussed in the introductory paragraphs of this chapter (4), IRT was selected as the most appropriate method for calculating a latent trait score, given the construct represented by the items can be viewed as unidimensional (Willcutt et al., 2012) and the data are categorical. As a first step, the data were tested to ensure the assumptions underlying IRT models were met.

1.2.5.1 Tests of assumptions

Toland (2013) summarised the assumptions for IRT procedures. These include:

- Adequate response per level
- Unidimensionality
- Item independence
- Model fit
- Normal distribution of the latent trait

To check for adequate response per level, the 37 BCS70 items identified as candidates to map to DSM-V criteria were checked for percentage of responses coded 0, 1, and missing. Two of these (j063-talkative with friends and j064-talkative with teacher) had responses of 1 for about 50% of the sweep, and since disorder-level ADHD is relatively rare, this was deemed excessive and these two items were removed.

The set of items should all relate to a single underlying dimension, i.e. be unidimensional. A factor analysis using tetrachoric correlations (most appropriate type for binary data; van der Eijk, 2016a), showed that correlations between a single assumed factor and each of the 14 variables were adequate, ranging from .44 to .88. Also, a Kaiser-Meyer-Olkin measure of sampling adequacy statistic showed the variables were all well related, with a value of .90. Cronbach's alpha was .83, which indicates a strong relationship, though not necessarily unidimensionality.

The assumptions of independence, model fit, and normal distribution of the latent trait were not tested for the pilot study but were addressed in the final study in the second part of chapter 4.

1.2.5.2 Selection of appropriate IRT model

Three types of IRT models were described in the literature reviewed: one parameter logistic (1PL), 2PL, and 3PL. 1PL IRT models assume discrimination is the same for all items, 2PL allows discrimination to vary, and 3PL estimates a guessing parameter (Baker, 2001; Reeve & Fayers, 2005). Since some DSM-5 items are likely to discriminate ADHD better than others, the 2PL IRT model is more appropriate than 1PL. 3PL is not necessary, because the 3rd parameter models the probability that a person would 'guess' the correct answer. There is no concept of 'correct' on these items; they indicate a child is more or less like a description on an ADHD symptom. It is unlikely a rater would be motivated to try and guess to select high scoring ADHD answers, so the 3rd parameter is not useful. Thus, the 2PL model is the best fit, and most parsimonious.

The 2-parameter logistic (2PL) model is based on the following equation:

$$\Pr(X = 1) = \frac{e^{a(\theta - b)}}{1 + e^{a(\theta - b)}}$$

Where: e = 2.718 (a constant), a = discrimination, b = difficulty), and theta is 'ability' level, or level of the latent trait (Baker, 2001, p.22).

This equation calculates the probability of responding correctly (positively, in the case of the ADHD scale) at a given ability (i.e. ADHD) level.

1.2.5.3 Results of IRT model

The 2PL IRT model on the 14 DSM symptom items (N=14,752) produced the results in Table 11. Observations were deleted only if all items were missing, because Stata IRT by default uses Full Information Maximum Likelihood (FIML) to handle missing data (Stata Corp LP, 2015; Yang & Zheng, 2018). Items in Table 11 were sorted in ascending order of discrimination, i.e. the least discriminating item, dh6, is listed first, and the most discriminating item, di8, is listed last. Di6 was the least difficult, and dh2 was the most difficult.

DSM-5 variable	**Disc(a)**	**z**	**p**	**Diff(b)**	**z**	**p**
dh6 - talks excessively	1.02	21.28	0.00	2.97	25.95	0.00
dh3 - climbs, restless	1.25	36.32	0.00	1.09	38.86	0.00
dh7 - blurts answers	1.37	35.57	0.00	1.34	43.10	0.00
dh2 - leaves seat	1.40	17.72	0.00	3.19	24.56	0.00

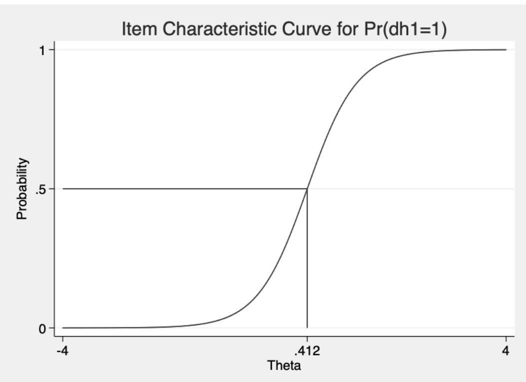
dh8 - trouble waiting turn	1.54	37.97	0.00	0.88	39.18	0.00
di9 - forgetful	1.63	37.18	0.00	1.07	44.31	0.00
dh4 – can't play quietly	1.77	34.52	0.00	1.52	50.31	0.00
di2 – can't pay attention	1.84	39.89	0.00	0.50	30.25	0.00
di6 - avoids complex tasks	1.91	40.91	0.00	0.14	9.23	0.00
dh9 - interrupts	1.93	36.36	0.00	1.16	49.14	0.00
dh1 - fidgets	2.23	40.51	0.00	0.41	27.78	0.00
dh5 - on the go/like motor	2.35	37.40	0.00	0.90	49.41	0.00
di4 – can't follow through	2.51	40.24	0.00	0.19	14.05	0.00
di8 - easily distracted	3.60	34.83	0.00	0.20	17.00	0.00

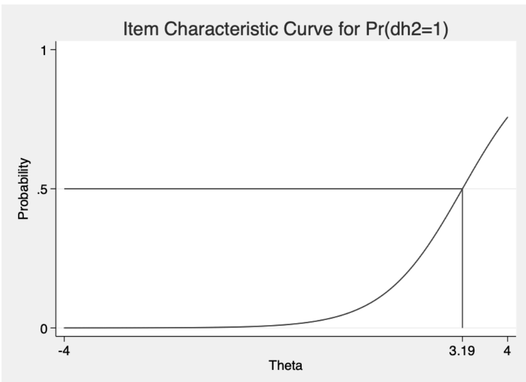
Table 11. 2PL IRT model results by item, ordered by discrimination

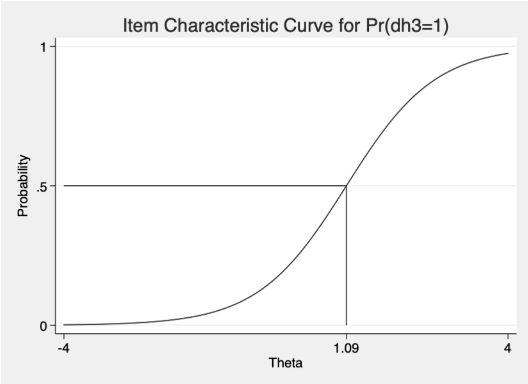
All p-values were (<.001), providing some assurance that the coefficients were significantly different from 0. A suggested interpretation of discrimination is that values < 0.65 are low, between 0.65 and 1.34 are moderate, and >1.34 are high (Baker, 2001). The discrimination parameter values here range from 1.02 – 3.60, so moderate to very high. Difficulty of the items ranges from 0.14 to 3.18, which means the probability of a 'yes' ADHD answer always hits 50% at theta values greater than zero (Baker, 2001). This is favourable for my purposes here, because the items function best for individuals who are higher on the latent ADHD trait (Baker, 2001).

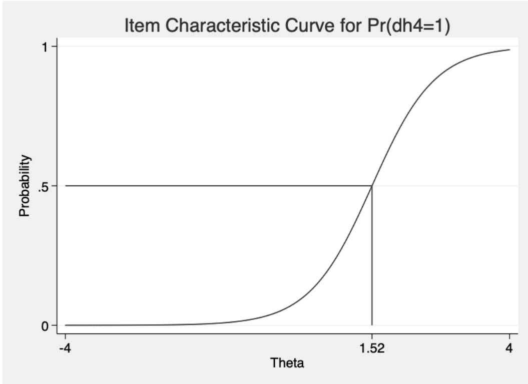
The Item Characteristic Curves (ICCs) for the mapped DSM-5 items are shown in Figure 11. Most of the curves were a strong 'S' shape, indicating good model fit, i.e. the items were useful in measuring an underlying trait (Toland, 2013). Dh2 (leaves seat) and dh6 (talks excessively) had the flattest curves and were thus the least informative items. However, since they are part of DSM-5, and the IRT model weights them accordingly, they were kept in the scale.

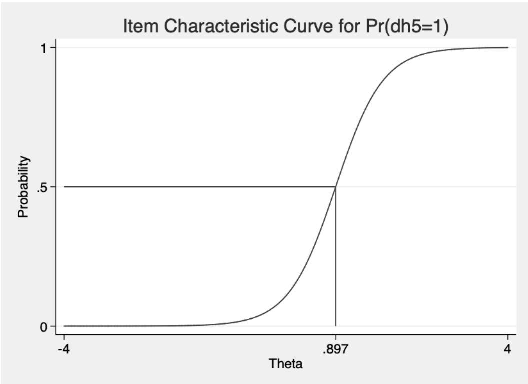
Hyperactive items: 1-9

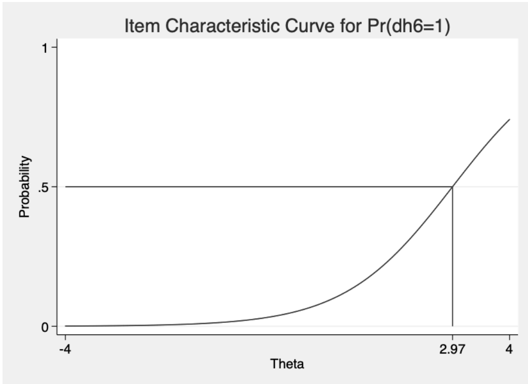


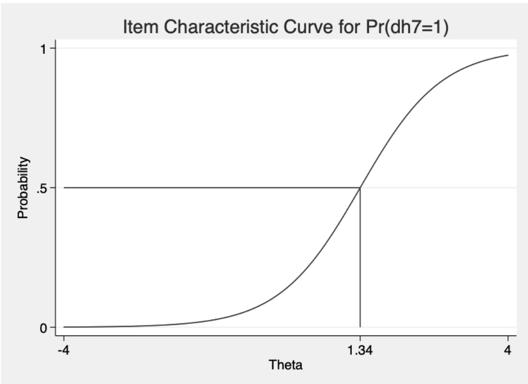


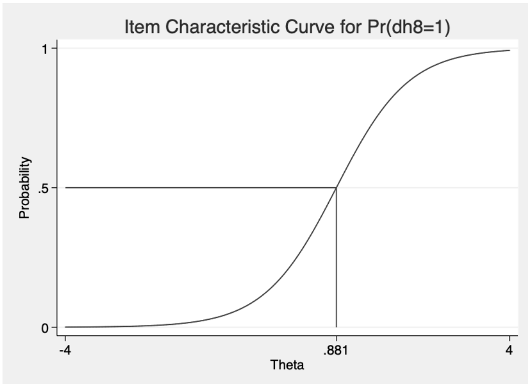

Item Characteristic Curve for Pr(dh3=1)
Probability
0
.5
1
-4
1.09
4
Theta

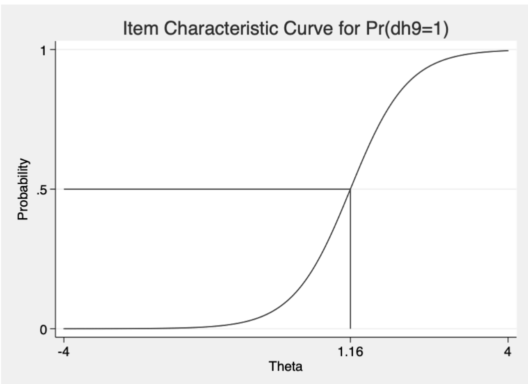

Item Characteristic Curve for Pr(dh4=1)
Probability
0
.5
1
-4
1.52
4
Theta


Item Characteristic Curve for Pr(dh5=1)
Probability
0
.5
1
-4
.897
4
Theta


Item Characteristic Curve for Pr(dh6=1)
Probability
0
.5
1
-4
2.97
4
Theta


Item Characteristic Curve for Pr(dh7=1)
Probability
0
.5
1
-4
1.34
4
Theta


Item Characteristic Curve for Pr(dh8=1)
Probability
0
.5
1
-4
.881
4
Theta


Item Characteristic Curve for Pr(dh9=1)
Probability
0
.5
1
-4
1.16
4
Theta

Inattentive items: 2, 4, 6, 8 & 9

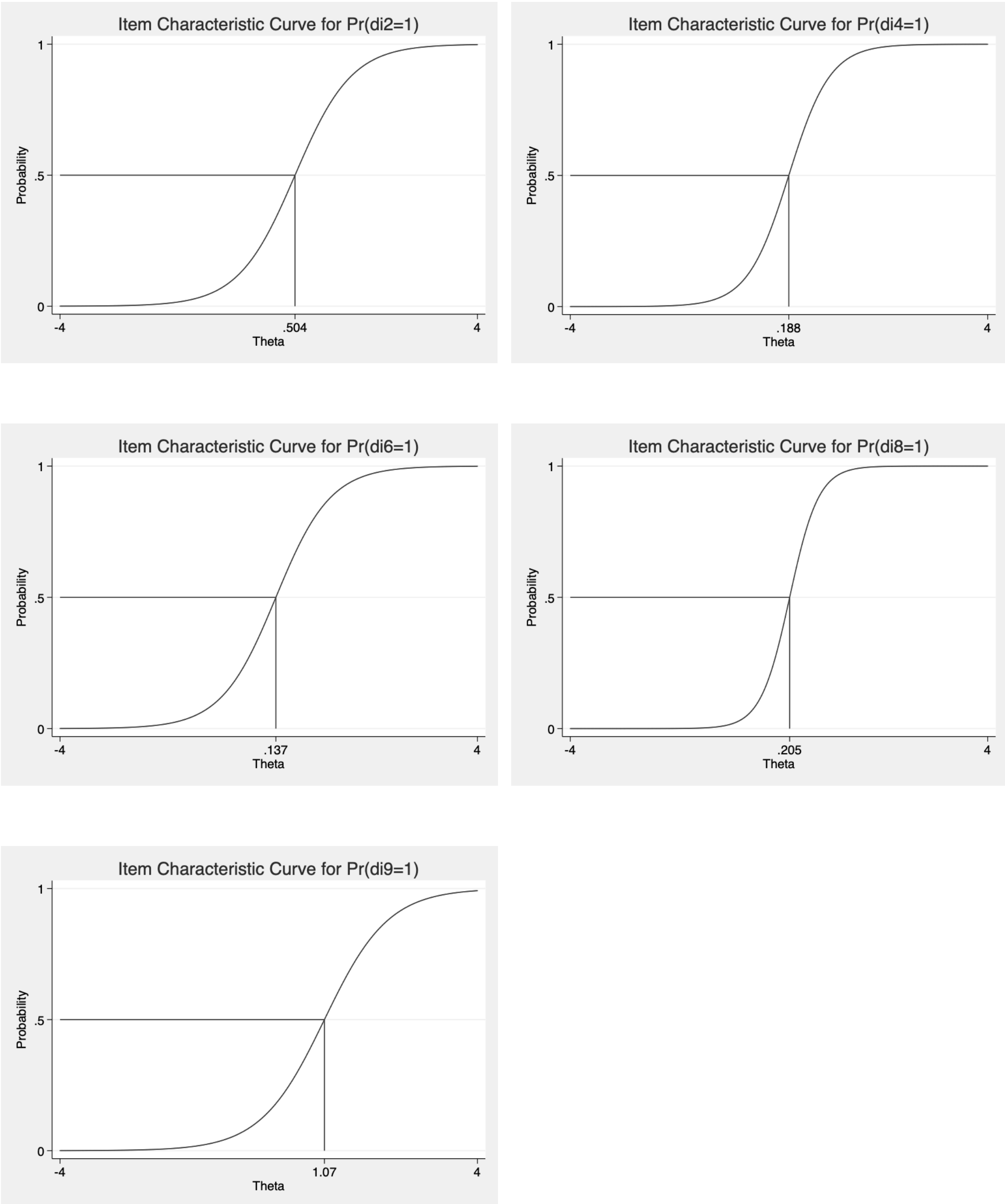


Figure 11. ICCs for the 14 mapped DSM-5 criteria

Figure 12 shows the Test Characteristic Curve (TCC) from the IRT model. It plots the expected score (number of ADHD symptoms indicated) against the theta scale created by the model. Lines are shown at theta = -1.96, 0, and 1.96 indicating low, average, and high expected scores (out of 14). Rounding these, (since decimal scores are not possible), indicates that 0 is very low, 3 is about average, and 11 is very high.

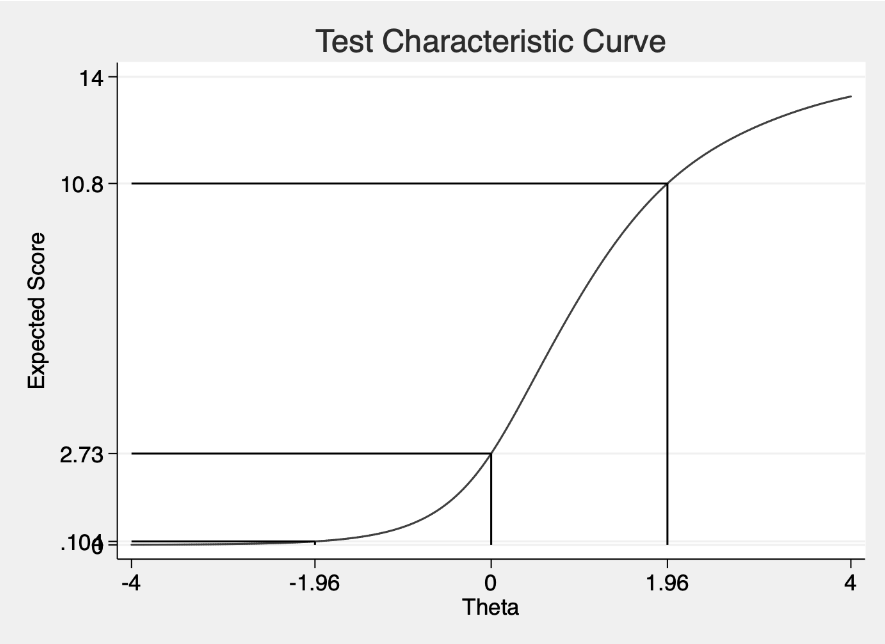


Figure 12. Test characteristic curve for 2PL IRT model of mapped and scored DSM-5 items, with theta lines

The next graph (Figure 13) shows the Test Information Function (TIF). This function is calculated by summing the Item Information Functions at each theta level, and thus it shows how much information is provided by the test overall for each theta level (Baker, 2001). In my 'test', or set of 14 DSM-5 criteria, the peak of the TIF shows that the most information is provided by the model just above theta levels of 0; the corresponding trough in the standard error plot reinforces this. The minimum theta value (or cut-off) associated with a BCS70 cohort member who met the DSM-5 criteria was 0.0176. This is also just above 0, so the point where the test provides the most information is ideal in terms of distinguishing between cohort members above and below that cut-off (Baker, 2001).

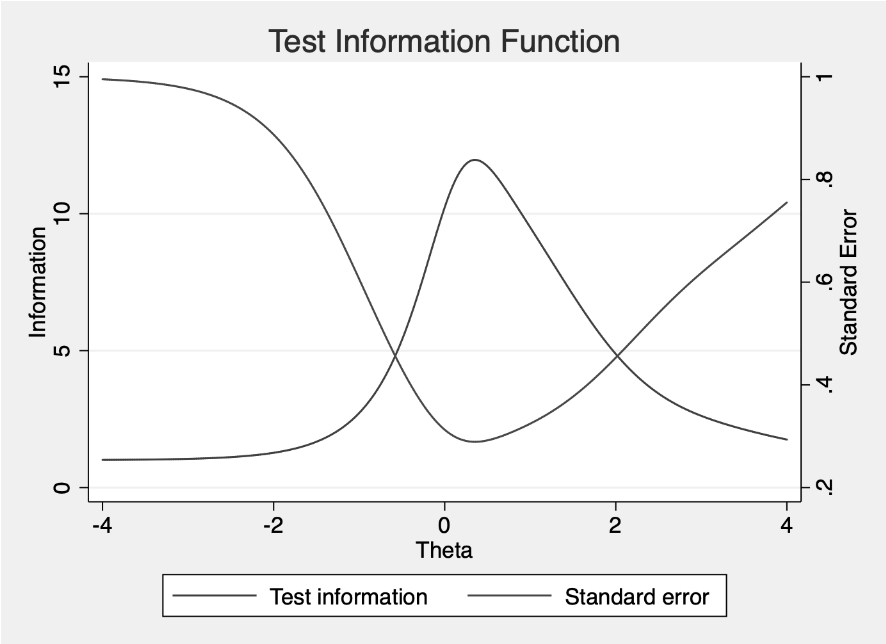


Figure 13. Test information function for 2PL IRT model

1.3 Answers to research questions

PRQ1: Is mapping DSM-5 items to BCS70 data and applying the criteria to identify an ADHD subgroup a viable approach?

Yes, a large proportion of the DSM-5 criteria could be mapped to BCS70 data items.

PRQ1a: How does the identified sample size compare to external sources of ADHD prevalence estimates?

The ADHD subgroup sample identified here of 889 was about 6% of the total sample, and ratio of boys to girls was about 2:1. These numbers are comparable to meta-analysis reports of prevalence based on non-clinical samples (Willcutt, 2012).

PRQ1b: How many of the identified ADHD subgroup responded to the BCS70 age 42 sweep (for evaluation of long-term outcomes)?

The ADHD subgroup of 889 was linked to the BCS70 age 42 data, and there were 472 matches. A larger sample would be better, but this is still an adequate number for building robust outcome models. However, power is limited to detect some effects.

PRQ2: Can an Item Response Theory (IRT) model with a good fit be built to score cohort members on a continuous latent construct of ADHD?

Yes, a continuous measure of ADHD was estimated using IRT and had indicators of acceptable model fit.

1.4 Conclusion

This pilot study developed a method for mapping BCS70 childhood behaviour questionnaire items answered by mothers and teachers at age 10 to an adapted version of current DSM-5 ADHD criteria. The adapted criteria were applied, and the resulting subgroup identified had a realistic size and ratio of males to females compared to recent epidemiological estimates, and also to an alternative method used to identify ADHD in BCS70 in previous research (Brassett-Grundy & Butler, 2004; Brassett-Harknett & Butler, 2007).

The method had limitations. For example, the exact same wording is not used on BCS70 items and DSM-5 criteria items. Independent raters should be enlisted to test mapping validity as a sensible next step. Also, a mapping of BCS70 items to SDQ hyperactivity subscale items could be used to further evaluate construct validity, since the SDQ subscale has been shown elsewhere to correlate to valid measures of ADHD (Ullebø et al., 2011).

This pilot study was a learning process, and through implementing the methods and supervisor feedback, I identified a number of improvements which were addressed in the final study. The improvements are summarised in Table 12.

No.	Limitation/feedback	Improvements
1)	Testing of assumptions not complete for IRT	Identify appropriate tests, test, and report results
2)	New mapping of DSM-5 items to BCS70 and resulting measure have limited validity	Identify and contact panel of experts on ADHD to validate mapping Compare to mapped SDQ and ADHD measure from (Brassett-Grundy & Butler, 2004)
3)	Distribution of ADHD latent construct is not normal	Identify distribution and build model with better fit
4)	Scale was not evaluated for measurement invariance	Test differential item functioning for males vs. females
5)	Compare categorical and dimensional measures	Compare derived subgroup to top 6% on ADHD severity scale, by key descriptive categories
6)	Link overall process to a methodological framework	Link to data mining process
7)	Two conditional items (behaviour problems and multiple settings) were not included in IRT model	Add to model for total of 16 items in scale

Table 12. List of limitations addressed in published study

2 Part 2: Copy of published study: A data mining and item response mixture modelling method to retrospectively measure DSM-5 ADHD in the BCS70

The objective of the published study was to answer the first research question of my thesis:

> RQ1: How can data science methods be used to retrospectively identify and validate robust categorical and continuous measures of DSM-5 ADHD in the BCS70?

The article citation is:

> *Cotton, J., & Baker, S. T. (2018). A data mining and item response mixture modelling method to retrospectively measure diagnostic and statistical manual-5 attention deficit hyperactivity disorder in the 1970 British Cohort Study. International Journal of Methods in Psychiatric Research.,*

which has been published in final form at https://doi.org/10.1002/mpr.1753. This article may be used for non-commercial purposes in accordance with the Wiley Self-Archiving Policy [http://www.wileyauthors.com/self-archiving].

N.B: the tables and figures in this section have been labelled with two sets of numbers; the first (far left) is the sequential number in the context of this thesis, and to the right of that, the number that was used in the published article.

The remainder of chapter 4 is a copy of the accepted version of the paper.

A data mining and item response mixture modelling method to retrospectively measure diagnostic and statistical manual-5 attention deficit hyperactivity disorder in the 1970 British Cohort Study

Joanne Cotton and Sara T. Baker
Faculty of Education, University of Cambridge

Corresponding author: Joanne Cotton, jmc236@cam.ac.uk

Acknowledgements

Mapping of DSM-5 ADHD symptoms to BCS70 items was reviewed via survey by 14 ADHD experts and two data analysis experts. Many thanks for taking the time to offer your views: Edmund Sonuga-Barke, Dave Coghill, Rosemary Tannock, Rafaela Marco, Cesar Soutullo, Eric Taylor, Hans-Christoph Steinhausen, Maite Ferrin, Jan R Wiersema, Lisa B Thorell, Mark Stein, Golam Khandaker, Anna Vignoles, Ricardo Sabates, and two anonymous reviewers. We are grateful to Edmund Sonuga-Barke and Tobias Banaschewski from the European Network for Hyperkinetic Disorders (EUNETHYDIS), and Laura Laughlin from the American Professional Society of ADHD and Related Disorders (APSARD), for facilitating distribution of the survey.

Grateful thanks also for assistance from Jung Yeon Park with adapting MPlus code from Wall, Park, & Moustaki, (2015), and Angela Brassett-Harknett with replicating Brassett-Grundy & Butler, (2008).

This study comprises part of Joanne Cotton's PhD, which is funded by the Economic and Social Research Council, award number 1652810.

Access to data was provided by the UK Data Service.

Abstract

Objective: To facilitate future outcome studies, we aimed to develop a robust and replicable method for estimating a categorical and dimensional measure of DSM-5 ADHD in the 1970 British Cohort Study (BCS70).

Method: Following a data mining framework, we mapped DSM-5 ADHD symptoms to age 10 BCS70 data (N=11,426) and derived a 16-item scale (α =0.85). Mapping was validated by an expert panel. A categorical subgroup was derived (n=594, 5.2%), and a zero-inflated IRT mixture model fitted to estimate a dimensional measure.

Results: Subgroup composition was comparable to other ADHD samples. Relative Risk Ratios (ADHD/not-ADHD) included: boys = 1.38, unemployed fathers = 2.07, below average reading = 2.58, depressed parent = 3.73. Our estimated measures correlated with two derived reference scales: SDQ hyperactivity (r=0.74), and a Rutter/Conners-based scale (r=0.81), supporting construct validity. IRT model items (symptoms) had moderate to high discrimination (0.90 – 2.81) and provided maximum information at average to moderate theta levels of ADHD (0.5 – 1.75).

Conclusion: We extended previous work to identify ADHD in BCS70, derived scales from existing data, modeled ADHD items with IRT, and adjusted for a zero-inflated distribution. Psychometric properties were promising and this work will enable future studies of causal mechanisms in ADHD.

Keywords: data-mining, IRT, ADHD, BCS70

1. Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is a disorder of inattention, impulsivity, and hyperactivity that interferes with functioning. It has three presentations: primarily inattentive, primarily hyperactive and impulsive, and combined (American Psychiatric Association, 2013), and affects approximately 6% of children worldwide (Polanczyk, de Lima, Horta, Biederman, & Rohde, 2007; Willcutt, 2012). Lifelong impairment often follows childhood ADHD, but about 50% are not significantly impaired as adults (Caye, Rocha, et al., 2016; Costello & Maughan, 2015). We can gain a better understanding of positive outcomes by studying causal mechanisms in the long term. However, methodological challenges have made it difficult to exploit existing longitudinal datasets to this end. Challenges include insufficient cohort age, sample biases, imprecise measures, and lack of psychosocial data. Here we propose a robust and replicable method to mitigate these challenges and facilitate future causal outcome analyses.

1.1 Methodological challenges

First, longitudinal data sources used in ADHD analyses are limited by cohort age. Most sources report adult ADHD outcomes between ages 18 and 25 (Cadman et al., 2016; Caye, Spadini, et al., 2016; Kuriyan et al., 2013; Lara et al., 2009; J. M. Swanson et al., 2017; van Lieshout et al., 2016). However, the brain continues to develop until about age 30 (Sowell et al., 2003), and imaging studies indicate that cortical development in ADHD is slower than average (P. Shaw et al., 2013). Additionally, there is a trend in Western societies to delay the traditional markers of 'settled' adulthood, such as stability of residence, marriage/partnership, and financial independence from parents (Arnett, 2000, p. 469). Thus, it is our view that long-term outcomes for ADHD should be evaluated after age 30.

Longitudinal data is needed from a cohort born in the mid-1980's or before to support post-age-30 outcomes analysis, but the current ADHD criteria have only been stable since 1987, or the DSM-III-R (American Psychiatric Association, 1987; Barkley, 2015). Yet, ADHD is a latent construct, i.e. not directly observable (Bollen, 2002), and latent constructs lend themselves to data mining, or "…the extraction of implicit, previously unknown, and potentially useful information from data." (Witten, Frank, Hall, & Pal, 2017, p. xxiii). Data mining could be used to retrospectively identify ADHD from data in a long-running, existing study, and mitigate the insufficient cohort age limitation.

Second, samples used for ADHD outcomes studies tend to be small, clinical, or based on retrospective recall (Caye, Spadini, et al., 2016; Cheung et al., 2015; Swanson et al., 2017). Small samples do not provide enough statistical power for complex modelling techniques needed to analyze long term trajectories (Wolf et al., 2013). Clinical samples tend to over-represent boys, cases with severe symptoms, and the combined type presentation of ADHD (Willcutt, 2012). Finally, non-clinical sample studies are often based on retrospective recall of childhood symptoms (Caye, Spadini, et al., 2016; Lara et al., 2009), which is affected by recall ability (Coughlin, 1990) and personality factors (Reuben et al., 2016). Accordingly, Caye et al. (2016) recommended that prospective cohort studies should be implemented. In the meantime, data-mining an existing long-running study could address all three of these biases.

Third, in studies of outcomes, ADHD is typically reported using an imprecise categorical indicator, i.e. 'ADHD' or 'not ADHD'. More sensitive dimensional measures are needed to detect individual differences (American Psychiatric Association, 2013; Gorter et al., 2015). A range of ADHD studies support this: in genetics, (Groen-Blokhuis et al., 2014; Thapar et al., 2013), neural connectivity (Elton et al., 2014), and performance on executive function tasks (Agnew-Blais et al., 2016; Salum et al., 2014). Derivation of a sensitive dimensional measure requires a large, minimally biased dataset.

Finally, identification of ADHD retrospectively in a rich dataset opens the possibility for longitudinal analyses on a variety of outcomes based on psychosocial factors, which are thus far understudied in the ADHD literature (Costello & Maughan, 2015).

In sum, insufficient cohort age, sample biases, imprecise measures, and lack of psychosocial data impede analysis of optimal ADHD outcomes. All could be mitigated by utilizing data from a large, long-term, population-based longitudinal cohort study, rich in psychosocial data. To this end, we short-listed candidate datasets, primarily based on data age, then reviewed in detail the following: Avon Longitudinal Study of Parents and Children (ALSPAC; 1991), 1970 British Cohort Study (BCS70) and Northern Finland Birth Cohort (NFBC) 1986. BCS70 was selected for preferable size, age, representativeness, and richness.

BCS70 is an ongoing population-based study of 17,198 children born from 5-11 April 1970. The study offers a rich array of health, psychological, social, and economic data from nine sweeps between ages 0 and 42 (Centre for Longitudinal Studies: UCL/IoE, 2019; Elliott & Shepherd, 2006). The third sweep at age 10 includes extensive data on behavior (Butler et al., 1997). Age 10 is ideal for assessing ADHD, because it is between 7, the most common age of diagnosis (Centers for Disease Control and Prevention, 2018a), and 12, the cut-off for diagnosis of childhood ADHD (American Psychiatric Association, 2013). Also, most of the ADHD-relevant questionnaire items in the BCS70 age 10 sweep were derived from the Rutter (Rutter, 1967) and Conners scales (Conners, 1969) (Butler et al., 1997), which are predecessors to current well-validated ADHD measures (American Academy of Pediatrics et al., 2002; Conners, 2008). Items were completed by both parents and teachers, providing valuable multiple-setting context (American Psychiatric Association, 2013; Butler et al., 1997). Finally, the age 10 sweep had 14,875 respondents and 11,426 with data on behavior, providing a plenteous sample to support complex statistical models and estimate a robust dimensional ADHD measure.

1.2 Literature review

We found only a handful of studies that derived a scale to measure ADHD, or a similar latent construct, in existing data. Brassett-Grundy & Butler (2004) derived a proxy measure for ADHD and evaluated outcomes at age 30 in BCS70. However, they used a combination of 23 Conners (Conners, 1969) and Rutter items (Rutter, 1967) to measure ADHD, including ten (e.g. "has difficulty using scissors"; Brassett-Grundy & Butler, 2004), which are not part of the current ADHD construct. Therefore, the construct they derived is unlikely to have specifically discerned ADHD as it is currently understood. Also, they calculated a simple sum and applied a clinical cut-off to create a categorical indicator, but did not estimate a dimensional measure.

Other researchers have derived measures of latent constructs like social and emotional skills (Goodman, Joshi, Nasim, & Tyler, 2015), self-control (Daly et al., 2015), and hyperactivity (Stuart-Smith et al., 2017) in BCS70 or similar datasets. They aggregated items and standardized as a general approach. Garcia-Barrera, Kamphaus, & Bandalos (2011) derived a scale to screen for Executive Function (EF) difficulties using items from the Behavior Assessment System for Children (BASC) in an existing dataset. They mapped BASC items to four EF domains, and estimated dimensional measures using factor analysis. Psychometric properties were evaluated using an expert panel to review the mapping, Cronbach's alpha, and measurement invariance by age and gender (Garcia-Barrera et al., 2011). A similar factor analysis approach has been used elsewhere to retrospectively measure intelligence, personality, and behavior factors (Gale et al., 2009; Prevoo & ter Weel, 2015b; von Stumm et al., 2009). These more complex methods address some of the key challenges faced with measuring ADHD in BCS70, including mapping items from an existing scale to an unmeasured construct, estimating with greater precision, and evaluating psychometric properties.

A more complex method is desired here, to provide a robust dimensional measure for use in future work. For our data, Item Response Theory (IRT) is a preferable modelling framework. IRT is a special case of confirmatory factor analysis which builds a model at the item level, leading to better generalizability across samples than other psychometric methods (Baker, 2001; Embretson & Reise, 2000). IRT fits here because the BCS70 age 10 dataset is large (N>500), the data are categorical (Embretson & Reise, 2000; van der Eijk & Rose, 2015), and factor structure evaluation indicates ADHD is most reliably measured as a unidimensional latent trait (Wagner et al., 2016). IRT models have been widely recommended for measuring psychiatric and health-related constructs (Edelen & Reeve, 2007; Gorter et al., 2015; Muthén & Asparouhov, 2006; Sturm et al., 2017). Importantly, other authors have used IRT to evaluate psychometric item properties of DSM ADHD criteria (Arias et al., 2018; Gomez, 2007, 2008, 2011, 2012; Gomez et al., 2011; Li et al., 2015), compare model fit in sub-samples (Polanczyk et al., 2010), and provide quantitative verification of diagnosis (Lindhiem et al., 2015). These IRT studies reported good indicators of model fit in a variety of clinical and non-clinical samples.

Whilst IRT models are robust to some non-normality, they assume an approximately normal distribution (Reise & Revicki, 2015). We should not assume a normal distribution for ADHD (or any psychiatric disorder) in a population-based sample (Kaat & Farmer, 2017; Reise & Waller, 2009; Wall et al., 2015). A large proportion of respondents are expected to have zero symptoms or very few (Finkelman et al., 2011; Reise & Waller, 2009; Wall et al., 2015). Simulation studies have shown that ignoring non-normality of a latent trait in IRT can lead to significant estimation errors (e.g. inflated discrimination parameters), and adjustments are recommended (Kaat & Farmer, 2017; Sass et al., 2008; Wall et al., 2015; Woods, 2015). There are a few ways to adjust for non-normality in IRT, including the Empirical Histogram, Ramsay Curve, (Woods, 2015), and Zero-Inflated Mixture Model (Wall et al., 2015; ZIMM). The latter method specifically adjusts for the zero-inflation we expect to find with ADHD in BCS70.

1.3 Present study

Our objective was to develop and demonstrate a robust method to derive a categorical and dimensional measure of ADHD in the BCS70 age 10 data, enabling future studies of outcomes. We aimed to incorporate a data-mining framework, apply approximate DSM-5 diagnostic criteria, develop an IRT model adjusted for zero-inflation, and evaluate psychometric properties.

2. Method

2.1 Data

Age 10 BCS70 data were collected in 1980 and 1981 in the United Kingdom. Ten questionnaires were completed by medical professionals, parents, teachers, and participants (Centre for Longitudinal Studies, 2015). Data was accessed through the UK Data Service (University of Essex et al., 2012).

In the age 10 sweep, cohort members (N=14,875) were 96% 'English, etc.', 51.5% boys, and 63.9% of their parents had jobs in the 'middle' social classes, designated in 1980 as 'III-manual', 'III-non-manual', and 'IV-partly-skilled'. All were born in April 1970. Children with parents born outside Britain, single mothers, teenage mothers, mothers over 40, unemployed fathers, and low parental education level were under-represented due to attrition (Butler et al., 1997, p. 35). The ADHD-relevant behavior questionnaire items were left blank by many respondents (n=3,449); these observations were excluded from our sample (N=11,426).

2.2 Ethics

An ethics checklist was approved by the Faculty of Education, University of Cambridge, based on British Educational Research Association (BERA) guidelines (BERA, 2011). Ethical procedures for the original study (BCS70) adhere to BERA and ESRC guidelines (Centre for Longitudinal Studies: UCL/IoE, 2014).

2.3 Tools

Analyses were conducted using Stata 14.2 (StataCorp LLC, 2015), MPlus 8 (Muthen & Muthen, 2017), Microsoft Excel, and Qualtrics (Qualtrics, 2017).

2.4 Measures

2.4.1 DSM-5 ADHD criteria

There are 18 symptoms: nine hyperactive/impulsive, and nine inattentive, plus six additional conditions, totaling 24 items. The diagnostic threshold requires at least six symptoms from either or both lists of nine to be observed 'often', along with all six conditions. Depending on which symptom thresholds are met, presentation types of Primarily Hyperactive and Impulsive (PHI), Primarily Inattentive (PI), or Combined (C) are applicable (American Psychiatric Association, 2013). In the present study we have used abbreviations to refer to the DSM-5 ADHD criteria; for example, 'dh1' refers to the 1st symptom in the DSM-5 list of hyperactive/impulsive symptoms.

2.4.2 BCS70 age 10 behavior items

53 items from the maternal self-completion form and educational questionnaire pertained to child behavior (Butler et al., 1997). The items were completed by a parent and teacher, respectively. Most were based on Rutter (Rutter, 1967) and Conners (Conners, 1969) items, though a handful were written, tested and added by the BCS70 study designers (Butler et al., 1997). An example item was 'Is squirmy or fidgety', and the respondent (parent or teacher) indicated the extent to which the statement applied to the child (see Figure 1).

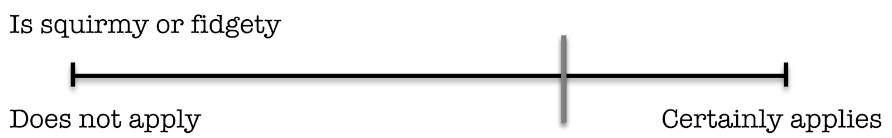


Figure 14. Figure 1. Example of Visual Analog Scale item used in BCS70 age 10 sweep
Respondent indicated the extent of their agreement with the item by marking a vertical line on the horizontal scale

2.4.3 Strengths and Difficulties Questionnaire (SDQ) hyperactivity subscale

The subscale for ages 4-17 consists of five items (abbreviated): restlessness, fidgeting, distractibility, impulsivity, and attention span (Goodman, 1997; youthinmind, 2012). The subscale has been validated for use as a diagnostic screener and in research as a proxy for ADHD diagnosis (Stone et al., 2010; Ullebø et al., 2011).

2.5 Approach

Our approach was guided by a data mining framework, and included three phases: 1) data assessment and preparation, 2) modelling, and 3) evaluation (Kurgan & Musilek, 2006, p. 6-7).

2.5.1 Data assessment and preparation

This first phase entailed item mapping, recoding, application of DSM-5 criteria, and model selection.

2.5.1.1 Item mapping and derived scale

Using the 24 DSM-5 ADHD items as a reference point, the 53 BCS70 behavior items were inspected visually for semantically similar content. Next, all the remaining (~2,900) data items from the age 10 sweep were checked for further mapping candidates using keyword searches and visual inspection. We successfully mapped 19 (79%) of the 24 DSM-5 items: five/nine inattentive, nine/nine hyperactive/impulsive, and five/six conditions, to BCS70 items. No mapping could be found for: di1-careless mistakes, di3-doesn't listen, di5-trouble organizing, di7-loses things, or dc6-symptoms > 6 months. Three of the conditions, dc1-symptoms by age 12, dc4-no other psychiatric disorder, and dc5-symptoms not part of another psychiatric disorder, were mapped to the BCS70 data, but had insufficient variation to be useful in a scale, so were excluded from the resultant 16-item scale.

A panel of 16 international experts completed an online survey to review the item mapping. Adjustments were made to reflect their views (Appendix A in the supporting information includes survey instructions, example questions and results, and details of adjustments). The final mapping of DSM-5 to BCS70 items and our derived 16-item scale is reported in Table 1.

No.	DSM-5 criteria	BCS70 questionnaire items †
Inattentive		

Di1	Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or with other activities.	No mapping found
Di2	Often has trouble holding attention on tasks or play activities.	R-j155 - Pays attention to what is being explained in class m82 - Has difficulty concentrating on any particular task though may return to it frequently j129 - Cannot concentrate on any particular task, even though the child may return to it frequently j077 - How well does this child concentrate on educational tasks, in comparison with the average 10-year-old?
Di3	Often does not seem to listen when spoken to directly.	No mapping found
Di4	Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., loses focus, side-tracked).	m76 - Fails to finish things he/she starts, short attention span R-j174 - Child completes tasks which are started j177 - Fails to finish things he starts
Di5	Often has trouble organizing tasks and activities.	No mapping found
Di6	Often avoids, dislikes, or is reluctant to do tasks that require mental effort over a long period of time (such as schoolwork or homework).	R-j139 - Shows perseverance; persists with difficult or routine work
Di7	Often loses things necessary for tasks and activities (e.g. school materials, pencils, books, tools, wallets, keys, paperwork, eyeglasses, mobile telephones).	No mapping found
Di8	Is often easily distracted	m65 – Inattentive, easily distracted j152 – Is easily distracted

Di9	Is often forgetful in daily activities.	j158 – Is forgetful when given a complex task
Hyperactive		
Dh1	Often fidgets with or taps hands or feet, or squirms in seat.	m44 - Is squirmy or fidgety j151 - Squirmy and fidgety m77 - Given to rhythmic tapping or kicking j165 - Given to rhythmic tapping or rhythmic kicking during class j082 - What percentage of the time is the child fidgeting and indulging other minor distracting activities, when he/she is expected to be working? (paraphrased)
Dh2	Often leaves seat in situations when remaining seated is expected.	j081 - What percentage of the time is the child moving around the classroom, when he/she is expected to be working? (paraphrased)
Dh3	Often runs about or climbs in situations where it is not appropriate (adolescents or adults may be limited to feeling restless).	m43 - Very restless. Often running or jumping up and down. Hardly ever still.
Dh4	Often unable to play or take part in leisure activities quietly.	m57 - Cannot settle to do anything for more than a few moments
Dh5	Is often "on the go" acting as if "driven by a motor".	m72 – Shows restless or overactive behavior j150 - Shows restless or overactive behaviour
Dh6	Often talks excessively.	j080 - What percentage of the time is the child talking to other children, when he/she is expected to be working? (paraphrased)
Dh7	Often blurts out an answer before a question has been completed.	m73 – Is impulsive, excitable

Dh8	Often has trouble waiting his/her turn.	m71 - Requests must be met immediately, easily frustrated j175 - Requests must be met immediately - easily frustrated
Dh9	Often interrupts or intrudes on others (e.g., butts into conversations or games)	m74 - Interferes with the activity of other children j142 - Interferes with the activities of other children
Conditions		
Dc1	Several inattentive or hyperactive-impulsive symptoms were present before age 12 years	True for all cases; criteria were evaluated at age 10
Dc2	Several symptoms are present in two or more settings, (such as at home, school or work; with friends or relatives; in other activities)	Both mother and teacher indicated three or more symptoms were present
Dc3	There is clear evidence that the symptoms interfere with, or reduce the quality of, social, school, or work functioning	As a proxy, criterion was considered met if the child was in the 'moderate' or 'severe' behavior problems group based on their (mother) Rutter items score.
Dc4	The symptoms are not better explained by another mental disorder (such as a mood disorder, anxiety disorder, dissociative disorder, or a personality disorder).	Cohort members were excluded if they had been diagnosed with another psychiatric disorder, as per the medical questionnaire, identified by ICD-9 codes. Only two children fulfilled this criterion.
Dc5	The symptoms do not happen only during the course of schizophrenia or another psychotic disorder	Assumed if no diagnosis - See item 4
Dc6	Symptoms should be present for at least six months	No mapping found

Table 13. Table 1: Mapping of DSM-5 criteria to BCS70 age 10 questionnaire items (paraphrased)

(American Psychiatric Association, 2013; Centre for Longitudinal Studies: UCL/IoE, 2019)
† Note on item codes: 'm' - Maternal Self Completion questionnaire, 'j' - Educational questionnaire, and 'R'- reverse coded

2.5.1.2 Recoding

Most of the mapped BCS70 items were presented to respondents using Visual Analog Scales (VAS; Figure 1). Post-completion, coders assigned values of 1-47 (teacher items), or 0-100 (mother items; Butler et al., 1997). Subsequently, studies have shown that VAS scales function as categorical rather than continuous variables because equal distance cannot be assumed between points; the likely maximum is three to four categories (Svensson, 2001; Wewers & Lowe, 1990). Hence, we recoded VAS items into more plausible categories. Visual inspection of histograms for raw VAS data indicated three roughly-equal-sized response levels. This is consistent with other measures of ADHD (e.g. the SDQ), which use 'not true', 'sometimes true' and 'certainly true' (or similar) as levels. However, the DSM-5 criteria are worded in a dichotomous way: symptoms occur 'often', or 'not often'. Accordingly dichotomous coding has been used in other IRT-based measures of ADHD (Gomez et al., 2011; Lindhiem et al., 2015). Therefore, we divided the scales into thirds and equated the bottom two-thirds to 'not true' and 'sometimes true', recoding both to 'not often' (0). The top third was equated to 'certainly true' and recoded as 'often' (1). Items were reverse coded as appropriate.

Three BCS70 teacher items (j080-talking, j081-moving around, j082-fidgeting) used a different scale ('what percentage of the time does the student spend…'). Precedent could not be found for categorically recoding this type of data. We coded only observations >=3 SDs from the mean as 'often' (1), which was difficult to achieve, but supported conservative inferences.

If more than one BCS70 item from parent or teacher mapped to a single DSM-5 criterion, the DSM-5 criterion was considered met if any of the mapped BCS70 items were met.

2.5.1.3 Application of DSM-5 ADHD criteria

Next, a categorical ADHD indicator and presentation type were derived by applying (approximated) DSM-5 diagnostic criteria (American Psychiatric Association, 2013) to our 16-item scale (Figure 2).

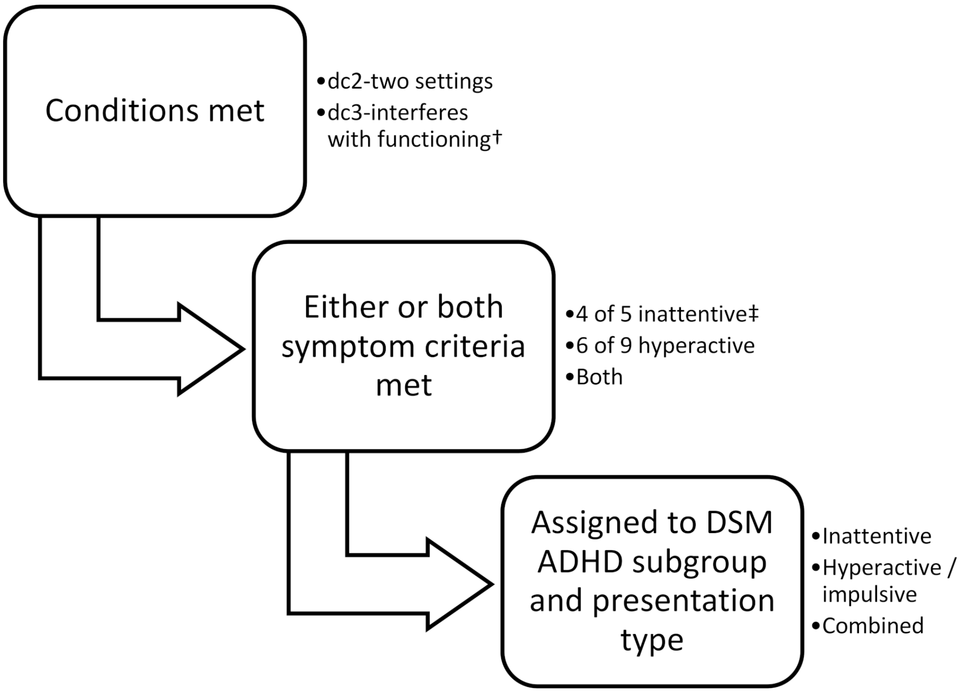


Figure 15. Figure 2. Process used to apply our approximation of DSM-5 ADHD criteria

† Conditions dc4 and dc5 (both based on another psychiatric diagnosis) were omitted from our scale due to insufficient variability. However, two children in our sample were explicitly excluded from the DSM-5-based ADHD subgroup due to another psychiatric diagnosis.

‡ 6/9 is two-thirds, so two-thirds of the of the 5 symptoms was used as a best approximation (3.35, rounded up to 4, to support conservative inferences)

2.5.1.4 Model selection

Descriptive statistics for a simple sum score of the 16 dichotomous items indicated a non-normal, zero-inflated distribution (i.e. a large proportion of the sample had zero symptoms: n=2,869, or 25%; see Figure 3). This supported use of a ZIMM model (Wall et al., 2015) for our analyses.

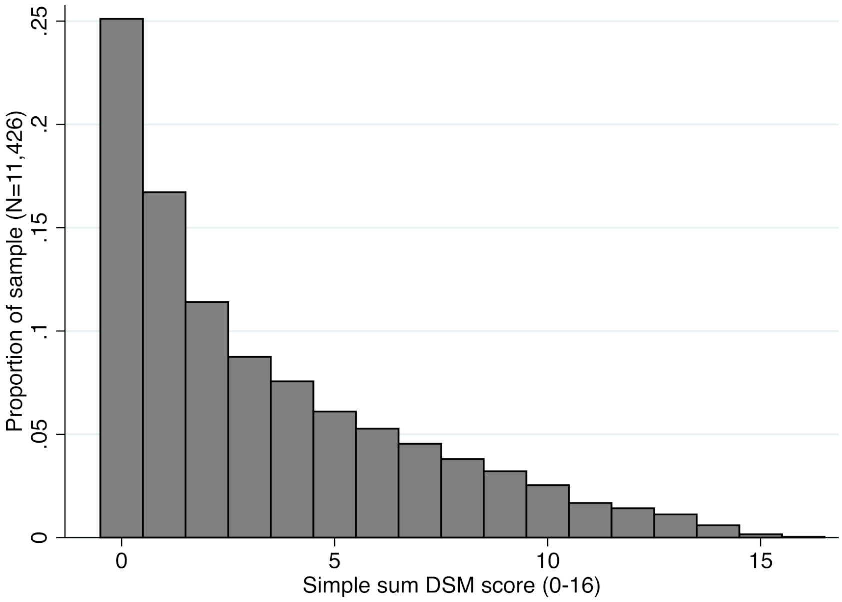


Figure 16. Figure 3. Histogram of mapped DSM-5 ADHD score (simple sum)
Demonstrates zero-inflated distribution

ZIMM is a zero-inflated mixture model, with 'mixture' referring to latent class and factor components. ZIMM uses a degenerate ('non-clinical') class, with an extreme fixed negative mean (μ= -100) and zero variance, to adjust for the influence of the large proportion of observations with zero symptoms (Wall et al., 2015). The second, 'clinical' class is then dominant in the estimation of model parameters, providing a dimensional measure of the latent trait that is less unduly biased by non-clinical cases (Finkelman et al., 2011; Magnus & Thissen, 2017; Wall et al., 2015).

For dichotomous data like ours, IRT models can estimate between one and four parameters: 1PL/2PL/3PL/4PL. The four parameters, building cumulatively, are: difficulty (i.e. location or threshold), discrimination, lower/guessing asymptote, and upper/fatigue asymptote (Embretson & Reise, 2000; Magis, 2013). DSM-5 ADHD items are unequal in their ability to discriminate (see Arias et al., 2018), so slopes vary and 1PL estimating difficulty only is not adequate. The third and fourth lower and upper asymptote parameters are relevant in educational tests measuring ability, where respondents are motivated to achieve a high score (Embretson & Reise, 2000; Magis, 2013). Accordingly, 3PL and 4PL are not appropriate for psychiatric constructs (Finkelman et al., 2011). Therefore, the two-parameter logistic (2PL) model (Birnbaum, 1968) was used here. The 2PL model is operationalized through an item characteristic curve (ICC) for each item, with the following equation:

$$\Pr(X = 1) = \frac{e^{a(\theta-b)}}{1 + e^{a(\theta-b)}}$$

Where Pr = probability, X = response to the item (either 0 or 1), a = item discrimination, b = item difficulty, and θ = individual scaled factor score. Discrimination is the slope of the ICC at the steepest point, indicating how dramatically the probability of a positive response increases over the range of factor scores (θ). Difficulty is the point on the ICC where the probability of either (0 or 1) response is 50% (Baker, 2001).

2.5.2 Modelling

Within the data mining framework, modelling comprised testing model assumptions, building plausible models, and selecting a model with the best fit to the data.

2.5.2.1 Validation of IRT assumptions

Unidimensionality and local independence were supported by factor analysis on a matrix of tetrachoric correlations for the 16 items, showing clear dominance on a first factor (4.9 times the second factor), and low (<0.30) correlation residuals for each item pair (Embretson & Reise, 2000; Hambleton et al., 1991). Tetrachoric correlations were used because they generate less error than Pearson's with categorical data (Embretson & Reise, 2000). Monotonicity was observed using Mokken's rule (Hardouin et al., 2011). The test indicated that item dh6-talks excessively, fell slightly short (H=0.26) of the criteria for being part of a strong scale (Loevinger's H>0.30; Hardouin et al., 2011).

The ZIMM models were based on Wall et al. (2015). We compared three variations (Table 2). The log likelihood, AIC, and BIC initially pointed to the ZIMM three class model as the best fit, but Entropy was low (0.45), indicating too many classes (Celeux & Soromenho, 1996). Thus, the ZIMM two-class model was selected, which aligns with findings from the Wall et al., (2015) study. Mplus code for the ZIMM two-class model is provided in the supporting information (Appendix B).

Model	No. of parameters	logL	BIC	AIC	Entropy
1. 2PL IRT/1 class	32	-71944.82	143953.64	144188.63	NA
2. ZIMM 2 class	33	-71930.19	143926.39	144168.73	0.80
3. ZIMM 3 class	35	-71898.60	143867.21	144124.23	0.45

Table 14. Table 2: Comparison of three item response models for dimensional measure

logL = log likelihood, BIC = Bayesian Information Criterion, AIC = Akaike's Information Criterion; fit statistics calculated in MPlus

3. Results

Results comprised an evaluation of psychometric properties for the derived 16-item scale, categorical measure based on DSM-5, and dimensional measure based on the ZIMM two-class model. Evaluation is the third phase of our data mining framework.

3.1 Derived 16-item scale

Reliability was good (Cronbach's α=0.85), and face validity was confirmed by an expert panel review (see section 2.5.1.1).

3.2 Categorical measure based on DSM-5 criteria (ADHD subgroup)

The derived ADHD subgroup (n=594) was 5.2% of the N=11,426 sample.
Since the data were collected in 1980-81 and no validated measures of DSM-5 ADHD were available (Butler et al., 1997), novel approaches were required to assess construct validity. These included comparisons to epidemiology and derived reference scales.

The DSM-5 ADHD subgroup had a similar composition to epidemiology/meta-analyses estimates of overall prevalence, gender, and subtype (Table 3). The subgroup was also comparable to epidemiology reports on ADHD samples, with over-representation of boys, health, social and economic disadvantages, and below average cognitive abilities (Table 4; Costello & Maughan, 2015; Loe & Feldman, 2007; Matza, Paramore, & Prasad, 2005; Willcutt, 2012).

Attribute	ADHD subgroup	Meta-analysis†
Prevalence	5.2%	6.1-7.1%
Ratio of boys to girls	2.3 : 1	2.4 : 1
Combined	35.6%	~32%
Hyperactive	12.4%	~18%
Inattentive	52.0%	~50%

Table 15. Table 3: DSM-5 categorical subgroup compared to recent meta-analysis estimates

†(Willcutt, 2012, p. 492), data based on estimates from Table 1, only using full DSM-IV criteria data from parents and teachers, as these were most comparable to the method used in the present study. Precise figures were not available for the subtypes, so the '~' symbol indicates an approximation based on the data available.

Attribute	% of ADHD subgroup †	% of non-ADHD subgroup	Relative Risk Ratio (RRR) ‡
Boys	69.90	50.50	1.38
Lives in residential institution	1.90	0.40	4.75
Attends special school	3.20	0.64	5.00

Any medical condition	51.80	24.10	2.15
Family demographics			
Single mother	9.91	5.50	1.80
Unemployed father	6.13	2.96	2.07
Low family income	11.70	6.70	1.74
Cognitive abilities			
Below average reading age (<-1SD)	43.40	16.80	2.58
Below average maths (<-1SD)	44.60	15.20	2.93
Social class			
Professional or Managerial & Technical	16.50	29.80	0.55
Non-manual & manual	52.50	52.40	1.00
Partly skilled or Unskilled	25.30	17.80	1.42
Parent Malaise Inventory			
Severe problems (95+ percentile)	15.30	4.10	3.73

Table 16. Table 4: Descriptive characteristics of DSM-5 categorically identified ADHD group compared to non-ADHD group

† ADHD Subgroup N = 594, non-ADHD subgroup N = 10,832; denominator in ratio varies as missing data are excluded

$$\ddagger\ RRR = \frac{Risk\ of\ factor\ in\ ADHD\ group}{Risk\ of\ factor\ in\ non\ ADHD\ group}$$

N.B. Relative Risk Ratio (RRR) > 1 indicates disadvantage, and < 1 indicates advantage (e.g. Professional and Managerial Social Class); RRR is also an effect size.

The SDQ hyperactivity subscale items were mapped (youthinmind, 2014b) to items from BCS70 (Table 5) and a sum score was derived for comparison. The simple sum score from our scale was highly correlated with the SDQ subscale score (r = 0.74, p < .001), supporting construct validity.

No.	SDQ	BCS70 questionnaire items
2	Restless, active, cannot stay still for long	M43 - Very restless. Often running about or jumping up and down. Hardly ever still.
10	Constantly fidgeting or squirming	M44 - Is squirmy or fidgety
15	Easily distracted, concentration wanders	M65 - Inattentive, easily distracted
21	Thinks things out before acting	R-M73 - Is impulsive, excitable
25	Sees tasks through to the end, good attention span	R-M76 - Fails to finish things he/she starts, short attention span

Table 17. Table 5: SDQ hyperactivity subscale mapping to BCS70 items

VAS scores were recoded as follows: 0-32 -> 0 not true, 33-67 -> 1 somewhat true, 68-100 -> 2 certainly true.

Additionally, we replicated part of a study that derived a proxy measure for ADHD in BCS70. Their measure was based on Conners (Conners, 1969) and Rutter (Rutter, 1967) items (Brassett-Grundy & Butler, 2004), including several that are not currently considered part of the DSM-5 ADHD construct (see Literature Review). The replication-based subgroup (N=1,102) was much larger than ours (N=594) and membership overlapped only 66.5%. However, the simple sum scores from their scale (mother + teacher) and ours were highly correlated (r=0.82, p<.001), also providing some support for construct validity.

3.3 ZIMM model and estimated dimensional score

The two-class ZIMM was used to estimate a factor score (theta) for our sample; (N=11,426, M = -0.06; SD = 0.91). For cases with zero symptoms (n=2,869), M=-1.14, and for the remainder (n=8,557), M=0.30. The overall distribution had a similar shape to the simple sum score, though substantially more nuanced in variation, as predicted (Figure 4; note contrast to Figure 3).

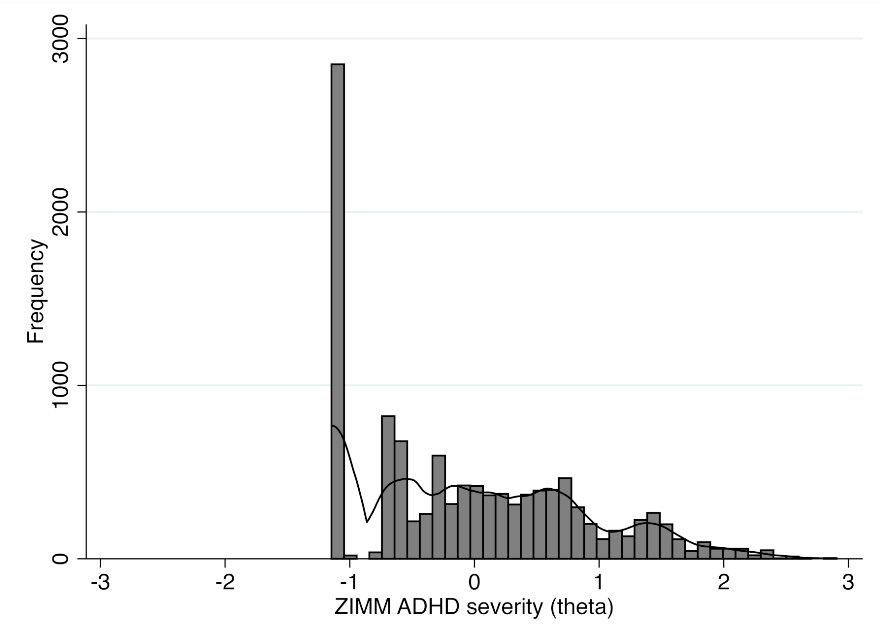


Figure 17. Figure 4. ADHD severity score estimated with ZIMM two-class model (N=11,426)

Showed expected zero inflation but with desired individual variation in ADHD severity

The IRT theta score correlated with the other measures derived, as expected. Logistic regression showed a significant association with the DSM-5-based categorical measure; (N=11,426, χ^2=3201.38, p<0.001, df=1; McFadden's R^2=0.69). Also, there was a large and significant positive correlation between theta and the derived SDQ subscale score (r=0.74, p<0.001), as well as the derived mother + teacher score (r=0.81, p < 0.001) calculated by part-replication of Brassett-Grundy & Butler (2004).

All the ZIMM two-class discrimination and difficulty parameters were significant, ($p<.001$; Table 6). Discrimination for symptoms ranged from 0.90 to 2.81, or moderate to very high (Baker, 2001). Difficulty ranged from 0.49 to 3.62, functioning best for individuals just above average to very high on the ADHD trait (Baker, 2001).

Item	Discrimination (α)	Difficulty (b)
Dh1 - fidgets or squirms	1.92	.53
Dh2 - inappropriately leaves seat	1.19	3.62
Dh3 – inappropriately runs about	1.19	1.09
Dh4 – cannot play quietly	1.73	1.57
Dh5 – on the go, 'driven by motor'	1.97	1.09
Dh6 – talks excessively	.90	3.27
Dh7 – blurts answers	1.30	1.38
Dh8 – trouble waiting turn	1.28	1.13
Dh9 – interrupts, intrudes	1.56	1.45
Di2 – trouble holding attention	1.49	.62
Di4 – doesn't follow through	1.74	.49
Di6 – avoids long tasks	1.37	1.05
Di8 – easily distracted	2.81	0.28
Di9 – often forgetful	1.27	1.25
Dc2 – symptoms interfere	1.31	1.40
Dc3 – multiple settings	5.09	1.24

Table 18. Table 6: ZIMM 2 class 2PL IRT parameters

3.3.1 Information Characteristic Curves (ICC)

All 16 ICC curves visually supported the moderate-to-high ability of the items to discriminate between respondents (Figure 5; Baker, 2001). The most discriminating symptoms were di8-easily distracted (α=2.81) and dh5-'on the go/motor' (α=1.97). The least discriminating was dh6-talks excessively (α=0. 90). Two items had high difficulty: dh2-inappropriately leaves seat (β=3.62) and dh6-talks excessively (β=3.27), only providing information at very high levels of ADHD. Low difficulty items were dh1-fidgets (β=0.53), di2-trouble holding attention (β=0.62) and di4-doesn't follow through (β=0.49).

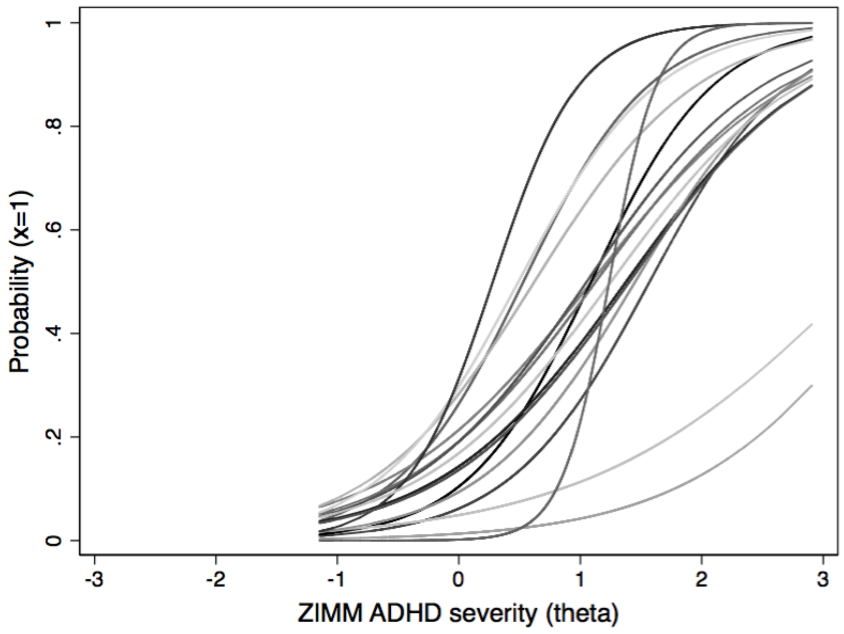


Figure 18. Figure 5. ICC curves of derived 16-item scale based on ZIMM two-class model
ICC = Item Characteristic Curve, ZIMM = Zero-inflated mixture model
Showed that items (other than the two flatter curves) discriminate well between individuals

3.3.2 Test information function

The Test Information Function shows how much information is provided by all items on the 16-item scale or 'test' at varying levels of the latent trait, based on the ZIMM two-class model (Figure 6). The curve shows our model provides the most information between theta values of 0.5 and 1.75, i.e. average to moderate levels of ADHD severity.

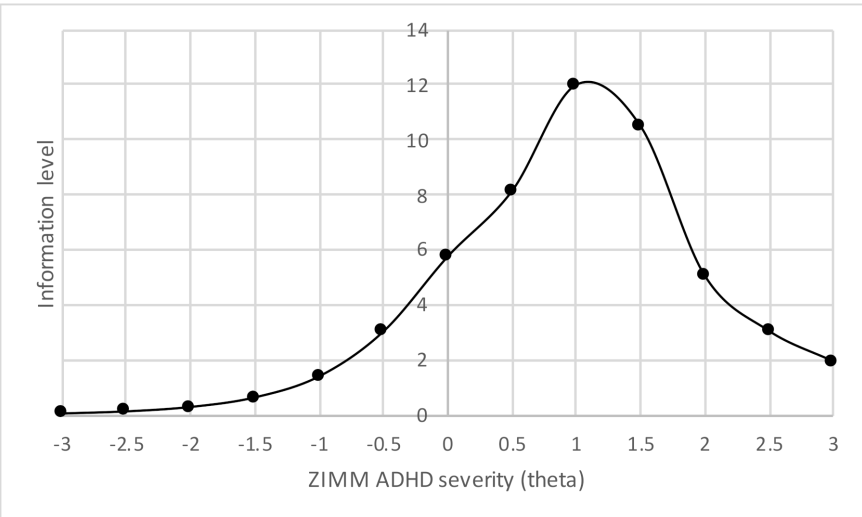


Figure 19. Figure 6. Test Information Function (TIF) for derived 16-item scale based on ZIMM two-class model
Showed the scale provides the most information at moderate levels of ADHD severity

3.3.3 Differential Item Functioning (DIF)

Other child mental health scales evaluate DIF (or measurement invariance) by gender, age and informant (e.g. the SDQ; youthinmind, 2014a). Age and informant were not applicable here because all participants were the same age, and our scale is based on combined responses from parent and teacher informants. Thus, we evaluated DIF by gender. According to the Mantel-Haenszel method, four items had significant DIF ($p<0.05$): two in favor of males and two in favor of females. However none had a large enough effect size to justify removal based on the Educational Testing Service (ETS) A/B/C classification method (Holland & Thayer, 1986).

3.4 Comparison of categorical and dimensional measures

Our DSM-5-based ADHD subgroup comprised 5.2% (n=594) of the sample (N=11,426). We compared this group to the top 5.2% (n=594) of the sample using the ranked IRT ADHD theta score (Table 7). 425 children (71.5%) were in both groups. Children in the IRT-based subgroup had slightly higher sum and theta scores, and were marginally more likely to be boys, have a medical condition, or a below-average reading age. They were less likely to have an unemployed father, or a parent with severe malaise (depression). 159 of the 169 children in the IRT-based group but not in the DSM-V-based group were missing the DSM condition dc3-symptoms interfere with functioning (based on the parent-rated Rutter behavior score). Nine were just under the threshold for both symptom lists (i.e. 3 inattentive symptoms and 5 hyperactive), and one had another psychiatric diagnosis, which was not taken into consideration in the IRT model.

Attribute	IRT ADHD subgroup (n=594)	DSM-5 ADHD subgroup (n=594)	RRR†
Average sum score	12.1	11.5	
Average IRT score	1.9	1.8	
	%	%	
Boys	73.4	69.9	1.05
Any medical condition	52.8	50.8	1.04
Below average reading age	35.0	31.8	1.10
Unemployed father	.04	.05	.90
Parent with severe malaise	13.1	15.3	.86

Table 19. Table 7: Comparison of the top 5.2% based on IRT factor scores to the DSM-5-based categorical subgroup

† *See notes on RRR (Relative Risk Ratio) with Table 4.*

4. Discussion

Our objective was to develop and demonstrate a method to derive a categorical and dimensional measure of ADHD in existing data. We chose the BCS70 to mitigate limitations of insufficient cohort age, sample biases, and imprecise measures typically found in longitudinal studies of ADHD. A data mining framework was used to guide the approach. DSM-5 ADHD criteria were mapped to age 10 data items from BCS70 to derive a 16-item scale, and the mapping was validated by an expert panel. An approximation of the DSM-5 ADHD diagnostic procedure was used to identify a subgroup of children with ADHD symptomatology (N = 594; 5.2%). Prevalence is slightly lower than epidemiology estimates of 6%, perhaps because disadvantaged groups were under-represented in our sample, and disadvantaged groups tend to be over-represented in ADHD samples (Russell et al., 2014). A ZIMM two-class model was selected as the optimal model for estimating a dimensional measure of ADHD, based on the non-normal, zero-inflated distribution, and comparison to two other plausible model variations. Psychometric properties tested for the 16-item scale, categorical ADHD measure, and dimensional ADHD measure were promising.

We included five of the six DSM-5 ADHD conditions, which is a strength given that most studies only evaluate symptoms (see Willcutt, 2012). However, four inattentive criteria and one of the conditions could not be mapped (Table 1). Nevertheless, the prevalence of inattentive type presentation in our sample was comparable to meta-analytic findings (Willcutt, 2012). This could be partially explained by findings from Li et al. (2015), who evaluated the full scale and found that two of the items missing from our scale had significant local dependence (di5 and di7; Li et al., 2015). Also Arias et al. (2018) analyzed the full scale and found that the most information was provided by three items (dh5, di2, and di8; Table 1), all of which were in our scale, possibly offsetting the absent items.

Two items, dh2-leaves seat and dh6-talks excessively, were based on BCS70 items from an unusual scale, and to be conservative we only coded an ‘often’ response for values 3SDs above the mean. Both items were accordingly high on difficulty parameters, and dh6 appeared as a weaker item per Mokken’s rule and Loevinger’s H. We accepted the high difficulty because it provides information at higher levels of the trait, which is desirable for our purposes. Regarding the relative weakness of dh6, we did not consider this an aberration, because others studies using typical levels of scale measurement also found dh6 to be a weaker item in terms of information provided (Arias et al., 2018; Gomez, 2011; Li et al., 2015).

The two approaches used to identify an ADHD subgroup (DSM diagnostic rules vs. top 5.2% based on IRT theta score) overlapped substantially in membership. Some difference was expected because the DSM-5 diagnostic rules assume all items are weighted equally, whilst the IRT model weights items according to their relative prevalence. Interestingly the IRT subgroup had a lower proportion of cases with an unemployed father or depressed parent. Non-overlapping cases were mostly (94%) explained by the parent rating of moderate to severe behavior problems (condition dc2-symptoms interfere). Children with an unemployed father or depressed parent may have been more likely to receive this rating, thus meeting the condition. This bias may indicate our mapped item dc2 is not an ideal indicator of the DSM condition. Moreover, endorsement for the mapping of this item, whilst acceptable, was somewhat mixed amongst expert panel members. These findings illuminate an interesting area for future work.

Our method extends previous work that aimed to identify ADHD in BCS70 (Brassett-Grundy & Butler, 2004) by adhering more closely to the current definition of ADHD, and estimating a more precise dimensional measure. We also built upon the work of Garcia-Barrera et al. (2011) by incorporating a data mining framework, more nuanced modelling technique, and validation through comparisons to mapped reference scales (e.g. SDQ), and epidemiology. Furthermore, we have replicated part of Wall et al. (2015) by re-using the ZIMM model, strengthening their findings, and applying the model to a different psychiatric construct (ADHD).

The present study adds to the literature on IRT models of ADHD, which has primarily focused on evaluating psychometric properties of items (e.g. Arias et al., 2018; Gomez et al., 2011; Li et al., 2015; Polanczyk et al., 2010). Our approach aimed additionally to minimize error and estimate a theta score as precisely as possible, through use of a large non-clinical sample and adjustment for the zero-inflated distribution of symptomatology. Also, building a model within the longitudinal context of the BCS70 provides a previously untapped opportunity for future exploration of a wide range of antecedents to long-term outcomes.

Finally, our method is clearly documented and uses mainstream software, making it easy to replicate or adapt (see Appendix C in the supporting information regarding sharing of data). Thus, in addition to supporting our future work on causal mechanisms in long-term outcomes for ADHD, similar knowledge gains could be pursued by other authors applying our method in existing large datasets with numerous unmeasured psychiatric constructs.

The authors have no conflicts of interest to declare.

References and supporting information

References and supporting information for the published paper are included with the references for the whole thesis and the appendices for chapter 4. A separate reference list for the paper is available in the online version.

Chapter 5 Pilot study of relationship between childhood ADHD and adult outcomes

1 Introduction

Chapter 4 reported on a method to identify new categorical and dimensional measures of ADHD at age 10 in BCS70. These measures opened up an exciting opportunity to exploit the long-running and rich data within BCS70 and further understanding of the effect of ADHD on adult outcomes. The purpose of chapter 5 is to review literature on methods that could be used to estimate robust effects of ADHD on outcomes, test use of those methods, and develop a plan for a subsequent study improved by the learning process. The results of the pilot were also reported in a poster presented at the Society for Research in Child Development (SRCD) conference in March 2019, and a copy of the poster is available on my ResearchGate page[27].

2 Literature review

The literature review section in chapter 5 is comprised of two parts. The first is a review of quasi-experimental methods, starting with some background on what makes them approximate an experiment, and how those features can be approximated in observational data like we have with BCS70. The second part is a review of literature on predictors of ADHD, which were mapped to BCS70 data where possible and used to control for confounding in outcomes analyses.

2.1 Quasi-experimental methods

2.1.1 Why quasi?

When only observational data is available, like we have here with the BCS70, quasi-experimental methods can be used to approximate some of the conditions of an experiment (Angrist & Pischke, 2010; Closer, 2019; Ho et al., 2007; King & Nielsen, 2018; Nichols, 2009). Experimentalists are generally sceptical about this approach (for a discussion see Imai et al., 2008). However, there is a substantial body of literature that makes robust supporting arguments. The key underlying principle is that features which minimise sampling and treatment assignment error (e.g. blocking on confounds, random treatment assignment) can be approximated using observational data (Angrist & Pischke, 2010; Hennekens & Buring, 1987; Ho et al., 2007; Iacus et al., 2011; Imai et al., 2008; King, 2015; King et al., 2017; Nichols,

[27] There were minor errors due to data corruption shown in the poster presented at SRCD. The size and direction of effects were not impacted, and the corrected version is on ResearchGate with a note about the errors.

2007; Pan & Bai, 2018; Rosenbaum & Rubin, 1983; Rubin & Thomas, 2000; Schneider et al., 2007; Stuart, 2010; Winship & Morgan, 1999). These authors and others argue that if an appropriate method is applied carefully, there is a sound theory about underlying mechanisms, and the dataset is large and rich enough, causal effects can be estimated using observational data. However, it is not possible to approximate random treatment assignment for evaluation of outcomes for ADHD. Assignment to the 'treatment' group, or the ADHD subgroup, is not random, but is associated with non-random factors such as socio-economic disadvantage, and parenting practices. Still, balancing the treatment and control groups on confounds is possible, and although effects cannot be inferred to be causal, strength of correlational findings can be increased. Thus, literature on quasi-experimental methods is reviewed next and evaluated for use in the pilot study.

2.1.2 Overview of quasi-experimental methods

Five widely used quasi-experimental methods are summarised in Table 20. Collectively they are referred to as treatment effects methods. The treatment group can be a group that was exposed to an intervention (e.g. medical treatment, educational programme, etc.), as would be the case with an RCT. Alternatively, treatment can be used to refer to membership in any group (e.g. an ADHD subgroup) identifiable in the data, if group membership is hypothesised to have an influence on outcomes (Vignoles & Alcott, 2018).

No.	Method	Key Features	Pros and Cons
1.	OLS	Treatment group is defined as dummy variable, and pre-treatment covariates controlled	Sample size requirements and complexity manageable Assumes linear form, risk of over-estimated effects due to unbalanced treatment and control groups
2.	Matching	Observations dropped or duplicated (weighted) to create quasi-control group similar to treatment group on confounding covariates (like blocking in RCT)	Removes bias from dissimilar participants in control group Inefficiency – some data is lost
3.	DID	Outcome measured before and after treatment for treatment and control groups	Use of two time points controls for fixed effects by participant over time, both X and U[±] Assumes effects of time same for both groups
4.	RD	Assumes a continuous observed variable has a qualitative breakpoint in its distribution that distinguishes (for example) between clinical and non-clinical cases	Isolates difference between treatment and control that is correlated with U Only mimics RCT close to breakpoint/threshold

5.	IV	Variable identified that correlates almost perfectly with treatment but not with outcome directly	Isolates part of difference between treatment and control that is correlated with U Treatment variable can be discrete or continuous Difficult to find credible IVs, large sample sizes needed

Table 20. Quasi-experimental methods for observational data

Abbreviations: OLS = ordinary least squares, DID = difference-in-difference, RD = regression discontinuity, IV = instrumental variable.

±X = observed covariates, U = unobserved covariates

(Nichols, 2007, 2009; Vignoles & Alcott, 2018)

Treatment effects methods assume that it is possible to estimate a counterfactual, i.e. what would have happened to the treatment group if they had not been exposed to the treatment (Rosenbaum & Rubin, 1983). This principle relies on an assumption that all confounding variables relevant to treatment group assignment and the outcome have been observed (X), or they are correlated enough with any unobserved confounds (U) to make systematic error approach zero (Imai et al., 2008; Rosenbaum & Rubin, 1983). Since the data source in the present study (BCS70) is observational panel data and contains a rich array of variables that could account for confounds or correlate with unmeasured confounds, using one or more of the treatment effects methods listed above should be a viable option.

It was noted above that quasi-experimental methods can be used to identify effects of group membership if it is hypothesised to effect outcomes. Rubin argued that manipulation or intervention is necessary to estimate causal effects from matching methods, because this makes the underlying assumptions more likely to be met (Rubin, 1986, p. 962). However, the estimation of counterfactuals and treatment effects have been reported elsewhere in a more general way to determine the effect of exposure to a specific condition, such as coming from a broken home (Boutwell & Beaver, 2010), playing violent computer games (Gunter & Daly, 2012), being a high-school dropout (Vaughn et al., 2011) or teenage mother (Zito, 2018), i.e., even when there was no active intervention. Treatment effects methods has also been used to study ADHD specifically, for example: the relationship between ADHD and child maltreatment (Ouyang et al., 2008), and ADHD vs. ASD performance on executive function tasks (Van Belle et al., 2015). Thus, whilst my ADHD subgroup does not satisfy one of the conditions set out for matching methods by Rubin, (1986), there is precedent for analysing similar constructs with matching, and if carefully applied, should still reduce bias and strengthen correlational findings.

Each of the five quasi-experimental methods was considered. Difference-in-Difference (DID) is not applicable with my data because it can only be used when outcomes are measured both before and after the 'treatment'. A Regression Discontinuity (RD) analysis is not feasible because there are is not a straightforward numeric cut-off point for the ADHD classification as defined using DSM-5 criteria[28]. Regarding Instrumental Variables (IV), no candidate instruments could be identified. OLS Regression and Matching remain, and these are the most widely used. Matching is preferred as a stronger method than OLS because it reduces bias from inclusion of dissimilar controls and is more robust to violations of normality (Gelman & Hill, 2006; King & Nielsen, 2018; Nichols, 2007). Monte Carlo simulations support the reduction in bias compared to OLS, particularly if the pool for selection of controls is large (Rubin, 1979; Rubin & Thomas, 2000). However, matching and OLS or other types of regression adjustments are not mutually exclusive approaches; in fact it is recommended that they are used together: matching first, followed by estimation of effects with multivariate regression, further controlling for confounds (Rubin, 1979; Rubin & Thomas, 2000). Both OLS regression and matching were used here to estimate treatment effects of ADHD on outcomes separately and combined, so comparisons can be made between approaches. Features and procedures of matching methods are discussed next.

2.1.3 Matching

Matching methods facilitate the estimation of treatment effects using a non-parametric pre-processing step (Imai et al., 2008; Rosenbaum & Rubin, 1983; Rubin, 1979). Confounding covariates must first be identified, then a procedure is used to prune observations (delete and/or duplicate) in a way that ensures the treatment and control groups have similar means for each of the matching covariates. If the original sample was random (not the case here for ADHD group), the resulting data approximates that collected from a blocked RCT (Iacus et al., 2011; Rosenbaum & Rubin, 1983; Rubin, 1979; Winship & Morgan, 1999). The method requires that the pool of non-treatment observations is large and comprehensive enough to contain close matches to the treatment group, and adequate matches to cover the covariate strata (Imai et al., 2008).

2.1.3.1 Propensity Score Matching (PSM)

The most widely-used matching method is Propensity Score Matching (PSM) (King & Nielsen, 2018; Pan & Bai, 2018). PSM is estimation of a single score based on multiple covariates (rather than matching on each covariate exactly) which increases the likelihood of finding a similar

---

[28] RD could be evaluated using the continuous ADHD measure, but this is not a clinically validated way of identifying ADHD symptomatology. However, it could still be interesting to model in future research.

match for each member of the treatment group. The score is a probability of membership in the treatment group estimated with a logistic regression on a dichotomous treatment variable (Rosenbaum & Rubin, 1983; Rubin, 1979; Rubin & Thomas, 2000).

$$ps = pr(y = 1|x)$$

Where ps is the propensity score, pr the probability, y the outcome, (in this case the treatment/ADHD subgroup), and x is a vector of covariates. The covariates should be selected to include in the regression based on literature and correlation with the treatment assignment and outcome (Brookhart et al., 2006; Caliendo & Kopeinig, 2005). Controls are then matched to each participant in the treatment group based on proximity of propensity score. Proximity can be determined using the nearest neighbour, and/or within a distance measure[29], such as a caliper (e.g. ¼ SD of the propensity score; Ho et al., 2007; Nichols, 2007; Stuart, 2010). After matching, balance is tested between the treatment and control groups on the matching covariates. If balance is not adequate, the procedure is repeated. T-tests between mean values of groups are often used to assess and confirm adequate balance (Caliendo & Kopeinig, 2005; Rubin & Thomas, 2000).

Recent evidence-based arguments[30] have been made that there are numerous inefficiencies and biases inherent in PSM, and researchers aiming to make causal inferences from observational data should ideally use other methods (King, 2015; King et al., 2017; King & Nielsen, 2018). Simulations and re-analyses of data from previous studies that used PSM have demonstrated that PSM in practice selects matches in a way that is similar to complete randomisation without blocking (King & Nielsen, 2018). This gave markedly inferior results compared to other methods such as Coarsened Exact Matching (CEM) and Multivariate Distance Matching (MDM) methods (King & Nielsen, 2018). The authors of the simulation studies also found that there is a paradox in PSM, where pruning beyond a certain point actually increases imbalance, and they argued that PSM violates the congruence principle, which states that the data space and analysis space should be the same (the PSM score is outside of the data space; King & Nielsen, 2018). Additionally, t-tests used to evaluate balance between matched treatment and control groups can show no group difference when there really is one, because of the inevitable loss of power from pruning/reduction of sample size (Imai et al., 2008).

Other authors responded to earlier arguments and argued that PSM is still a sound method, and problems only occur if assumptions are not met in the data (Pan & Bai, 2018), or if pair

---

[29] For a fuller discussion of PSM matching approaches see Caliendo & Kopeinig (2005).
[30] Albeit the arguments have been made by the same group of authors.

matching without replacement is used, thereby increasing data loss (Jann, 2017). However, in their most recent paper, King & Nielsen (2018) addressed these arguments and maintained that based on their simulations and replications, PSM is overly biased regardless of data characteristics (i.e. assumptions) and specific type of matching used. PSM has been widely used for decades and is familiar to many researchers. However, the arguments made about bias and inefficiency are substantive, so coarsened exact matching was considered here as an alternative approach.

2.1.3.2 Coarsened Exact Matching (CEM)

Coarsened Exact Matching (CEM) is similar to PSM in that both are used to match a treatment and control group in observational data, groups are balanced on relevant observed covariates, and they serve as a pre-processing step before estimation of treatment effects. However, instead of calculating a propensity score based on the probability of being treated given observed covariates, CEM is more similar to exact matching on each covariate. Exact matching can make identifying matches difficult and reduce power. However with CEM, matching can be 'coarsened' by matching on categories that are more broad than those in the original data, thus increasing the probability of matches (Blackwell et al., 2009; Iacus et al., 2011). The coarsening can be custom-defined by the analyst based on knowledge of the construct(s), or based on automatic binning procedures derived from the distribution of the data (Blackwell et al., 2009). CEM is a monotonic imbalance bounding (MIB) method, meaning that balance adjustments to one variable in the covariate set does not affect the others, like they do using a composite propensity score (Blackwell, Iacus, King, & Porro, 2009, p. 524). Once CEM has been applied, a simple mean comparison can be used to evaluate balance (Blackwell et al., 2009; Iacus et al., 2011).

CEM has been recommended as superior to PSM particularly when confounding covariates are not all continuous (Blackwell et al., 2009; Iacus et al., 2011; King & Nielsen, 2018). CEM is easier to understand and use, has fewer assumptions, and reduces model dependence inherent in PSM's iterative post-matching balance adjustments (Blackwell et al., 2009; Iacus et al., 2011; King & Nielsen, 2018).

2.1.4 Estimating a treatment effect using matching methods

Estimating a treatment effect is a multiple step process (Rosenbaum & Rubin, 1983), and Figure 20 shows the six broad-level steps. The purpose of the process is to ensure assumptions required to estimate a treatment effect are met (Harris & Horst, 2016; Iacus et al., 2011; Pan & Bai, 2018; Rosenbaum & Rubin, 1983; Staffa & Zurakowski, 2018). All six steps are used for both PSM and CEM.

Matching process	1. Identify key covariates
	2. Address missingness
	3. Evaluate balance
	4. Match
	5. Re-evaluate balance
	6. Estimate treatment effects

Figure 20. Six steps in estimating treatment effects using matching methods (Harris & Horst, 2016)

As noted previously, the first step is the most important: to identify covariates that are known to have a significant relationship with membership in the treatment group, and with the outcome(s) of interest (Caliendo & Kopeinig, 2005; Garrido et al., 2014; Pan & Bai, 2018; Rajeev & Wahba, 2002; Rosenbaum & Rubin, 1985). Identification of a robust and relevant set of covariates supports all three main assumptions for matching methods: ignorable treatment assignment, (a.k.a. unconfoundedness or strong ignorability), stable unit treatment value assumption (SUTVA), and common support (Caliendo & Kopeinig, 2005; Pan & Bai, 2018; Rosenbaum & Rubin, 1983).

2.1.5 Matching assumptions

The first of the three assumptions, ignorable treatment assignment, requires that there is independence between assignment to the treatment group and the outcome; i.e. any confounders that significantly affect the relationship between the treatment assignment (e.g. ADHD subgroup) and outcome (e.g. health and wellbeing) must be observed, and accounted for in estimations of treatment effects (Rosenbaum & Rubin, 1983; Stuart, 2010). Ignorable treatment assignment is untestable, but supporting evidence can be provided by theory, literature, and sensitivity analysis (Pan & Bai, 2018; Rajeev & Wahba, 2002; Rosenbaum & Rubin, 1983; Sutherland, 2016). There is a trade-off between including too many variables, which reduces the likelihood of achieving balance/common support, and too few, which reduces support for the strong ignorability assumption (Caliendo & Kopeinig, 2005). The covariates should be selected based on theory and from literature review, and can be further reduced to a core set using an iterative or stepwise regression procedure. (Caliendo & Kopeinig, 2005; Pan & Bai, 2018; Rosenbaum & Rubin, 1983; Sutherland, 2016).

The second assumption, SUTVA, requires that when covariates are controlled, the outcome would be the same regardless of how treatment was assigned. Example violations of SUTVA include different versions of treatments that have an effect on the outcome, or participants across groups interacting in a way that affects the outcome (Rubin, 1986, p. 961).

The third assumption is sufficient common support, which should be satisfied by the matching process. Common support requires overlapping distributions for the propensity scores (or matched values) in the treatment and control groups; however it is acceptable for the shape of the functions to differ (Nichols, 2007; Pan & Bai, 2018; Stuart, 2010).

2.1.6 Validity

For treatment effects methods, external validity, or generalisability to other samples, relies on the random selection of participants from a population. It is a limitation in the present study that random selection cannot be relied upon, because selection into the 'treatment' group is defined based on a set of characteristics (ADHD symptoms), which are not random. However, it is not necessarily an objective here to generalise results to the general population, but it is an objective to generalise to other ADHD populations. Thus external validity is reasonable to infer for those with ADHD symptomatology. Internal validity within the sample relies on meeting the assumptions for the quasi-experimental method (Hahs-Vaughn & Onwuegbuzie, 2006).

2.2 Predictors of ADHD

This second section of the literature review comprises a search for significant predictors of ADHD. The is driven by the requirement to identify important predictors of ADHD (the treatment group) so they can be evaluated as potential covariates in the matching process and subsequent regressions.

A literature search was conducted by using the following search terms in Google Scholar: "ADHD", "Attention Deficit Hyperactivity Disorder", "Attention Deficit/Hyperactivity Disorder", "predictor", "risk factor", and "cause", and published since 2000. A short-list of 10 studies was selected by excluding small studies (N < 100), those conducted with only boys or girls, with preference given to reviews and meta-analyses. Abstracts (and full texts if needed) were reviewed for findings of significant predictors of ADHD. Results of the review are summarised in Table 21.

No.	Study	N	Significant predictors

1.	(Hanć et al., 2018)	278	APGAR score, post-term birth, and low birth weight
2.	(Sagiv et al., 2013)	604	Low paternal education, prenatal smoking and drug use, maternal depression, low home quality score, low income, male
3.	(Russell et al., 2014)	19,519	Poverty, housing, maternal education, income, lone parenthood, young motherhood, family conflict (mediates SES)
4.	(Groen-Blokhuis et al., 2014)	N=19,210 (GWAS)	ADHD polygenic risk score
5.	(Touchette et al., 2007)	N=1,492	Early short sleep durations (<41 mos)
6.	(Russell et al., 2015)	N=8,132	Mediators – adversity and parental involvement
7.	(Thapar et al., 2013)	Review (last 15 yrs)	No single factor, risks related to extreme early adversity, pre and post-natal exposure to lead, low birth weight and pre-maturity
8.	(Silva et al., 2014)	N=43,062	Young, single, smoking mothers, induced labor, threatened or pre-term labor, pre-eclampsia, UTI, early delivery - both genders
9.	(Thapar et al., 2009)	N=815	Smoking not a predictor in unrelated mother and child (from fertility clinic data)
10.	(Chronis et al., 2007)	N=108	Maternal depression increases and positive parenting decreases

Table 21. Recent studies reporting predictors of ADHD

Each predictor identified through this search was be evaluated for use in matching and regressions, and the process was documented in the methods section of chapter 5.

The matching approach was assessed in a pilot, which is discussed next in a Methods, Results, and Discussion format.

3 Method

3.1 Data

The BCS70 age 42 sweep data was used to measure adult outcomes. The sweep was conducted in 2012 using face-to-face interviews, questionnaires, and skills assessments, N=9,841. An overall description of all BCS70 sweeps along with background on non-response can be found in chapter 2.

A review of attrition in BCS70 sweeps up to age 34 reported that male and low SES cohort members were more likely to attrit (Ketende et al., 2010). ADHD is associated both with being male (Willcutt, 2012) and low SES (Russell et al., 2015). Thus the larger attrition rate for the ADHD group is not surprising. Given the associated risk of bias, I evaluated representativeness the pilot sample using categorisations by sex, low SES (age 10 free school meals indicator), and ADHD severity score. Comparisons between the age 10 and age 42 samples, overall and for the ADHD subgroup, are summarised in Table 22.

	Full samples			ADHD sub-samples		
	S1(10)	S2(42)	RRR	S3(10)	S4(42)	RRR
N	11,426	7,242	(or t/p)*	594	298	(or t/p)*
Boys	51.5	47.7	1.08	69.9	63.1	1.11
Low SES	15.3	13.4	1.14	27.5	25.8	1.06
ADHD Severity	-.06	-.13	5.2/<.001	1.81	1.80	.49/.62

Table 22. Comparison of representativeness by sex, SES, and ADHD severity for the age 10 samples, and for samples with data from both age 10 and 42

RRR = Relative Risk Ratio, S1 = full age 10 sample scored for ADHD, S2 = age 10 sample with data available at age 42, S3 = Full ADHD subgroup from age 10, S4 = ADHD subgroup from age 10 with data available at age 42
** For the continuous ADHDness variable, means were compared using a two-sample t-test so the t statistic and p values are reported instead of RRR.*

The relative risk ratios (RRRs) were too small to be considered important; a rule of thumb for a minimum cut-off indicating a small and substantial effect for a non-rare event is RRR > 1.32 (Olivier et al., 2017). Also, within the two ADHD subsamples, the mean ADHD severity measure was not significantly different. It is helpful that the available sample is similar to the original on ADHD severity. However, the mean ADHD severity did differ significantly between the full age 10 sample (N=11,426) and the one limited by available outcomes at age 42 (N=7,242). For the pilot, these differences are stated as caveats, but no specific remedy implemented. Attrition-related bias is revisited in chapter 6.

3.2 Measures

3.2.1 ADHD

The ADHD measures used here are the same as those reported in Chapter 4. One is an ADHD subgroup indicator, which is a binary/categorical variable, and was assigned using an adapted version of the DSM-5 diagnostic criteria applied to a 16-item scale. The value of '1' is associated with meeting the ADHD criteria. The second measure indicates ADHD severity; it is a continuous variable and was estimated using a zero-inflated item response mixture model of the 16-item scale. Both ADHD measures had 0% missing within the age-10 behaviour dataset (N=11,426).

3.2.2 Health and wellbeing

As discussed in chapter 3, wellbeing has been defined by the ONS and WHO as a multidimensional construct with subjective and objective aspects. dimensions, including personal wellbeing, health, relationships, and 'what we do'. In the pilot, I wanted to test the idea of creating a composite measure, taking into account both subjective and objective aspects, and separate from socioeconomic status, to provide a broader view of functioning (i.e. what if a person has anxiety and a lower SES, but otherwise functions well?), and also to minimise bias from multiple tests/comparisons. I selected six related measures to represent non-SES aspects of both subjective and objective wellbeing in a single score. The six components were: the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) score, an abbreviated Rutter Malaise scale score, a self-rated life satisfaction score, a self-rated general health score, a binary indicator for living with a partner, and a binary indicator for 'working' (working part-time or full-time). Each is described in the paragraphs that follow.

The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) was administered at age 42 in BCS70. It is a measure of subjective psychological wellbeing, and has been validated in several clinical and non-clinical samples (e.g. Bass, Dawkin, Muncer, Vigurs, & Bostock, 2016; Smith, Alves, Knapstad, Haug, & Aarø, 2017; Tennant et al., 2007). The 14-item version of the scale was used. Response levels ranged from 1-5; none of the time, rarely, some of the time, often, all of the time (Centre for Longitudinal Studies: UCL/IoE, 2012). The statements were worded in a positive way, e.g. "I've been feeling confident" and framed in the context of the past two weeks. Higher scores indicate higher levels of well-being. The scores ranged from 14-70, and were stored in the BCS70 variable BD9WEMWB (Brown & Hancock, 2014). The variable was recoded ('not enough information' set to missing) and renamed O42WarwickWB. The variable had significant missingness (17.4%) in the matched age 10-42 subsample (N=7,242). Including other indicators in a single composite measure should help compensate for the missingness.

The abbreviated 9-item Rutter Malaise Inventory score (Rutter et al., 1970) was the second indicator of subjective wellbeing included from the age 42 sweep. The scale was developed to measure psychological distress or depression. Low malaise scores have been used elsewhere to indicate wellbeing in BCS70 (Sacker & Cable, 2006; Sigle-Rushton et al., 2005; Steptoe & Butler, 1996; White et al., 2012). The nine items were administered as a self-completion questionnaire to cohort members at age 42, and responses were recorded in variable BD9MAL. Example items included: 'do you feel tired most of the time?' and 'do you often feel miserable and depressed?'. Scores ranged from 0-9, with higher scores indicating higher levels of depression/malaise, and inversely, lower scores indicating higher levels of wellbeing. The variable was recoded ('not enough information' values set to missing) and renamed O42Mal. It had moderate-high missingness (12.3%) in the matched age 10-42 subsample (N=7,242).

The third component was a life satisfaction rating, which has also been used elsewhere to measure subjective wellbeing (Layard et al., 2014; Schoon & Kneale, 2013). Cohort members (CMs) were asked to rate from 0-10 how satisfied they were with how life has turned out so far, with 0 meaning completely dissatisfied, and 10 meaning completely satisfied. The scores for variable B9LIFST1 were recoded ('not enough information' recoded to missing) and renamed O42Life, with a range of values from 0-10. Missingness was minimal at 1.12%.

Fourth, a self-reported general health score was included with five response levels: poor, fair, good, very good, and excellent, numbered from 1-5, where 1 was excellent and 5 was poor. The original variable was B9HLTHGN, responses of refused, not applicable, and don't know, were recoded as missing, and a new variable created: O42GHlth. The general health rating had almost no missing observations (0.19%, or 14 CMs).

The fifth component tested as part of this measure was an indicator of whether or not the cohort member (CM) was working at the time of the age 42 sweep. This served as a rough measure of the ONS wellbeing dimension 'what we do'. The variable BD9ECACT in BCS70 provided categorisations of economic activity for cohort members. A new binary variable was created called O42Working and coded as 1 if the CM was working full-time or part-time (employed or self-employed, codes 1-4), and 0 otherwise, including CMs in education or training (only 33 CMs). This variable had few missing observations (0.39%, or 28 CMs).

Finally, an indicator of whether or not the cohort member was living with a partner at the time of the sweep was used as a measure for the ONS wellbeing dimension 'relationships'. A derived BCS70 four-level variable called BD9PARTP contained four-level data indicating whether a CM did not report a partner, or lived with a spouse, civil partner, or cohabiting partner. A new variable O42LwPart was created and recoded to 1 if the CM was reported as living with a spouse, civil partner or cohabiting partner, and 0 otherwise. This variable had no (0%) missing data.

A factor analysis model on these six indicators was significant (χ^2 (15) = 6703.89, $p < .001$). The scree plot supported a single Health and Wellbeing (HWB) factor (eigenvalue of first factor = 1.81, no others close to 1.0). However, the uniqueness was relatively high (> 0.70) and loadings on the first factor were fairly low for general health (.50), living with partner (.29), and working (.28), so the fit was questionable. However, I decided to proceed with the HWB factor modelled in this way for the pilot, and revisit alternatives in chapter 6.

```
sem (HWB -> O42Life, O42Working, O42WarwickWB, O42GHlth,
O42LwPart, O42MalR), method(mlmv) latent(HWB) nocapslatent
```
[31]

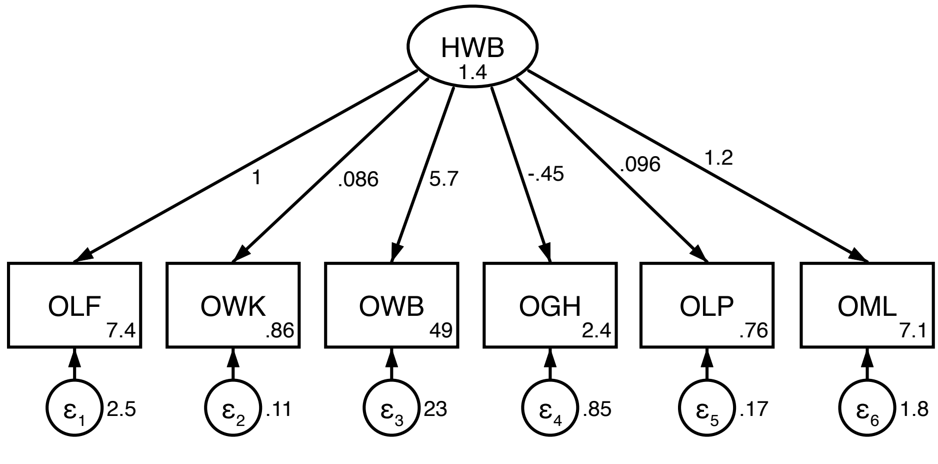


Figure 21. Factor analysis model of six wellbeing indicators loading on a single factor of Health and Wellbeing (HWB)
OLF = Outcome Life Satisfaction, OWK = Working, OWB = WEMWBS Wellbeing, OGH = General Health, OLP = Living with Partner, OML = Malaise (score reversed)
N.B. The mlmv option was used to handle missing data, which is Stata's implementation of FIML.

3.2.3 Educational attainment: a proxy for SES as a measure of objective wellbeing

The second outcome measured was educational level (EDL). The BCS70 variable BD9HACHQ values ranged from 0 (no academic qualifications) to 8 (higher degree). The variable did not contain any values that needed to be recoded as missing, but was copied to a new variable O42Educ, to maintain the 'O42' naming convention for outcome variables in these analyses. This variable had no (0%) missing data.

3.2.4 Social class of job: a proxy for SES as a measure of objective wellbeing

A third outcome evaluated in this pilot was a classification of the cohort member's job. There were multiple classification schemes for occupations reported in BCS70, because the ONS changed them over the course of time the study has been active. For simplicity and consistency with the earlier data (ages 0, 5, and 10) the social class reported here was based

[31] All Stata code was formatted using `this non-proportional font.`

on the scheme in effect in 1970, which comprised six classifications of jobs: 1-professional, 2-managerial and technical, 3-non-manual, 3-manual, 4-partly-skilled, and 5-unskilled. The variable from BCS70 was B9CSC, and non-missing responses of not enough info to code, not applicable, and 'others' were recoded as missing in a new variable called O42Social. 'Others' is technically non-missing, but it is not known where 'others' would fit on an ordinal scale of one to six, and there were only 23 CMs classified this way, so they were coded as missing. The O42Social variable had fairly high missingness (15.2%).

N.B. Income data was briefly considered as an SES measure, but it is contentious for respondents and accordingly at age 42 had extensive (>80%) missingness.

3.2.5 Family-wise error rate (FWER)

Testing three outcomes: HWB, Educational Attainment, and Social Class allowed for a significance threshold of p<0.017 (0.05/3).

3.2.6 Covariates

3.2.6.1 Mapping of variables to BCS70

Abstracts and full text of the 10 studies selected in the literature review section of chapter 5 were reviewed, and 30 unique constructs were identified as significant predictors of ADHD. I attempted to map all of them to BCS70 data either from age 0 or age 5 sweeps. I used the sweeps prior to age 10 because variables used in matching procedures should not be related to assignment to the treatment group (ADHD subgroup), so should be fixed in time or collected before assignment to the group (Caliendo & Kopeinig, 2005)[32]. It was not possible to accurately map the BCS70 age 0 or age 5 items to the following: APGAR score, prenatal drug use, family conflict, adversity, low income, parental involvement, ADHD polygenic risk score, exposure to lead, early differences in orienting and attention, early short sleep durations, threatened labour, and UTI during pregnancy. The 17 remaining predictors of ADHD were mapped, cleaned, recoded into new variables (Table 23) and merged into a chapter 5 working dataset.

No.	Item	Age	BCS70 variable(s)	Mapping notes
1.	Post-term birth	0	A0159b	F0PostTerm (0/1) gestational age >41 weeks
2.	Low birth weight	0	A0278	F0LBW (0/1) birth weight < 2500g

[32] In my data, ADHD group assignment was based on teacher and parent ratings of child behaviour at age 10.

3.	Low paternal education	5	E189b	F5DadEd (0-5) father education level (no qualifications/vocational or other/o-levels/a-levels/certified nurse or teacher/degree)
4.	Prenatal/maternal smoking	0	A0043b	F0PregSmoke (0/1)
5.	Maternal prenatal smoking level	0	A0043b	F0SmokeLevel (0/1/2/3, 0, 1-4, 5-15, >15 per day)
6.	Maternal depression	5	D122a	F5MumMal (0/1) Mother Rutter Malaise score indicated psychological problems, or not
7.	Low home quality	5	E264, e265	F5Home LS (0/1) Health visitor rated home furnishings low/v low standard F5HomeUntidy (0/1) HV rated home untidy or chaotic
8.	Sex (male)	10	Sex10	Assumed unchanged since birth
9.	Poverty, low income, poor housing	5	E267b	F5PoorNbhd (0/1), Home visitor rated house in 'poor neighbourhood'
10.	Maternal education	5	E189a	F5MumEd (0-5) mother education level (see F5DadEd for scale)
11.	Lone parenthood/single mother	0	A0012	F0Unmarried (0/1) marital status at birth = not married
12.	Premature/pre-term birth	0	A0195b	F0PreTerm (0/1) gestational age <37 weeks
13.	Young mother	0	BD1MAGE	F0MumAge (actual age)
14.	Induced labour	0	A0245	F0Induced (0/1) labour induced
15.	Pre-eclampsia	0	A0226, A0227, A0228	F0PreEInd (0/1) if proteinuria, oedema, or pre-eclamptic fits reported
16.	Child intelligence	5	E268	F5HVOCIQ (0/1) Home visitor opinion of child's intelligence being 'backward' or not
17.	Positive parenting	5	D124g	F5AuthCRV (0/1) Authoritarian parenting indicated (reverse)

Table 23. Mapping of ADHD predictors from literature to BCS70 items

N.B. cut-off points for low birth weight, pre-term, and post-term, and indicators of pre-eclampsia from NHS, (2018).

As discussed previously, guidance in literature advises that when selecting covariates for matching, there are disadvantages to including too few or too many (Caliendo & Kopeinig, 2005). In order to select an optimal list of the most important factors, iterative logistic regressions were modelled to test the relationships between the 17 candidate predictors, membership in the ADHD subgroup, and the outcomes of interest.

3.2.6.2 Iterative regressions to select matching covariates

First, univariate logistic regressions were modelled for each variable separately as predictors of the ADHD subgroup (treatment). The following variables were not significant at α=0.05: pre-term birth, post-term birth, induced labour, low birth weight, pre-eclampsia, and authoritarian child rearing views. Some of these were surprising given other findings of associations between the factors and ADHD, in particular low birth weight and premature birth (Silva et al., 2014; Thapar et al., 2013). Next, a multivariate logistic regression was modelled with all the variables. The same variables that were not significant individually, were also not significant within the multivariate model. Additional variables became clearly not significant (p > 0.20) in the multivariate set: mother's education level, untidy home, mother's age at birth, and unmarried mother. The second smoking variable, indicating the amount smoked, remained just significant at the 95% level (p = 0.05). A new regression was modelled with the clearly insignificant (p > 0.20) variables removed, which left sex, father education level, low-standard home, mother malaise, smoked during pregnancy, amount smoked, unmarried mother, and home visitor rated child's intelligence as 'backward', as follows:

```
logistic adhd_sg sex10b F5DadEd F5HomeLS F5PoorNbhd F5MumMal
F0PregSmoke F0SmokeLevel F5HVOCIQ
```

In this model, poor neighbourhood became clearly insignificant. The binary smoking indicator was again insignificant, whilst the smoking level indicator was significant. The regression was run again without poor neighbourhood and the smoking indicator.

```
logistic adhd_sg sex10b F5DadEd F5HomeLS F5MumMal F0SmokeLevel
F5HVOCIQ
```

This combination allowed for N=5,077 to be included in the model (listwise deletion). All predictors had p values < 0.05 except F5HomeLS (home low standard, p = 0.06), and the overall model chi-square test was significant (χ^2 (6) = 145.91, p < 0.001).

The variables previously removed were then added back in one at a time, to see if significance values changed. Using this process, significance was achieved for two more variables: unmarried mother, and induced labour.

This set of variables appears to be a good set of predictors for ADHD based on literature, data available in the BCS70, and statistically significant relationships. However, the literature on PSM suggests that predictors should only be included in the prediction of a propensity score if they are also predictors of the outcomes in question (Caliendo & Kopeinig, 2005). So, additional univariate regressions were used to test the relationship of each of these predictors with the three outcomes that are part of the planned analysis: health and wellbeing, education level, and social class. Induced labour was consistently not a significant predictor (α=0.05) of any outcome, so was removed.

Based on this multi-stage iterative regression analysis, the predictor set to be used for running a matching procedure to create a quasi-control group is set out in Table 24.

Variable name	Description	Coding
Sex10b	CM1 sex	0=Female, 1=Male
F5DadEd	Father's education level, @CM age 5	0=no qualifications, 1-other, 2=O level, 3=A level, 4=Nurse or Teacher, 5=Degree
F5HomeLS	Home visitor reported home furnishings as 'low standard' @CM age 5	0=No, 1=Yes
F5MumMal	Mother self-report Rutter Malaise questionnaire high score indicating psychological problems @CM age 5	0=No, 1=Yes
F0SmokeLevel	Mother self-reported number of cigarettes smoked per day during pregnancy with CM (age 0)	0=0, 1=1-4, 2=5-15, 3=15+
F0Unmarried	Mother self-reported marital status unmarried at CM birth (age 0)	0=No, 1=Yes
F5HVOCIQ	Health visitor reported opinion of CM development progress as 'backward' @CM age 5	0=No, 1=Yes

Table 24. Covariates to be used in matching procedures

[1] *CM = cohort member*

The sex variable had no missingness, but the others had moderate to high levels (9.4 to 29.5%). Accordingly, a missing data analysis was conducted to identify an appropriate strategy to mitigate bias due to missingness.

3.2.6.3 Missing data analysis on matching covariates

A description of missing data patterns showed there were four patterns that described 92% of the data. The other patterns each accounted for 2% or less of observations, so those were not evaluated specifically. The top patterns are shown in Figure 22.

```
misstable patterns F5DadEd F5HomeLS F5MumMal F5HVOCIQ F0Unmarried F0SmokeLevel

      Missing-value patterns
        (1 means complete)

              |   Pattern
    Percent   |  1  2  3  4    5  6
  ------------+-----------------------
      70%     |  1  1  1  1    1  1
              |
      11      |  1  1  0  0    0  0
       7      |  1  1  1  1    1  0
       4      |  0  0  0  0    0  0
       2      |  1  1  1  1    0  1
       2      |  0  0  1  1    1  1
       1      |  1  1  0  1    1  1
      <1      |  1  1  1  0    0  1

  Variables are  (1) F0Unmarried  (2) F0SmokeLevel  (3) F5MumMal  (4) F5HomeLS  (5) F5HVOCIQ
                 (6) F5DadEd
```

Figure 22. Missing value patterns for six ADHD predictor variables (pilot)
N.B. This figure is a partial screen-grab of a larger table with several more rows of <1%

The first pattern of missingness showed that 11% had values for the first two variables: F0SmokeLevel and F0Unmarried from the age 0 sweep, but none of the variables from the age 5 sweep. These were probably non-responses for the age 5 sweep (but they did respond at age 10). The second pattern, with 7% missing, had values for all variables other than the father's education level at age 5, and for the third pattern, 4% were missing for all predictor variables. The most likely respondent for the survey at these younger ages is the cohort member's mother, so the missing father's education level may indicate an absent or partially absent father. The 4% missing pattern, where all variables are missing at ages 0 and 5 (but there is data at age 10 and 42), could be cohort members who were recruited after age 5, as some recruitment was done after age 5 in BCS70 to try and account for immigration into the UK. The 7% and 4% missing patterns could also be the result of a systematic bias because people with specific characteristics avoid answering specific types of questions.

In order to look at possible systematic bias in father's education level specifically, a new variable based on F5DadEd missingness was created, and named F5DE_m (observed = 0, missing = 1). A logistic regression with the ADHD subgroup indicator as the dependent variable and F5DE_m as the predictor indicated a significant relationship between missingness of father's education level and membership in the ADHD subgroup (N=8,519) χ^2 (1) = 8.85, p= 0.003). Next, logistic regressions were run to test the other predictors in the analysis for a significant relationship with father's education level missingness (F5DE_m).

Cohort member sex was not a significant predictor, but each of the other variables were when tested individually (unmarried mother at birth, smoking, mother malaise, low standard home, and child development backward). When combined into a single multivariate regression, only unmarried, smoking, and mother malaise remained as significant. Since the other observed variables in the analysis were related to father's education level, and could be used to estimate the missing values, it is reasonable to assume Missing at Random (MAR) (Baraldi & Enders, 2010).

The MAR assumption suggests it is possible to estimate missing values using an imputation method. Method choices are single imputation (SI), multiple imputation (MI) or Full Information Maximum Likelihood (FIML). Single imputation is usually replacement of all missings with a mean value, and is normally only advisable when missingness on a variable is less than 2%. This is because with larger proportions of missingness, considerable bias is introduced by the inflated number of mean values and results in artificially lowered standard errors (Sainani, 2015). However, in the context of implementing matching procedures, recent simulations comparing SI and MI found that SI performs acceptably, as long as covariate balance is achieved in the matched samples (for a discussion, see Leite, Stapleton, & Bettini, 2018). Thus, single mean imputation was used here in the pilot study, and a more robust method was implemented in the final study.

To implement the single imputation, a mean was calculated for each variable with missingness, separately by ADHD subgroup. As an additional measure to allow models to control for missingness, new dummy variables were created to indicate which observations had missingness in each covariate, as follows: father's education level (F5DE_m), home low standard (F5HLS_m), mother malaise (F5MM_m), mother's smoking level during pregnancy (F5SL_m), mother unmarried at birth (F0UM_m) and health visitor reported 'backward' development (F5HI_m). Finally, the covariate missings were updated with calculated mean values, as described in Table 25. A separate version of the working data file was saved with these singly imputed mean values, to preserve the original data with missingness for use in other analyses.

	ADHD (1)	Non-ADHD (0)
F5DadEd	.990	1.56
F5HomeLS	.086	.028
F5MumMal	.404	.210
F5HVOCIQ	.114	.027
F0SmokeLevel	1.14	.819

F0Unmarried	.097	.053

Table 25. Means used by ADHD subgroup for single imputation of missing data

This competes the measures needed for the analysis of outcomes, including ADHD, outcomes, and covariates.

3.2.7 Procedures

3.2.7.1 Naïve regression

As a first step, separate naïve univariate regressions were used to model the relationship between the two ADHD measures and the three outcomes. In this context naïve refers to the fact that the sample is unmatched, i.e. probably imbalanced across treatment and control groups. The models also did not control for confounding. The dataset used for the naïve regressions contained the single (mean) imputed values, but these variables were not used in any of these models, so they had no effect on the results.

```
Dataset name: Chapter5Work_cem_pilot
```

The HWB outcome is continuous, so simple Ordinary Least Squares (OLS) regression was used:

$$y_{i=} \beta_0 + \beta_1 x_1 + \varepsilon$$

```
regress HWB_FS adhd_sg, beta[33]
regress HWB_FS ADHDness, beta
```

Educational Attainment and Social Class have ordered levels (nine and six, respectively), indicating ordinal measures. However, the proportional odds (or parallel lines) assumption for ologit was not met, i.e. each level of outcome does not have the same slope (Muthén & Schultzberg, 2017). Here for simplicity and comparability I used OLS regression in Stata (rather than ologit or a more robust procedure) to estimate naïve treatment effects and note this as a limitation of the pilot that was addressed in subsequent chapters.

```
regress O42Educ adhd_sg, beta
regress O42Educ ADHDness, beta
regress O42Social adhd_sg, beta
regress O42Social ADHDness, beta
```

Next, effects were evaluated based on matched samples.

[33] The beta option produces standardised coefficients in the output

3.2.7.2 Coarsened Exact Matching (CEM)

The covariates identified through iterative regressions were used with CEM procedures for matching. They were sex, father's education level, low-standard home, mother malaise, mother smoking level, mother unmarried at birth and health visitor rated child's development as 'backward'.

```
sex10b F5DadEd F5HomeLS F5MumMal F0SmokeLevel F0Unmarried
F5HVOCIQ
```

Coarsening is intended to increase the probability of matches for the treatment group, but when the categorical data only have two or a few categories, coarsening may not improve the chances of a match (Blackwell et al., 2009). So, although the method is called coarsened exact matching, the coarsening is not necessarily needed. The categories within each variable are outlined in Table 26. Only two could be coarsened; F5DadEd and F0SMokeLevel, because the others were already binary. The former was reduced to coarsened categories as shown, based on my assessment of categories that were similar to each other.

Variable	Description	Categories	Coarsened Categories
F5DadEd	Father's education level at CM age 5	0-no qualifications	0=0
		1-vocational	1=(1,2,3)
		2-O level	2=(4,5)
		3-A level	
		4-nurse or teacher	
		5-degree	
F0SmokeLevel	No. of cigarettes smoked per day during pregnancy	0-0	0=(0,1)
		1-(1-4)	1=(2,3)
		2-(4-15)	
		3-(15+)	

Table 26. Coarsened covariates

The dataset used for this procedure was Chapter5Work_cem_pilot.dta. The cem procedure (Iacus et al., 2014) was coded with the two coarsened variables (the others were not coarsened, as denoted by the (#0) after the variable name), the option to match each treatment observation to only one control (k2k option), and to treat missing values as a category (matches missing to missing) (Iacus et al., 2014). Because of the last option (a default), imputation of missing variables was not required. This is a relative advantage over other procedures (e.g. PSM) given the high missingness in the variables.

```
cem sex10b(#0) F5DadEd(0 3 5) F5HomeLS(#0) F5MumMal(#0)
F0SmokeLevel(1 3) F0Unmarried(#0) F5HVOCIQ(#0),
treatment(adhd_sg) k2k
```

The distance statistic was zero[34] for all but two of the predictors: F5DadEd and F0SmokeLevel, the two that were coarsened. The multivariate balance statistic was not close to zero (L1 = 0.30), which indicates the matching was not successful overall (Blackwell et al., 2009). Using the k2k option resulted in 281 treatment cases (17 were dropped) and 281 controls.

Given this was not an ideal match, the model was run again without the k2k option.

```
cem sex10b(#0) F5DadEd(0 3 5) F5HomeLS(#0) F5MumMal(#0)
F0SmokeLevel(1 3) F0Unmarried(#0) F5HVOCIQ(#0),
treatment(adhd_sg)
```

This model matched 283 from the ADHD (treatment) subgroup to 6,523 controls. All matched observations were assigned weights in the cem_weights variable, to compensate for different strata sizes (Blackwell et al., 2009). The multivariate balance was not much improved (L1 = 0.28), so the model was re-configured with no coarsening for any of the variables. The weights complicate use of the data slightly, so the k2k (1 to 1 match) option was added back.

```
cem sex10b(#0) F5DadEd(#0) F5HomeLS(#0) F5MumMal(#0)
F0SmokeLevel(#0) F0Unmarried(#0) F5HVOCIQ(#0),
treatment(adhd_sg) k2k
```

This third model resulted in n = 273 for both treatment and controls (25 treatment observations were dropped, total n = 546) and had an ideal multivariate balance (L1 = 0). Per G*Power software (Faul et al., 2009), this sample had the power to detect an effect size of about 0.3, assuming a one-tail test and 95% confidence level.

Since this last CEM model had good balance and adequate power to detect a relatively small effect, it was selected to evaluate the treatment effect of ADHD on the three outcomes. The sample average treatment effects (SATT) can be estimated simply with a regression of the treatment group on the outcome, where cem_matched is true (1).

```
regress HWB_FS adhd_sg if cem_matched==1
regress O42Educ adhd_sg if cem_matched==1
regress O42SocialR adhd_sg if cem_matched==1
```

There was no missing data for the HWB and EL outcome variables, but 21% were missing SC. They were deleted listwise in the regression.

[34] Zero is the ideal value for the distance statistic (Blackwell et al., 2009)

The regress command denotes OLS regression, which assumes a continuous outcome variable. The HWB_FS is continuous, but O42Educ and O42Social are not. I used linear regressions here because the direction and significance of the effect were the same as those found with ordinal logistic regressions, and also for consistency with the naïve regressions so comparisons can be made. Models that better accommodate the distribution of the outcome variables were developed for the final studies in chapter 6.

Finally, multivariate regressions were also modelled with the CEM-matched sub-sample including both the ADHD subgroup and ADHD severity score, based on the formula:

$$y_{i=} \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_n x_n + \varepsilon$$

```
regress HWB_FS adhd_sg ADHDness if cem_matched==1
regress O42Educ adhd_sg ADHDness if cem_matched==1
regress O42SocialR adhd_sg ADHDness if cem_matched==1
```

4 Results

4.1 Naïve regression

This section reports results for separate univariate regressions predicting each of the three outcomes. Missing values in these models were deleted listwise.

Health and Wellbeing (HWB)

	β	B	SE(B)	Adj R^2	BIC
HWB					
ADHD	-0.08	-.40***	.06	0.01	21291.76
Severity	-0.16	-.19***	.01	0.03	21137.44

Table 27. Univariate regressions showing the relationship between ADHD measures at age 10 and Health and Wellbeing (HWB) measure at age 42

**** $p < .001$*
N =7,242

Both coefficients were significant, with membership in the ADHD subgroup predicting a 0.08 SD lower Health and Wellbeing score, and a 1 SD increase in ADHD severity predicting a 0.16 SD drop in the HWB score. Both models explained very little of the variance in HWB, but ADHD severity explained more than ADHD subgroup membership, and the smaller Bayesian Information Criterion (BIC) indicated the ADHD severity model was a better fit.

Education Level

	β	B	SE(B)	Adj R^2	BIC

EL					
ADHD	-0.11	-1.56***	0.16	0.01	35356.13
Severity	-0.27	-0.85***	0.04	0.07	34897.78

Table 28. Univariate regressions showing the relationship between ADHD measures at age 10 and Education Level (EL) measure at age 42

**** p < .001*
N = 7,242

Both coefficients were significant at the 99.9% confidence level. The model with ADHD severity as a predictor had a smaller BIC indicating a better fit. Membership in the ADHD subgroup was associated with a 0.11 SD lower Education Level, and 1 SD higher in ADHD severity was associated with a 0.27 SD lower Education Level. $R^2 = 0.01$ indicated the variance explained by membership in the ADHD subgroup was not practically important (Ferguson, 2009), but $R^2 = 0.07$ indicated a small and practically important effect for ADHD severity.

Social Class of Job

	β	B	SE(B)	Adj R²	BIC
SC					
ADHD	-0.08	-0.37***	0.06	0.01	15379.32
Severity	-0.18	-0.18***	0.01	0.03	15210.91

Table 29. Univariate regressions showing the relationship between ADHD measures at age 10 and Social Class of job (SC) measure at age 42

**** p < .001*
N = 6,140; SC variable based on CM job at age 42 had 15% missingness
The social class variable was reversed (O42SocialR) so higher values were associated with higher social class

Both models were significant at the 99.9% confidence level, and the smaller BIC indicated the model with ADHD severity as a predictor of Social Class (reversed) had a better fit. Membership in the ADHD subgroup was associated with a 0.08 SD lower Social Class of job, and 1 SD higher in ADHD severity was associated with a 0.18 SD lower Social Class. $R^2 = 0.01$ and 0.03 indicated that neither the variance explained by membership in the ADHD subgroup nor ADHD severity were practically important.

4.2 Coarsened exact matching (CEM)

OLS regressions of ADHD subgroup membership on outcomes using the matched observations identified in the selected CEM procedure produced the estimated treatment effects are reported in Table 30.

	β	B	SE	Adj R²	R	d

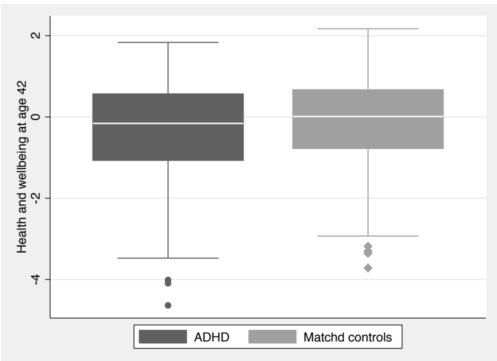
HWB	-0.10	-0.24*	0.10	0.01	0.09	0.18
Education level (EL)	-0.23	-1.20***	0.22	0.05	0.23	0.47
Social class (SC)	-0.21	-0.36***	0.08	0.04	0.20	0.41

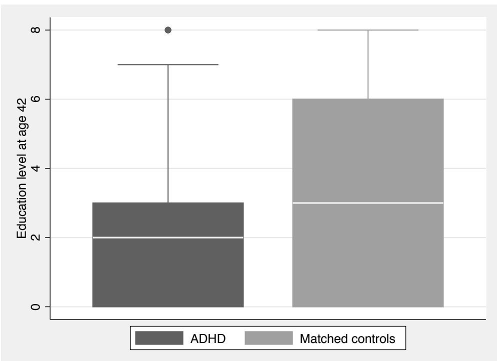
Table 30. Sample average treatment effect of ADHD on outcomes, based on CEM model

**** $p < .001$*
N.B. Cohen's d was derived from adjusted R^2 and R values using an online effect size calculator (Lenhard & Lenhard, 2016)

The effect of ADHD subgroup membership was significant on all three outcomes in the matched sample at the adjusted confidence level (p > 0.017). Adjusted R^2 (0.01) for HWB was too small to be practically important, but could be interpreted as small and important for EL (*Adj* $R^2 = 0.05$) and on the border of importance for SC (*Adj* $R^2 = 0.04$). The sample sizes were large enough to detect a Cohen's d effect size of 0.30, so had the power to detect the effects shown here for both EL at $d = 0.47$ and SC at $d = 0.41$, but not HWB at $d = 0.18$.

Distributions were examined visually using box plots.





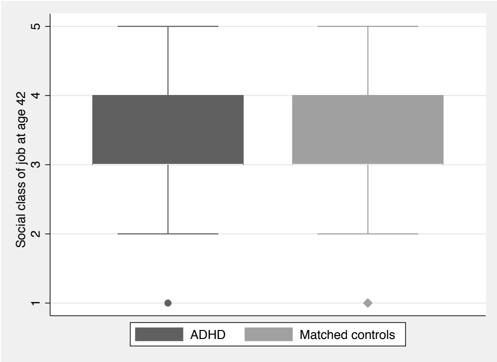


Figure 23. Box plots of CEM-matched ADHD and controls on outcomes

The box plots showed that distributions for the ADHD and matched control groups were similar for HWB and SC. There was significant overlap between the groups for EL, but the range was markedly different, with lower education levels for the middle 50% of the ADHD group.

4.3 CEM matched sample models with ADHD subgroup indicator and severity score as predictors

When the ADHD severity score was added as a second predictor to the regression models for all three outcomes, the ADHD subgroup indicator became insignificant for all three and switched signs for HWB and EL. The severity score was negative and significant for all three outcomes, Adjusted R^2 suggested that a small but important amount of variance was explained for EL and SC, but this should be interpreted with caution since OLS was not the optimal form of regression to use for the EL and SC variables.

	β ADHD	B ADHD	β Sev	B Sev	Adj R^2
HWB	0.13	0.30	-0.28	-0.28***	0.03
Education level (EL)	0.06	0.33	-0.34	-0.78***	0.08
Social class (SC)	-0.01	-0.01	-0.23	-0.17**	0.05

Table 31. Treatment effect of ADHD on outcomes, based on CEM matched sample (N = 546)

*** p < .01, *** p < .001*

The change in direction for two outcomes and significance of effects for all three indicates a collinearity problem and this should be explored further and adjusted for in subsequent analyses.

5 Discussion and conclusion

5.1 Pilot study findings

Naïve univariate regressions on a preferred sample (N=7,242) indicated that belonging to the ADHD subgroup (n=298) and (separately) ADHD severity score as measured in chapter 4, corresponded to significantly worse outcomes at age 42 on all three outcomes: a composite Health and Wellbeing measure, Education Level, and Social Class of job. The significance held at an adjusted alpha level 0.017 (0.05/3). Adjusted R^2 values indicated that only the variance explained by ADHD severity on Education Level ($Adj\ R^2 = 0.07$) could be considered small but practically important; none of the others were large enough to conclude importance. However, OLS was not the optimal regression procedure for the ordinal variables, so the effects should be interpreted with caution. The results of these models support the findings in other literature (see chapter 3) for associations between ADHD and worse long-term outcomes (Brassett-Grundy & Butler, 2004; Erskine et al., 2016).

In the selected CEM exact-matched sample ($N = 546$), the treatment and control groups were balanced on covariates selected through literature review and iterative regression analysis. Here the effects were still significant for all three outcomes. The effect size (per Adj R^2 was too small to be important for HWB (0.01), on the border of importance for SC (0.04), and small and important for EL (0.05). Unstandardised coefficients were smaller in the matched sample, but standardised coefficients were actually larger.

When the ADHD subgroup indicator and ADHD severity were both included in the same model, the sign changed on the indicator and the coefficient became insignificant. This indicates problematic collinearity, which is not surprising (the two variables are be highly correlated), but were investigated further and accommodated in subsequent analyses.

5.2 Strengths

The pilot study successfully extended previous work on ADHD in BCS70 (Brassett-Grundy & Butler, 2004) by using a more refined measure, as well as a continuous severity measure, developed in chapter 4. Objectives were also met to evaluate fewer, more person-centred outcome measures to avoid bias from a large number of tests, and to avoid a narrow and negative focus on specific events. Additionally, literature was reviewed on quasi-experimental methods, and approaches were tested and evaluated. The main objective for the testing of methods was to learn about them and develop a plan for a more robust final study of outcomes. The next section on limitations summarises the lessons learnt from the pilot and brief remediation plans.

5.3 Limitations, and improvements planned for subsequent analyses

A key objective of the pilot study was to test measurements of wellbeing outcomes and use of quasi-experimental methods to improve balance between treatment (ADHD) observations and controls (non-ADHD). Lessons learnt were used to develop more robust methods for evaluating the relationship between ADHD and long-term outcomes. Feedback was gathered from my supervisors and added to my own reflections on improvements that should be made in the final study. Limitations and planned improvements are documented in Table 32 and addressed in chapters 6, 7, and 8.

No.	Limitation	Improvements
1.	ADHD predictors literature review only used Google Scholar, thus the process for selecting studies was not systematic enough	Repeat review following library guidance for database searches and PRSIMA-P systematic review protocol
2.	A large percentage of observations with age 10 data were lost when matching to the age 42 data for outcomes	Evaluate data available on outcomes from previous sweeps after age 30 (ages 34 and 38) and use either/or, if feasible
3.	Preferred sample (N=7,242) was not compared to an external source to assess representativeness	Compare sample to 2011 Census, consider/discuss use of weights
4.	Literature on social class is too complex to engage with properly within resource limits, and educational attainment is a better proxy for SES	Use educational attainment only as proxy for SES and measure of objective wellbeing in subsequent analysis – explain reasoning in chapter 6
5.	Psychometric properties of composite health and wellbeing measure were not ideal	Drop indicators not well-supported by literature or model fit and align measure with subjective wellbeing concept supported by ONS
6.	Educational attainment levels were not well-balanced in terms of response	Collapse into fewer categories with more balanced response levels
7.	Key covariates known to be predictors of outcomes for ADHD were not included as covariates	Add age 10 measures of IQ, comorbidity, and low SES (Costello & Maughan, 2015)
8.	Measures to test state regulation theory (stress and protective factors) were not included as predictors or covariates	Add age 10 measures of chronic stressors, life event stressors, locus of control, self-esteem, and engagement in leisure activities
9.	Reporting of descriptive statistics was limited	Create table of descriptive statistics for all analysis variables, including missingness, add to measures section
10.	Selection of covariates for matching relied too much on p-values and would be difficult to reproduce.	Use vselect Stata procedure to simplify selection of best fit covariate set and reduce reliance on p-values
11.	Exact matching resulted in too much data loss and could have increased bias instead of reducing	Use weighted matching procedure

12.	OLS assumptions of normality and linear/continuous outcomes were violated, and missingness was not handled using a robust procedure	Test assumptions, moved analysis to Mplus, used MLR estimator with FIML
13.	There is reason to expect slopes of functions may differ for boys and girls (Brassett-Grundy & Butler, 2004), and they were not reported separately	Run and report separate regression models for boys and girls
14.	Treatment effect comparison between naïve and matched samples was biased because covariates were not controlled in the naïve regressions, and the one-to-one matching approach resulted in significant data loss	Include matching variables as covariates in regression on unmatched sample so direct comparison between methods could be made. Use weighted match CEM procedure to utilise more of the available data.
15.	The ADHD indicator and ADHD severity variables behave in unexpected ways when in the same regression, due to collinearity	Analyse relationship and interaction of variables, accounted for in models
16.	Composite wellbeing measure was not validated against existing measure that has been previously validated (WEMWBS)	Compare outcome results for CM's with WEMWBS data to their outcome results with composite/derived WB measure
17.	Findings were not compared in detail to previous studies of ADHD outcomes	Compare findings to outcomes studies discussed in chapter 3 literature review

Table 32. Limitations in pilot and planned improvements for final outcomes analyses

Learning from the pilot process was extensive and beneficial. The planned remediations of limitations were incorporated into a methods redesign in chapter 6, and separate analysis chapters for subjective wellbeing (chapter 7) and educational attainment (chapter 8).

Chapter 6 Enhancements to measures and methods for study of outcomes

1 Introduction

Chapter 5 reported on a pilot study of methods to evaluate the relationship between ADHD and long-term outcomes. Based on feedback on the pilot and my own reflections on the approach, several changes were planned to increase the robustness of the analysis, which were listed at the end of chapter 5. Chapter 6 documents implementation of these methodological improvements.

1.1 Refined ADHD predictors literature review

The first action from the pilot was to improve the robustness of the literature review process for ADHD predictors. The new process was based on updated University of Cambridge Faculty of Education library online guidance (Cambridge Libraries, 2019). I also consulted guidance for PRISMA-P systematic literature reviews (Moher et al., 2015).

Library guidance recommends that databases should be used in addition to Google Scholar because they allow for more precise search strings with field-level searches and complex combinations of Boolean operators (Cambridge Libraries, 2019). Specifically it is advised that literature searches situated within the psychology and education subject area (like the present thesis) should interrogate the British Education Index, ERIC, PsychInfo, PubMed, Scopus, and Web of Science databases (Cambridge Libraries, 2019). Since this search is specifically about predictors of ADHD, which is a medicalised construct, I limited the searches to PubMed, Scopus, and Web of Science. I started with Wed of Science because it is the most comprehensive of the three.

Web of Science is a collection of 20,000+ high quality peer-reviewed journals, as well as books and conference proceedings, dating back to 1900 (Clairivate Analytics, 2019). The initial search terms were:

```
TI = (ADHD OR Attention Deficit Hyperactivity Disorder OR
Attention Deficit/Hyperactivity Disorder)
AND TI = (predictor OR risk factor OR cause)
AND publication date between 2000 and 2019
```

This search produced 385 results. On review of titles for the first 100 results, keywords for non-relevant studies were identified, and it was noted that studies before 2008 were covered by reviews and meta-analyses. Thus, the following exclusions were added:

```
NOT TI = (genetic OR gene OR methylphenidate OR drug OR
medication OR outcome OR intervention OR adult OR epilepsy OR
autism OR obesity OR EEG OR "college student*" OR mouse OR animal
OR addiction* OR remission OR comorbid*)
AND publication date between 2008 and 2019
```

This search produced 200 results. Titles of these results were reviewed, and further exclusions made for studies reporting on ADHD as a risk factor for something else (e.g. injury, treatment response, driving accidents, etc.), philosophical pieces or editorials, and risk factors that were not relevant to or not collected in the BCS70 age 0 or 5 sweeps (e.g. birth month (all the same for BCS70), bacterial meningitis, soy-based formula, exposure to lead or other toxins). Also studies specific to the ICD-defined Hyperkinetic Disorder were not included, because these are a subset of the most severe cases of ADHD as it is most widely defined in the DSM.

The exclusions left 52 studies. Abstracts were reviewed for all 52, and 17 were identified with data that could be relevant to my analysis, i.e. data that could potentially be mapped to BCS70 age 0 and 5 data. Similar searches were conducted in Scopus, PubMed, and Google Scholar, titles were reviewed, and no additional relevant studies were identified. The Google Scholar search procedure used in study 2 produced 10 viable studies, so the alternative approach used here resulted in seven new studies available to be used as sources of ADHD predictors. The new studies are listed with sample sizes and significant predictors in Table 33.

No.	Study	N	Predictors
1	(Chen et al., 2014)	21,756	Any atopic disease (asthma, dermatitis, allergies), before age 3. Higher number of comorbidities = higher assoc. with ADHD
2	(Cherkasova, Ma, Ba, Pondé, & Hechtman, 2013)	Review (11 studies)	genetic factors, maternal smoking, low birth weight, premature birth, maternal stress and psychosocial adversity
3	(Kim et al., 2009)	2,673	Maternal stress and alcohol use during pregnancy, parental marital discord, changes in caregivers, not breastfeeding.
4	(Oerlemans et al, 2016)	696	Low parental age, maternal diseases, smoking, stress
5	(Park et al., 2014)	649	Maternal stress, postpartum depression, changes in caregiver, less prenatal check-ups, postnatal illness. No differences between subtypes in genes tested.

6	(van de Weijer-Bergsma, Wijnroks, & Jongmans, 2008)	Review (25 studies)	Preterm birth - influenced also by SES, mother education level, home environment, early individual differences in orienting and sustained attention
7	(Wustner et al., 2019)	1,384	Parent mental health, migration status

Table 33. Summary of seven new studies found through refined literature review process: ADHD predictors

17 ADHD predictor variables had been identified in the chapter 5 pilot study literature review and mapped to BCS70 items. I reviewed the seven new studies and was able to identify and map seven additional ADHD predictor variables. The new variables indicated the following: breastfed, sleeping problems in the first 6 months, eczema, hayfever, wheezing, separation from mother of one month or more (changes in caregiver), and count of mother's previous miscarriages, as a proxy for maternal stress during pregnancy (Woods-Giscombé et al., 2010). I also temporarily added a 'migration status' variable (Wüstner et al., 2019), but it was so rare in the data that it was not a useful predictor, so I removed it.

Finally, I added another proxy measure of IQ, a widely reported predictor of ADHD outcomes (e.g. Costello & Maughan, 2015). It was the age-5 score from the English Picture Vocabulary Test (EPVT), to serve as a more precise pre-age-10 measure than the health visitor's 'backward' indicator. With the eight new variables, 25 ADHD predictor and ADHD outcome predictor variables were identified in total for use in the subsequent outcome analyses. The new variables and BCS70 mappings are shown in Table 34, and descriptive statistics for all 25 are included in Table 54 at the end of the Measures section.

No.	Item	Age	BCS70 variable	Varnames and mapping notes
	Breast fed	5	E020	F5Bfed (0/1) Mother's recollection of breastfeeding child (ever)
	Sleep problems	5	E079	F5SlPr1st6 (0/1) Mother's recollection of sleep problems in child's 1st 6 months
	Eczema	5	E067	F5Ecz (0/1)
	Hayfever	5	E068	F5HayF (0/1)
	Wheezing	5	E087	F5Wheez (0/1)
	Separation from mother > 1 month	5	E030	F5MumSep1m (0/1)
	Number of miscarriages mother had before CM pregnancy	0	A0167	F0PMisc (count)

	English Picture Vocabulary Test score	5	F120	F5SEPVT (continuous, normal)

Table 34. Eight new ADHD predictors identified via refined literature review

2 Method

2.1 Data

2.1.1 Recovery of missing age 42 outcome responses using age 34 data

In the pilot study, there were 11,426 children with data at age 10, and 7,242 of them had corresponding outcome data at age 42. Within the ADHD subgroup, there were 594 at age 10, and 298 with data at age 42. The loss of 36.6% of observations overall and 49.8% of ADHD subgroup members was significant, and since there were other BCS70 sweeps post-age-30 (settled adulthood), a sensible improvement was to recover some of the lost observations using data on outcomes from another sweep. The sweep prior to age 42 was conducted at age 38, but it was a telephone interview with relatively high non-response, and after a review of the documentation I determined that age 38 data on wellbeing outcomes was inadequate.

Prior to age 38, a full face-to-face sweep was conducted at age 34. I reviewed the age 34 sweep documentation and found that all of the relevant adult outcome data I used in chapter 5 from age 42 was also measured at age 34, with the exception of the Warwick Edinburgh Mental Wellbeing Scale (WEMWBS). A query of the age 34 data revealed that 1,277 of the age 10 observations lost to attrition at age 42 could be recovered overall, 71 of those from the ADHD subgroup, increasing the sample sizes to N=8,519 and n_{ADHD}=369. Although it would be preferable to have the WEMWBS data, the age 34 sweep was the only viable option over 30 with a full range of other outcome data and offered a sizable data gain.

Literature differs on whether or how wellbeing changes with age. An analysis of three large longitudinal surveys found that age explained very little of the variance in another validated psychological/subjective wellbeing measure (Ryff's scales; Springer & Hauser, 2006). Other studies have found the relationship between age and wellbeing to be a function that is either u-shaped (Blanchflower & Oswald, 2008) with the low-point in mid-life, or separate spline functions across three stages of adulthood (Wunder et al., 2009). Regardless, age 34 and 42 are usually considered to be part of a similar life stage (e.g. early middle-age), and wellbeing across the two time points should be similar, or perhaps slightly lower at age 42 (Blanchflower & Oswald, 2008; Wunder et al., 2009). Thus, the age 34 BCS70 wellbeing data was selected to augment outcome data using the variables listed in Table 35.

Variable name	Description
b7khlstt	Self-reported general health
b7lifet1	Self-reported life satisfaction
b7pachk	Lives with partner
b7sc	Social class
b7khldl2	Registered disabled
BD7MAL	Rutter Malaise score
BD7HACHQ	Educational attainment level
bd7dunit	Units of Alcohol per week
bd7dgrp	Alcohol use group
b7region	UK region of residence

Table 35. Outcome variables extracted from the age 34 sweep

Up to this point, my working dataset was comprised of data from the age 0, 5, 10, and 42 sweeps, and contained a large number of variables that I had cleaned, recoded, or derived for use in in the pilot. So, many variables did not exist in my dataset for the observations I now wanted to include using outcome data from age 34. Consequently, I had to rebuild my working dataset, starting with a clean version of the age 10 data from Chapter 4 (N=11,426). I created a copy of the data, then re-ran Stata code to merge in predictor variables from the age 0 and 5 sweeps and generated other variables that had been added for the pilot/chapter 5. Finally, I merged in the age 34 outcomes data, and dropped all observations (from the N=11,426) without outcome data available from EITHER age 42 OR 34. The resulting dataset contained 655 variables (N=8,519, n_{ADHD} = 369). About 200 of these were temporary variables created for chapters 4 and 5 and not needed going forward, so were dropped. The new working file name was:

```
Chapter5Work_out_short.dta
```

Finally, new variables were created to allow for analysis using the outcome from age 42 if it was available, OR if not, the outcome from age 34. New variables were prefixed "OE", for Outcome from Either (42 or 34). Age 42 was given preference over 34 (if outcomes from both were available) because it was the most recent. The code for rebuilding and incorporating new data was stored in:

```
GetAge34OutcomesData.do; and
RebuildChapter5Work.do.
```

2.1.2 Comparing representativeness of the preferred sample to an external source

The third action from the pilot was to investigate representativeness of the preferred sample more thoroughly, by making comparisons to an external source of data on the target population. The objective was to gain evidence about the strength of conclusions that could be drawn from analysis of the preferred sample. There were two 'target' populations that my sample (N=8,519) could be compared to: the original set of cohort members with data collected in 1970 (i.e. the original cohort), and the population of Great Britain near the time the most recent outcomes were measured (2012). Representativeness of the 1970 cohort compared to the 2012 sample was not analysed in the present study because it has been done elsewhere (see chapter 2 for discussion; Mostafa & Wiggins, 2015). Thus, a comparison was made of the preferred sample to external data from the UK census.

The 2011 census was completed close in time to the 2012 BCS70 sweep, so was an ideal candidate for comparison. Based on previous analysis of BCS70 non-response (see chapter 2), the groups most likely to be under-represented in our 2012 preferred sample should be males and low SES (Ketende et al., 2010; Mostafa & Wiggins, 2015). Accordingly, a comparison was made between the BCS70 new preferred sample (N = 8,519) and 2011 census data grouped by age, sex, low SES job, and low SES qualifications.

2.1.2.1 2011 census comparison

2011 census data was accessed using a publicly available Office of National Statistics website (Office for National Statistics et al., 2017). I found that the 2011 UK Census data had reporting categories that were similar to BCS70, but would not allow exact, like-for-like comparisons. Data was available by age group (not exact age), sex, National Statistics Socioeconomic Classification (NS-SEC, broadly comparable to the Social Class reported in BCS70), and highest level of academic qualifications. The latter two are both widely used as measures of SES, or relative (dis)advantage. BCS70 data pertains to Great Britain only (not the whole of the United Kingdom), so I used census data either from Great Britain, or England and Wales combined, if Great Britain numbers were not available.

	BCS70 (N=8,519)	2011 Census	RRR	Notes
Males	48.50%	49.39%	0.98	BCS70: age 42 or 34 outcomes available Census: ages 40-44
Lower SES job	27.12%	28.96%	0.94	BCS70: age 42 or 34, SC IV (Partly skilled), V (unskilled), or job not coded or reported [35]

[35] The 'not currently employed' category in BCS70 includes cohort members in full-time education, but the numbers are very small so did not affect analyses significantly (n=6). Also, this BCS70 group includes any cohort member who did not report working part-time or full-time at the time of the 2012 survey, whilst the census data from 2011 only includes long-term unemployed (2 years or more).

				Census: age 35-49, NS-SEC 6, 7, 8, semi-routine, routine, never employed and long-term unemployed
No academic qualifications	28.42%	20.27%	1.40	BCS70: age 42 or 34 highest academic qualification level, GB Census: age 35-49, England and Wales only, highest academic qualification

Table 36. Comparison of Census 2011 data to BCS70 new preferred sample including age 42 (or 34) outcomes

RRR = % in BCS70 / % in 2011 census

Table 36 shows that the subset of participants in my preferred sample did not uniformly or dramatically under-represent the proportions of males or low SES participants reported in the 2011 census. The proportion of males in both samples was similar, the BCS70 sample was 6% less likely to be in a low socioeconomic job classification, and the BCS70 sample was 40% more likely to have no academic qualifications.

The latter finding was unexpected. A possible explanation for the greater proportion of no qualifications in the age 42 BCS70 group is the broader 35-49 age range included in the census statistic. The BCS70 cohort reached school leaving age (16) in 1986, and the youngest persons in the census group at age 35 in 2011 would have reached school leaving age in 1992. During that time frame, a dramatic change took place in the UK with regards to absence of academic qualifications: between 1985 and 1995; the proportion of the workforce with no qualifications dropped 13.2 percentage points for women and 14.7 for men, based on General Household Surveys (Machin & Vignoles, 2005). This would naturally make the no qualifications statistic lower for the broader age group from the census, so I concluded this was not a source of substantial bias.

There were no obvious reasons for the BCS70 age 42 group in 2012 to have fewer low SES jobs than the age 35-49 group in 2011, other than again the wider age group represented by the census data. There were also slight differences between the categorisation of low SES in BCS70 and the census. Regardless, the difference was very small, and not expected to cause bias in related estimates.

Thus far I have made no comparison on ethnicity, which is commonly evaluated for representativeness in large studies. However, as a birth cohort study, BCS70 should be representative of the ethnic mix in the population when the cohort was born, not necessarily in the decades that followed. That was indeed the case with the BCS70 preferred sample. The ethnic groups reported in census population statistics were again not exactly comparable to those reported in BCS70. For example, census ethnic groups included: white British, mixed-ethnicity British, Asian/Asian British… etc. BCS70 groups included: 'English etc', Irish, Other European, West Indian, other, etc. So, the census categories described race/ethnicity and country of origin, whilst the BCS70 categories only described country of origin. However, if I assume that the English, etc., Irish, and Other European categories in BCS70 all describe a 'white' ethnicity, the original BCS70 sample was probably more than 97% white. In the 2011 census, the UK-born population was reported as 92% white, and including non-UK born reduces the figure to 86%. Thus, by this crude comparison, and in spite of some attempts to recruit non-UK born participants in later BCS70 sweeps (Butler et al., 1997; Elliott & Shepherd, 2006), our preferred sample is not at all representative of the ethnic mix in the UK in 2011. This is most likely due to increased immigration to the UK after the study started in 1970, particularly in the late-1990's onwards (Office for National Statistics, 2016). This disparity was and should be accepted as an inherent characteristic of the BCS70 data[36].

2.1.2.2 Evaluation of bias from attrition in the ADHD subgroup

Whilst it is important to understand the composition of the preferred sample as a whole, it is also is important to understand what effect matching the age 10 to age 42 and 34 data had on the ADHD subgroup identified at age 10. Of the age 10 ADHD subgroup, n=594, n=369 also responded at age 42 or 34, or 62%. There were 32 (or 24) years between the childhood measures and outcome measures, so many random factors, such as death, emigration, and unsystematic refusal to participate, could explain the attrition. However, to assess for possible sources of bias, I cross-tabulated the ADHD subsample by categories that could be expected to affect outcomes.

In a review that has been referenced throughout the chapters of the thesis, important factors in ADHD outcomes were reported to be intelligence, socio-economic status, comorbid conduct problems, and severity of ADHD symptoms (Costello & Maughan, 2015). Accordingly, these factors were compared for representativeness between the original and reduced ADHD subgroup. Categorical variables from age 10 were used as proxies as follows: Maths test z-score $<= -1.645$ ($\alpha=0.95$) for lower intelligence, a Free School Meals (FSM) indicator for lower SES, and a mother-reported externalising problems z-score $>= 1.645$ for conduct problems. Both reading and maths scores were available, but maths was used here because it has been reported to be more reliably related to general intelligence than reading (Floyd et al., 2003; Primi et al., 2010; Taub et al., 2008). Free School Meals was used because it has been validated as a good proxy for low income/socio-economic deprivation (see Hobbs

[36] Effects on generalisability are discussed in chapter 9.

& Vignoles, 2010). Both mother and teacher ratings of externalizing items were available, but mother ratings were used here because they mapped more precisely to the SDQ externalising subscale. Teacher ratings are generally seen as more objective than parent ratings, so these scores are likely to contain some bias, but mapping to the widely validated SDQ scale was given priority. Comparisons of the variables were made between the Full ADHD subgroup identified at age 10, and the subset with outcome data from either age 42 or 34 (Table 37).

	Full age 10 ADHD subgroup (n=594)	ADHD subgroup with outcomes data at age 42 or 34 (n=369)	RRR	t/p
Lower IQ (maths z <= -1.645)	17.00%	14.36%	1.18	
Lower SES (FSM)	27.44%	24.66%	1.11	
Conduct Problems (externalising z >= 1.645)	50.67%	46.61%	1.09	

Table 37. Comparison of full ADHD subgroup at age 10 and matched adults with outcomes at age 42 or age 34

RRR = % in ADHD group identified at age 10 / % of original ADHD group with outcomes

The ADHD subgroup with outcomes data at age 42 or 34 (n=369) was less representative of children with lower intelligence, lower SES, and higher externalising/conduct problems than the original subgroup identified at age 10 (n=594). The effect sizes indicated by the relative risk ratios were however very small. Thus, here again I concluded that attrition was not a significant source of bias for the ADHD subgroup based on representativeness within influential variables.

2.2 Measures

2.2.1 ADHD

The ADHD categorical indicator and dimensional severity score derived in chapter 4 were used as the ADHD measures in subsequent analyses. There was no missingness for either variable because only age 10 children who could be evaluated and scored for ADHD were included in the preferred sample (N=8,519; n_{ADHD}=369).

2.2.2 Outcomes

Learning from the pilot reported in chapter 5 inspired changes/improvements to measures of outcomes. Combining subjective and objective wellbeing was not well supported by literature, and so perhaps unsurprisingly the health and wellbeing measure (HWB) I tested did not have good psychometric properties. So, I reduced the set of indicators to three that were highly correlated and supported in literature as measures of psychological subjective wellbeing (SWB): self-rated life-satisfaction, Rutter malaise score, and WEMWBS score.

I also learned via the pilot that using both social class of job and educational attainment as proxy measures of SES and indicators of objective wellbeing was problematic. The social class of job classification system (levels 1-5; professional - unskilled) contains ambiguities and use of it requires engagement with a complex literature. For example, the system is a point-in-time measure, based on current job only, so does not accommodate a housewife/mother role, temporary, or elective under or unemployment. Accordingly, social class had high (~16%) missingness, high enough to reduce power and introduce bias. Highest educational attainment level, however, is almost always stable after age 25, and in my preferred sample had 0% missingness. The stability and high completeness added to a strong positive association with income, wealth, health behaviours, and access to opportunities make educational attainment a highly preferred single proxy measure for SES (Carlton, 2012; Galobardes et al., 2006; Hackman & Farah, 2009; Stevens et al., 2009). Additionally, the relationship between ADHD educational attainment is of particular interest, since several other studies have found that ADHD is linked to lower attainment (e.g. Erskine et al., 2016; see discussion in chapter 3).

Therefore, two outcomes were measured in adulthood: subjective wellbeing (SWB) and educational attainment level (EDL), the latter as a proxy for SES and indicator of objective wellbeing. Outcomes were sought for the age 10 sweep respondents first from age 42 data, because it was more recent, and contained the validated subjective wellbeing measure (WEMWBS). If data was not available at age 42, outcomes from age 34 were used.

2.2.2.1 Subjective wellbeing (SWB)

The three selected measures, Warwick Edinburgh Mental Wellbeing Scale (WEMWBS), self-rated life satisfaction, and the adapted Rutter Malaise score[37], all had good psychometric fit in the pilot, and, as noted in the literature review in chapter 3, all three have been used in other studies to measure (psychological) subjective wellbeing in BCS70 (Goodman et al., 2015;

[37] For more detailed descriptions of WEMWBS, Malaise, and life satisfaction, see chapter 5.

Layard et al., 2014; Sacker & Cable, 2006; Schoon & Kneale, 2013; Sigle-Rushton et al., 2005; Wood et al., 2017). I briefly considered using only WEMWBS as a measure, because it has the advantage of a validation history (e.g. Bass, Dawkin, Muncer, Vigurs, & Bostock, 2016). However, the variable had high (17%) missingness in the sample linked to age 42 (N=7,242), which increased to almost 30% when the age 34 data were added, because WEMWBS was not administered in the age 34 sweep.

The 9-item version of the Rutter Malaise inventory (Brown & Hancock, 2014; Rodgers et al., 1999; Rutter et al., 1970) and self-rated life satisfaction, were available from both the age 42 and 34 sweeps. Amongst the three indicators there were seven missing data patterns, but over 90% of observations had values for at least two indicators. The three measured variables were used to build a confirmatory factor analysis model of a single latent subjective wellbeing variable and estimate a factor score.

None of the three indicators were measured as continuous variables, but rather as discrete sets of values that could be assumed to be ordered, but not necessarily to have equal distance between points. Thus, they were technically ordinal, or ordered categorical. However, all three variables had 10 or more levels. It is sometimes controversial but also common practice to model variables as if they were continuous when they are ordered categorical and have 10 or more levels or values. In fact, MPlus does not allow variables to be declared as categorical unless they have less than 10 levels. Accordingly, the variables were declared as continuous, and means, SDs, medians, and Pearson's correlations are reported (Table 38 and Table 39).

Measure	Variable	N	Range	Median	Mean	SD	Missing
Life satisfaction	OELife	8,428	0-10	8	7.34	1.97	1.07%
Malaise	OEMalR	7,618	0-9	8	7.16	1.98	10.58%
Wellbeing	O42WarwickWB	5,981	14-70	50	49.25	8.26	29.79%

Table 38. Descriptive statistics for indicators of subjective wellbeing (SWB)

(R) – Reversed, so correlations between the three measures are all positive
(Total N=8,519: includes 7,242 observations with age 42 outcomes, and 1,277 with age 34 outcomes)

	OELife	OEMalR	O42WarwickWB
OELife	1.000		
OEMalR	0.398	1.000	
O42WarwickWB	0.494	0.623	1.000

Table 39. Correlations for SWB indicator variables

All correlations significant, $p < .001$

Exploratory factor analysis was first done in Stata to evaluate properties of the three items together. Eigenvalues and a scree plot indicated a single factor, though uniqueness was fairly high for life satisfaction (0.68), indicating a lower relevance to the construct. However, life satisfaction had the lowest proportion of missing data, and a minimum of three indicators is preferable for a factor score (Worthington & Whittaker, 2006), so it was retained.

Descriptive statistics and normal quantile plots showed that all three distributions were non-normal. In this case, Maximum Likelihood Robust (MLR) is the recommended estimator for a measurement model in Mplus (Brown, 2006). Thus, a confirmatory factor analysis was constructed using the structure in Figure 24. The MPlus MLR model produced good global fit indices (χ^2 = 14.848, p < 0.001, *CFI* = 0.996, *TLI* = 0.988, *RMSEA* = 0.04) based on widely-used thresholds (Hu & Bentler, 1999).

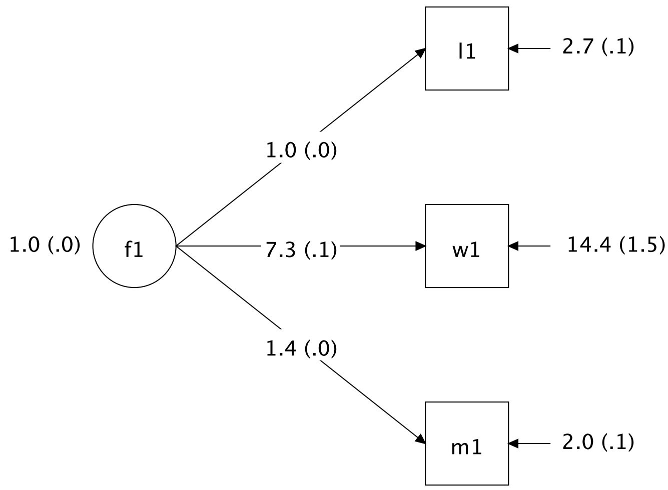


Figure 24. Diagram of confirmatory factor analysis model used to estimate subjective wellbeing

Life Satisfaction (l1), Warwick-Edinburgh Wellbeing (w1), and Rutter Malaise (m1; reversed) to measure the latent factor (f1) Subjective Wellbeing.
Statistics: estimates and (SEs) for endogenous variables, residual variances and (SEs) for exogenous

2.2.2.2 Educational attainment level (EDL) as a proxy for SES and indictor of objective wellbeing

In the chapter 5 pilot, educational attainment was based on the BCS70 variable BD9HACHQ (highest academic qualification achieved) measured in 2012 at age 42. To reduce missingness of outcomes in the ADHD subgroup, data was added where available from the same BCS70 variable (BD7HACHQ) measured at age 34. It is unlikely that the data would have changed between ages 34 and 42 because obtaining qualifications in later life is relatively rare. A new combined (age 42 or 34) variable was created with nine levels labelled from 0-8, as follows:

0 - no academic qualification
1 - GCSE D-E[38]
2 – CSE 2-5, other Scottish quals
3 – GCSE A-C, good O levels, Scottish standards
4 - AS levels or 1 A level
5 - 2+ A levels, Scottish higher/6th
6 - Diploma
7 - Degree level
8 - Higher degree

Exploratory data analysis of the (age 42/34) 0-8 level variable showed that the response to Level 1 (GCSE D-E or similar) and Level 4 (AS levels or 1 A-level) was very low compared to other levels (Figure 25).

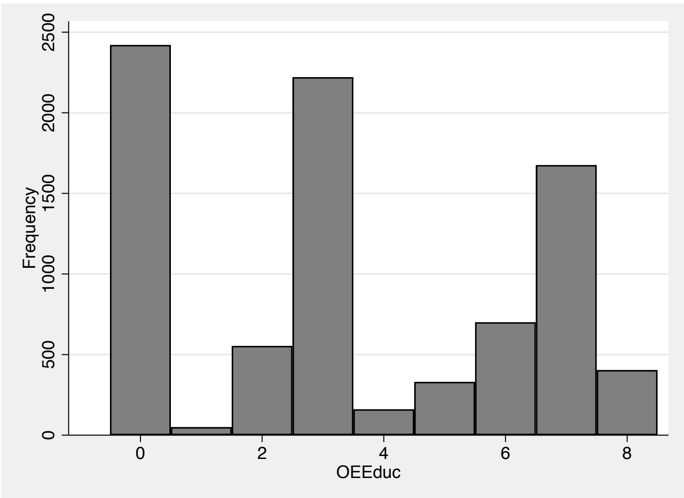


Figure 25. Histogram of highest academic qualification
(8 levels) from BCS70 at age 42 (or 34), N=8,519

A paper on classifying educational qualifications specifically in British survey data reports that others have used a collapsed version of this scale with six levels (labelled 0-5), combined as follows: 1+2, 4+5, and 6+7 (see Jenkins & Sabates, 2007). In the interest of improving balance across categories, a new variable was created and recoded in this way. However, cross-tabulations of the 0-5 level measure revealed there were no girls and only 2 boys in the ADHD subgroup with a higher degree (level 5) in my sample (Table 40).

[38] The BCS70 cohort reached school leaving age in 1986, and would have taken O-levels, not GCSEs. The education level descriptions include only the most common or recognizable qualifications, i.e. they can and do include other similar qualifications. For example, the GCSE D-E category also includes poor O-level results, and the equivalent Scottish qualifications. See Dodgeon & Parsons, (2011) for a full description of the level derivations.

Level	Description	Girls	Boys	Total
0	No academic qualifications	56	114	170
1	GCSE D-E or CSE 2-5	17	30	47
2	GCSE A-C	29	57	86
3	AS levels, 1 A level, or 2+ A levels	8	12	20
4	Diploma or Degree	17	27	44
5	Higher degree	0	2	2
		127	242	369

Table 40. Academic qualifications (5-level coding) by sex, for the ADHD subgroup only (n=369)

To avoid instability of analysis with empty or very small cell sizes, I collapsed the levels further by combining 4+5 in another new 0-4 level variable. The name in Mplus was OEDUC4, and the distribution across levels for each sample is provided in Table 41.

0 - no academic qualification
1 – GCSE D-E or similar
2 – GCSE A-C or similar
3 - AS levels, any A levels, or similar
4 – HE diploma, degree or higher degree

OEDUC4

Level	All		ADHD	
0-no quals	2,421	28%	151	45%
1-GCSE D-E	604	7%	43	13%
2-GCSE A-C	2,221	26%	80	24%
3-A levels	492	6%	18	5%
4-Degree+	2,781	33%	41	12%
	8519	100%	333	100%

Table 41. Academic qualifications (4-level coding) by sample

Matched sample N = 6,207 because it was pruned to be more comparable to the ADHD group, and observation counts are not integers because the coarsened exact matching weights were applied

Based on this 0-4 highest academic qualifications variable, the proportion of cohort members in the ADHD subgroup with no qualifications (45%) was much higher than in the full unmatched sample (28%). Cohort members with a highest qualification that was vocational were reported in this variable as having no (academic) qualifications. I wondered how a different measure that included vocational qualifications might affect results, particularly for those in the ADHD subgroup. So, I created an additional measure of education level using the BCS70 variables BD7HNVQ (age 42) and BD9HNVQ (age 34), which accounted for both vocational and academic qualifications. A full description of the qualifications mapping can be found in Jenkins & Sabates, (2007). It is important to note there is controversy about whether vocational qualifications should be assumed equivalent to academic qualifications, due to wide variation in quality control structures and standards (Wolf, 2011). Data from the two five-level NVQ-based variables was cleaned and combined. Again, there were very few observations in the highest NVQ level for the ADHD group, so I combined categories 4+5 into a final variable with 0-4 NVQ levels. The variable name in Mplus was NVQ, and the distribution is shown in Table 42.

NVQ4

Level	All		ADHD	
0-no quals	952	11%	60	18%
1-nvq1	674	8%	42	13%
2-nvq2	2,190	26%	108	32%
3-nvq3	1,273	15%	57	17%
4-nvq4	3,427	40%	66	20%
	8516	100%	333	100%

Table 42. Academic and vocational qualifications (4-level coding) by sample

N = 8,516 because 3 observations were missing the NVQ4 data

The disparity between the full (All) and ADHD samples using the NVQ was smaller, particularly in levels 0 and 4.

Finally, I created two binary 'high' and 'low' educational attainment variables to allow for simpler analysis and use of additional statistics (Table 43 and Table 44):

OEDUC2: Academic only, with higher cut-point: high = A-levels (3) or above

OEDUC22: Academic only, with lower cut-point: high = GCSE A-C (2) or above

OEDUC2

Level	All		ADHD	
Low/GCSE A-C or below	5,246	62%	274	82%
High/A-levels or above	3,273	38%	59	18%
	8,519	100%	333	100%

Table 43. Academic qualifications (2-level coding) by sample – higher cut-off

OEDUC22

Level	All		ADHD	
Low/GCSE D-E or below	3,025	36%	194	58%
High/GCSE A-C or above	5,494	64%	139	42%
	8,519	100%	333	100%

Table 44. Academic qualifications (2-level coding) by sample – lower cut-off

N.B. I created two versions of the academic variable to allow for comparisons to the previous study of ADHD in BCS70 (Brassett-Grundy & Butler, 2004; they used the lower cut-off).

2.2.3 Other adult wellbeing measures

In chapter 3 (see Table 5), twelve possible measures of adult wellbeing were identified within BCS70 data based on the ONS-defined domains. After exploratory analysis of the data and the pilot work in chapter 5, four were accounted for in the SWB and EDL measures. Other measures, including general health, disability indicator, alcohol problems indicator, social class of job, and lives with partner indicator were retained in the working dataset and included in some of the multivariate regression models in chapter 7 to capture their contribution to variance in the prediction of subjective wellbeing. Income and satisfaction with home were not used due to high missingness, and working indicator was not used due to insufficient variation. Further background details on the other measures are available in chapter 5.

2.2.4 Key confounding covariates

IQ, comorbidity, and low socioeconomic status were selected as key confounding covariates because they were reported as influential on long-term outcomes for ADHD in recent reviews and a large longitudinal follow-up (Costello & Maughan, 2015; Erskine et al., 2016; Roy et al., 2017). All three of the constructs are complex, the subject of extensive literature, and can be controversial. No measures currently considered robust for these constructs were available in the BCS70 data at age 10. Thus to minimise engagement with complex bodies of literature, I elected to use simple proxy measures for them: standardised maths for IQ (and reading as a secondary measure), mapped SDQ subscales for comorbidity, and free school meals indicator for low socioeconomic status. The use of the maths (and secondarily reading) score for IQ may seem contentious and oversimplified from a psychology point of view, so next I discuss the reasoning for the decision.

2.2.4.1 Reasoning for proxy measures of IQ

The reviews did not consistently specify the measure(s) used for IQ, but use of the term implies either a specific IQ score measure (e.g. the Wechsler Abbreviated Scale of Intelligence, or WASI-II; Wechsler, 2011), or a similar validated measurement of a general intelligence factor. As noted above, there is no specific validated IQ measure (e.g. WASI-II) in BCS70. Other researchers have studied intelligence in BCS70 at age 10. I reviewed a selection of these found by searching the CLS Bibliography for 'intelligence' within the BCS70. I found that other authors used composite factor scores, either incorporating data from all or from a subset of eight tests administered to measure cognitive ability and learning (Daly & Egan, 2017; Furnham & Cheng, 2017; Gale et al., 2007, 2009; von Stumm et al., 2009). The eight tests were: Shortened Edinburgh Reading Test, Friendly Maths Test, Pictorial Language Comprehension Test, and four modified subscales of the (21 available) version I British Ability Scales (Elliott et al., 1979): Word Definitions, Word Similarities, Digit Recall, and Matrices (Butler et al., 1997; Parsons, 2014). There was variation amongst approaches. A possible standard was suggested by a data note prepared within the Centre for Longitudinal Studies that used principal component analysis of all eight measures to confirm a hypothesis for a single underlying cognitive ability factor (Parsons, 2014). The highest loadings were for the Friendly Maths Test and Edinburgh Reading Test (both ~0.86). However, the resulting predicted factor score was not shared with the BCS70 data. I noted that the factor loading for BAS recall of digits was fairly low (0.53), and also that the BAS scoring procedure used for BCS70 was 1) not provided with the study documentation, and 2) criticised for reliance on subjective judgements. Thus replicating the composite measure would add considerable complexity to my work here, and may not necessarily measure IQ per se.

I noted in other literature that important aspects of IQ have been shown to predict maths and reading ability in schoolchildren under age 10 (Mayes et al., 2009). In particular maths predicts domain-general fluid reasoning (Green et al., 2017), and working memory (Alloway & Passolunghi, 2011). Maths and reading were the highest loading factors in the principal component analysis of 'general cognitive ability'(Parsons, 2014), and use of these separate scores is more straightforward to interpret than a composite score. Of the two, maths s thought to be the stronger predictor of IQ, as it draws on a wider range of abilities (Green et al., 2017). Thus the Friendly Maths Test score was selected as a primary proxy measure for IQ, and the Edinburgh Reading Test score as a secondary proxy measure for use in selected models.

2.2.4.2 Maths

The Friendly Maths Test score from age 10 was based on the BCS70 variable BD3MATHS. The excerpt below from the age 10 data guide provides a helpful description of the test:

> *"The lack of a fully acceptable mathematics test appropriate for ten year olds led to the development of a special test for the BCS70 Ten-year Follow-up. This was done in BCS70 collaboration with Colin Appleton and John Kerley, specialists in primary mathematics. It was piloted in two halves in Bristol schools each on 400 children. It consisted of a total of 72 multiple choice questions and covered in essence the rules of arithmetic. number skills, fractions, measures in a variety of forms, algebra, geometry and statistics."*
>
> *(Butler et al., 1997, p. 1.9-1.10)*

BD3MATHS had a roughly normal distribution (Figure 26), so was standardised to a z-score for use in analysis as a proxy measure of IQ.

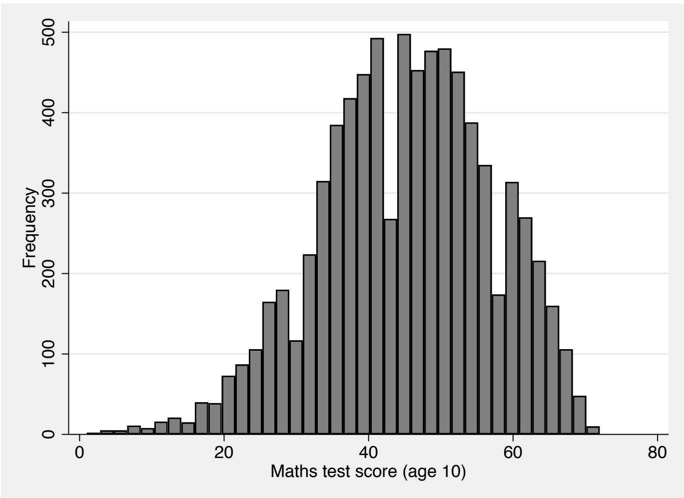


Figure 26. Histogram of maths test score variable at age 10
N=8,519

2.2.4.3 Reading

Reading skills were measured at age 10 using the Shortened Edinburgh Reading Test, which was adapted and normed specifically for BCS70 and published for use elsewhere. A derived reading 'age 'variable (BD3RAGE) was standardised to provide a relative measure within the cohort. The distribution of the variable is shown in Figure 27.

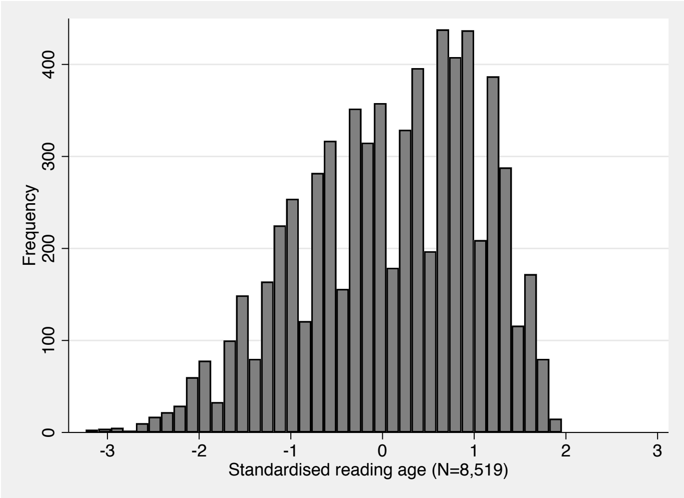


Figure 27. Histogram of standardised reading age score at age 10

2.2.4.4 Comorbidity – externalising and internalising problems

Diagnoses of comorbid psychiatric disorders were also not available (as was the case with ADHD), because culturally, the practice of diagnosing children with disorders was so rare at the time the data were collected. However it was possible to map some of the BCS70 age 10 behaviour items to short sub-scales of internalising and externalising behaviours from the Strengths and Difficulties Questionnaire (SDQ; youthinmind, 2012).

Externalising and internalising behaviours are defined as:

> *"…a broad classification of children's behaviors and disorders based on their reactions to stressors. Externalizing behaviors and disorders are characterized primarily by actions in the external world, such as acting out, antisocial behavior, hostility, and aggression. Internalizing behaviors and disorders are characterized primarily by processes within the self, such as anxiety, somatization, and depression."*

(APA, 2020)

These broad constructs have played a central role in the measurement of childhood difficulties for decades and feature in widely-used scales such as the Rutter Behaviour Questionnaires (Rutter, 1967) and the Child Behavior Development Checklist (CBCL; Achenbach & Edelbroch, 1983; Achenbach & Ruffle, 2000). DSM-5-defined constructs such as Conduct Disorder and Oppositional Defiant Disorder are often (about 50% of cases; Ter-Stepanian et al., 2017) comorbid with ADHD, and are comprised of externalising, or 'acting out' problems. DSM-5 anxiety and depression are also often (about 25-33% of cases; Ter-Stepanian et al., 2017) comorbid with ADHD, and are comprised of internalising, or 'within the self' problems. Thus, although high scores on internalising and externalising behaviour scales are not diagnoses, they are indicative of the most common ADHD comorbidities.

The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997; youthinmind, 2012), which was discussed in chapter 4, is a current, widely-used and well-validated scale for measuring externalising and internalising problems in school-age children. SDQ items were derived from Rutter scales (Goodman, 1997), so align well to the Rutter (Rutter, 1967) items included in BCS70. Thus, I chose to use a mapping of SDQ subscales to measure externalising and internalising behaviour. Teacher ratings are generally preferred to parent ratings to measure child behaviour, but I used parent ratings here because they mapped more precisely to the SDQ subscales (0).

Externalising SDQ conduct problems scale	Mapped BCS70 item
ITEM 5: Often has temper tantrums or hot tempers (I get very angry)	M80-displays outbursts of temper, explosive or unpredictable behaviour
ITEM 7: Generally obedient… (I usually do as I am told)	M56-is often disobedient (rev)
ITEM 12: Often fights with other children… (I fight a lot)	M46-frequently fights with other children
ITEM 18: Often lies or cheats (I am often accused of lying or cheating)	M60-often tells lies
ITEM 22: Steals from home, school or elsewhere (I take things that are not mine)	M52-sometimes takes things belonging to others

Internalising SDQ emotional problems scale	Mapped BCS70 item
ITEM 3: Often complains of headaches… (I get a lot of headaches…)	M15-child complains of headaches often
ITEM 8: Many worries… (I worry a lot)	M48-often worried, worries about many things
ITEM 13: Often unhappy, downhearted… (I am often unhappy….)	M51-often appears miserable, unhappy, tearful, or distressed
ITEM 16: Nervous or clingy in new situations… (I am nervous in new situations…)	M58-tends to be fearful or afraid of new things or situations (mapped to both SDQ items 16 and 24)
ITEM 24: Many fears, easily scared (I have many fears…)	

Table 45. Mapping of SDQ internalising and externalising subscale items to age 10 parent-rated BCS70/Rutter items

Whilst ADHD, externalising, and internalising, problems often coexist (i.e. are comorbid), there is no overlap across any of the three constructs in terms of DSM symptoms or the mapped BCS70 questionnaire items.

The BCS70 items were recoded into three levels: not true/somewhat true/certainly true, and scored 0, 1, and 2, like the SDQ. The items were measured using Visual Analog Scales, retrospectively coded from 0-100, as described in chapter 4. These were divided into thirds and recoded as categorical data: cut-off points used for levels 2 and 3 were 32 and 67. This was the same approach used in chapter 4 for ADHD symptoms, and similar to the previous measurement of ADHD in BCS70 (Brassett-Grundy & Butler, 2004). Raw scores were summed and then used to calculate a z-score. Since externalising and internalising problems are relatively rare in a non-clinical sample, similar to ADHD, the distribution was not normal but zero-inflated. Thus the z-score is not an ideal measure, but it has been used elsewhere to measure similar constructs (e.g. Goodman, Joshi, Nasim, & Tyler, 2015).

2.2.4.5 Free school meals indicator as a proxy for low socioeconomic status

A free school meals indicator at age 10 was used as a proxy measure for low socioeconomic status/family income (Hobbs & Vignoles, 2010). This was selected in preference to a more detailed categorical income or SES measure because it applied to a 12-month period (rather than a single point-in-time, like the available income and job measures), finer details of income or SES were not of particular interest, and it simplified regression analysis.

The data on free school meals was reported in item m126 on the parent (maternal) questionnaire. A 'yes' answer is indicative of low socioeconomic status/income. Free school meals were more than twice as likely in the ADHD sample than in the whole sample (Table 46).

Free school meals	Whole sample		ADHD sample		RRR
No	7,062	86%	278	75%	0.87
Yes	1,063	13%	91	25%	1.92
Missing	25	<1%	0	0%	
	8,519	100%	369	100%	

Table 46. Distribution of Free School Meals indicator in full and ADHD samples

2.2.5 Covariates per State Regulation Theory

As described in chapter 3, state regulation theory was operationalised as stressful life events, chronic stressors, and three protective factors: self-esteem, locus of control, and engagement in leisure activity.

2.2.5.1 Stressful life events

Stressful life events were data-mined from BCS70 in a similar way to the DSM-5 ADHD criteria in chapter 4. Most stressful life events scales descend from Holmes & Rahe (1967), and only a handful of scales have been adapted for childhood and adolescence. Scales used in four studies assessing children's life events were reviewed here to capture as complete a list as possible. They were the Coddington Life Events Record from 1972 (LER), as used in (Williamson et al., 1995), the Stressful Life Events Schedule for children and adolescents (SLES; Williamson et al., 2003), the Life Events Questionnaire (LEQ) developed also based on Coddington's LER and used in Berden, Althaus, & Verhulst, (1990), and an intake questionnaire used in schools in Hungary (HIQ; Mayer et al., 2009).

Life events were identified by using a union of items from the two most comprehensive scales: LEQ and HIQ, and, and searching for semantically similar indicators in the BCS70 data. Items were also compared to the LER (many overlapped with LEQ), and one item was also added from the top 20 stressful items reported by children and adolescents on the SLES (Williamson et al., 2003). Four of the BCS70 age 10 questionnaires were reviewed and labelled as the source in the 'From' column of Table 47. They were the Maternal Questionnaire (MQ), Parental Interview (PI), Medical Questionnaire (MeQ) and Educational Questionnaire (EQ), completed by a parent, interviewer, medical professional, and teacher, respectively. Mapping items from multiple raters reduced the risk of single-rater bias.

		Mapped from	To BCS70 variables	From
LE1	Birth of younger sibling	LEQ (1) and HIQ(17)	a4a_5, a4a_9, a4a_13, a4a_17, a4a_21, a4a_25, a4a_29, a4a_33, a4a_37 2nd to 10th person in household, with a relationship code of 13 (younger brother) or 14 (younger sister). Total count, may be > 1	PI
LE2	Serious illness in household	LEQ(2) and HIQ(1-9)	e3_1, e3_2, e3_3 Since age 5, mother, father, or other person in home with severe or prolonged illness. Total count, may be >1	PI
LE3	Father unemployed	LEQ(4)	c2_3	PI

			Father seeking work	
LE4	Mother unemployed	HIQ(14)	c2_11 Mother seeking work	PI
LE5	Death of sibling	LEQ(8) and HIQ(11)	e3_5, e3_12, e3_14, e3_21, e3_23, e3_30 Other person in household with illness = sibling, and outcome of illness = death (code for both = 3). Total count, may be >1	PI
LE6	CM hospital admissions	LEQ(10)	b16_2 Number of child hospital admissions since 5th birthday, overnight or longer	PI
LE7	Parents separated	LEQ(11)	a5_3, a6_3 Living situation change – parents separated (3)	PI
LE8	Parents divorced	HIQ(22)	a5_3, a6_3 Living situation change – parents divorced (4)	PI
LE9	Father died	LEQ(12) and HIQ(10)	a5_3, a6_3 Father died	PI
LE10	Mother died	LEQ(12) and HIQ(10)	a5_3, a6_3 Mother died	PI
LE11	Financial problems	LEQ(13) and HIQ(12)	m126 Free school meals last 12 months LEQ refers to decrease in financial status, HIQ financial problems.	MQ
LE12	Parent in prison	LEQ(15)	a4b_1, a4b_5 Father away from home (1), reason = prison (6)	PI
LE13	Sibling left home for stressful reason	LEQ(18)	a4b_1, a4b_6, a4b_11, a4b_16 Older or younger sibling away (11/12/13/14) AND a4b_5, a4b_10, a4b_15, a4b_20 Reason = institutionalised, hospitalised, in care, in prison, fostered (1/2/5/6/8)	PI
LE14	House move	HIQ(13)	all_1 Number of moves since birth	PI
LE15	CM in foster care	HIQ(21)	a9_2 Number of times in care since birth	PI

LE16	Accidents requiring medical attention	SLES(11)	b18_7 Since age 5	PI
LE17	CM suspended from school	HIQ(26)	j116 During the last school term only	EQ

Table 47. Stressful life events mapped from existing instruments to BCS70 items measured at age 10

It was not possible to specifically map the following LEQ and HIQ items: parental conflict, increased absence of father, parent gets new partner, increased absence of mother, another adult moved into home, and death of (child's) friend, because the data was not available in BCS70. However, most of the items that were not mapped are likely to be correlated with items that were mapped (e.g. parent conflict, moves, and absences are likely to correlate with parent separation and divorce). Death of a child's friend would not be expected to correlate to any of the existing items and could be expected to have a considerable (unmeasured) impact on a child.

A total of 17 stressful life events were identified at age 10 in BCS70 from the maternal, educational, and medical examination questionnaires. The total number of life events was counted and stored in a simple sum score variable: LETot, Range= 0 – 25, Median = 3. Some of the life events were counts in and of themselves, e.g. LE6 was a count of hospital admissions, and LE15 was a count of number of times in care since birth. Each instance was counted as a separate stressful life event, which is why the top of the range of total life events (25) is greater than the number of life event items (17). A histogram is shown in Figure 28.

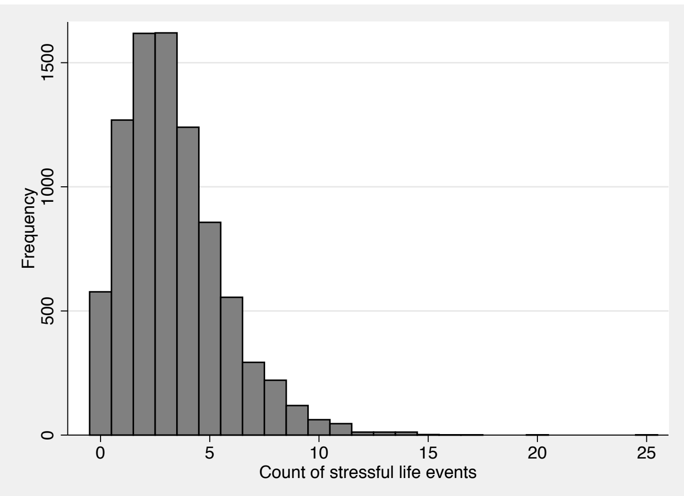


Figure 28. Histogram of stressful life event count (age 10)

Validated measures of stressful life events sometimes have objective weights that are multiplied to create a score, (e.g. Holmes & Rahe, 1967), and/or include a weighting of perceived severity that is provided by the participant (Brown, Sklair, et al., 1973). Adequate data is not available in BCS70 to support either type of weighting. However, a simple count of life events has been found in multiple studies provide similar results to weighted scores (see review: Berden et al., 1990, p. 951). Thus, the simple count was used in the present study.

2.2.5.2 Chronic stressors

Chronic stressors were identified as ongoing problems (i.e. not one-off events). They included problems at school or with health, which would be likely to be perceived as negative and increase stress for the child/cohort member. It is possible that chronic stressors could contribute causally to ADHD, or that ADHD could cause some of the stressors (e.g. being bullied). The items were not derived directly from an existing and validated questionnaire, but were derived based on themes I identified from three of the existing stress measures discussed above in the life events section: HIQ, LER, and SLES:

- teasing from peers (HIQ-24)
- having a visible congenital deformity (LER-3/14)
- school performance problems (SLES-3)
- fights/arguments at school (SLES-4)
- general health problems (SLES-7)
- being bullied (SLES-8)
- increased arguments with parents (SLES-14)

The numbering of the SLES items is from the top 20 most reported stressful items by children and adolescents in (Williamson et al., 2003), and reflects the ranking.

The items identified in BCS70 as chronic stressors related to these themes were mined from three age 10 sweep questionnaires: the educational questionnaire (completed by the teacher; 10EQ), the maternal questionnaire (completed by a parent; 10MQ) and the medical examination form (completed by a medical professional who examined the child; 10MeQ). Here again the use of multiple raters reduces the risk of single-rater bias. There were eight teacher-rated items, seven parent, and seven medical, for a total of 22 (Table 48).

	Chronic stressor description	To BCS70 variables	From
CS1	Teacher reports child's knowledge is extremely limited	j011	EQ
CS2	Child receiving special help at school	j031	EQ
CS3	Teacher reports mother has a hostile attitude towards child	j101	EQ
CS4	Teacher reports father has a hostile attitude towards child	j107	EQ
CS5	Child has unusually high absences from school	j111 z-score > 2	EQ
CS6	Teacher reports child has no friends	j123	EQ

		Top third of VAS	
CS7	Teacher reports child wets pants at school	j130	EQ
		Top third of VAS	
CS8	Teacher reports child soils pants at school	j167	EQ
		Top third of VAS	
CS9	Mother reports child has any medical, behavioural, or educational problem	m13	MQ
CS10	Mother reports child wets bed at night	m21	MQ
CS11	Mother reports child has a stammer	m24	MQ
CS12	Mother reports child has sleep difficulties	m36	MQ
CS13	Mother reports child has great difficulty with maths	m114	MQ
CS14	Mother reports child has great difficulty with reading	m115	MQ
CS15	Mother reports child has great difficulty with writing	m116	MQ
CS16	Medic reports child has squint	meb9	MeQ
CS17	Medic reports child has vision defect that interferes with functioning	meb11_1	MeQ
CS18	Medic reports child has hearing loss that interferes with functioning	meb12_15	MeQ
CS19	Medic reports child has an abnormal appearance	meb21_1	MeQ
CS20	Medic reports child has early puberty	meb22_1	MeQ
CS21	Medic reports child has any disfigurement	meb23_1	MeQ
CS22	Medic reports child has moderate or severe clumsiness	meb33_1	MeQ

Table 48. Chronic stressors mapped to BCS70 at age 10 (based on stress questionnaire themes)

A simple sum score produced a variable (CSTot) with a zero-inflated (Poisson/count) distribution, as shown in Figure 29. Range = 0 – 12, Median = 1.

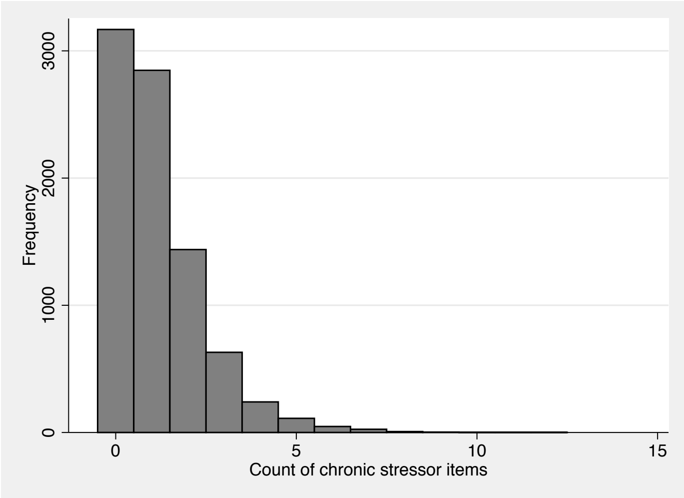


Figure 29. Histogram of chronic stressor count (age 10)

For consistency with the life events count and parsimony, the simple count was also used in the present study as the measure of chronic stressors.

2.2.5.3 Self-esteem

Self-esteem was measured in the BCS70 at age 10 on the pupil questionnaire, completed by the cohort members. The Lawrence Self-Esteem Questionnaire, or LAWSEQ (Butler et al., 1997; Lawrence, 1973, 1978) was the basis for the items. There were 16 questions, but only 12 are part of the scale; items 4, 7, 9, and 12 were distractors. Possible responses were Yes, No, and Don't know.

Don't know (DN) responses have sparked considerable debate in in psychometrics, because the meaning and interpretation is not straightforward. It has been reported that DN responses are more likely for questions that have a sensitive nature, a relationship to education level, and for female respondents (Durand & Lambert, 1988; Young, 2012). Accordingly, explicit DN options are often omitted from questionnaires. However, the present analysis is constrained by the available data, so assumptions need to be made about the treatment of the collected DN responses.

It is not clear what the child answering actually intended with the 'don't know' answer (Durand & Lambert, 1988). With this self-esteem scale, if 'don't know' were to be scored with a '1', as suggested in the Rae et al. (2011) study, it would indicate that the respondent had more self-esteem than a respondent who answered Yes, but less than someone who answered No. This is probably not a valid assumption. Both methods of scoring were implemented and evaluated, and the distribution of the method that excluded don't know responses was closer to normal. So, in the present research, 'don't know' answers were not assigned a score, but instead treated as missing data. Thus, values were recoded to 0 for Yes, and 1 for No, except item 1 ('parents listen to your ideas') was reversed. Higher scores indicated higher self-esteem.

Descriptive statistics of the 12 item LAWSEQ scale recoded with 'don't know' as missing had an approximately normal distribution based on a histogram, p-p plot and q-q plot. The scale had just-acceptable internal reliability, Cronbach's $\alpha = 0.72$. Confirmatory factor analysis of a tetrachoric correlation matrix for the items indicated unidimensionality with a single dominant factor. Tetrachoric correlations of standard errors indicated that item 11 ('feel foolish with parents') was not locally independent (rho >0.05). A Mokken's rule test indicated item 1 was very low on Loevinger's H (0.13), and items from a strong scale should have H >= 0.30 (Hardouin et al., 2011). An item response model was fitted, and ICC curves indicated that item 1 did not discriminate well between low and high levels of self-esteem. Thus, items 1 and 11 were removed, and descriptive statistics assessed again. The 10-item scale was also approximately normally distributed, with internal reliability still at $\alpha = 0.72$. The IRT assumption of unidimensionality held, and local independence was met. IRT ICC curves were all strongly 'S' shaped, indicating an informative relationship between the probability of a 'correct' (or positive) response, and the level of the latent construct measured; in this case, self-esteem. This adapted 10-item scale had better psychometric properties than the 12-item scale, so the 10-item scale was used for the measure of self-esteem in subsequent analyses.

Vname	BCS70 LAWSEQ Item (abbreviated)
SE2	k011-feel lonely at school
SE3	k012-people fall out with you
SE4	k014-people say nasty things about you
SE5	k015-feel shy with teacher
SE6	k017-feel sad that no playmates
SE7	k019-want to change yourself
SE8	k020-feel foolish with peers
SE9	k022-feel foolish with teacher
SE10	k023- often have to find new friends
SE12	k025-other people think you lie

Table 49. Subset of 10 LAWSEQ items used to score self-esteem

Vname = variable name, in Stata

To better account for the variation in contribution of information of each of the items, a 2PL IRT model was used to predict a score. The latent trait (theta) was predicted using empirical Bayes means, and named SEtheta (N = 8,519, Range = -2.48 – 1.29, M = 0.01, SD = 0.82).

2.2.5.4 Locus of control

The second measure of a protective factor from stress was locus of control. Locus of control was measured in the BCS70 age 10 sweep on the pupil questionnaire, completed by the cohort members. The CARALOC scale (Gammage, 1974, 1982) was used, consisting of 20 items. Possible answers were 'Yes', 'No', and 'Don't know'. Following an approach like the one taken in Murasko, (2007) five distractor items were removed (4, 7, 11, 15, 19), and the remaining 15 items were recoded to 'no' = 1, 'yes' = 0, (except item 10, which was reversed). All values of 'don't know' were treated as missing, like the approach for the self-esteem items. A sum score was then calculated, which had a range of 0-15 and a normal distribution (M=7.23, SD = 2.91). This adapted scale was close to acceptable internal reliability ($\alpha = 0.67$), according to a rule of thumb that a minimum of 0.70 is preferable (Tavakol & Dennick, 2011).

Vname	BCS70 CARALOC items (abbreviated)
LoC1	k075-not worth trying
LoC2	k076-wishing helps
LoC3	k077-people are good to you
LoC4	k079-useless to try
LoC5	k080-high mark is luck
LoC6	k082-tests are guesswork
LoC7	k083-blamed for things
LoC8	k084-believe in planning (reversed)
LoC9	k086-bad things are others' fault
LoC10	k087-making friends is impossible
LoC11	k088-nice things are luck
LoC12	K090-arguments are others' fault
LoC13	k091-surprised by teacher praise
LoC14	k092-get low marks even when you study
LoC15	k094-studying for tests is a waste of time

Table 50. Non-distractor CARALOC items used to score Locus of Control

Tetrachoric correlations indicated a single dominant factor, and standard errors were not correlated > 0.05, so the independence assumption was met. To better account for the variation in contribution of information of each of the items, and for consistency with the self-esteem measure, a 2PL IRT was modelled, and all item characteristic curves indicated a reasonable fit. A latent trait (theta) score was predicted using empirical Bayes means, and named LoCtheta (N = 8,519, Range = -2.15 – 2.03, M = 0.03, SD = 0.85). This measure was used in subsequent analyses.

N.B. All variables that were scored based on data at age 10 (i.e. z-scores and IRT theta scores) were modelled in the full (preferred) age 10 dataset (N = 11,426). Descriptive statistics (see Table 54) however only report the distributions within the preferred sample (N = 8,519)

2.2.5.5 Engagement in leisure activity

As discussed in chapter 3, engagement in leisure activity is thought to function as a protective factor against childhood difficulties/stressors, so was included in my operationalisation of state regulation theory. The measure was based on a section about leisure activity on the age 10 maternal self-completion questionnaire. Parents were asked if their child participated in various leisure activities, and answer choices were never, sometimes, or often. The list included 14 items including “plays sports”, “rides a bicycle”, and “plays a musical instrument” (Butler et al., 1997; Centre for Longitudinal Studies: UCL/IoE, 2019). The items were not documented as belonging to a scale, and some items were not necessarily likely to correlate with each other (e.g. sports and music), but could be scored and used to estimate a relative ‘amount’ of engagement in leisure activity for the cohort. Three items were excluded because they could be associated with academic ability and/or would not involve active engagement: m86-reads books, m88-watches television, and m95-goes to library. The remaining items are listed in Table 51.

Vname	BCS70 spare/leisure time activities
P4	m84-plays sports
P5	m85-listens to records
P6	m87-rides a bicycle
P7	m89-goes to a club or organisation
P8	m90-goes for walks
P9	m91-goes to the cinema
P10	m92-listens to the radio
P11	m93-goes to a museum (any kind)
P12	m94-goes swimming
P13	m96-plays a musical instrument
P14	m97-plays with constructional toys (e.g. LEGO)

Table 51. Items used to measure leisure activity

The ratings of never, sometimes, and often, were scored 0, 1, and 2 respectively. As a set the items had poor scale reliability, Cronbach’s $\alpha = 0.54$, which is not surprising given it was not intended to be a scale, and the types of activities represented in the items are quite different from each other (e.g. sports and music). Nonetheless, when items were aggregated into a sum score, the result had a roughly normal distribution, so it was standardised into a z- score (N

= 8,505, Range = -3.65 – 3.10, M = 0.04, SD = 0.98) to provide a relative measure of an 'amount' of engagement in leisure activity within the cohort.

2.2.6 Covariates that are predictors of ADHD

As discussed in the refined literature review of ADHD predictors, I identified eight new variables to use as predictors, adding to the 17 identified in chapter 5, for a total of 25. This set was used to select subsets for matching and regression analyses.

2.2.6.1 Selection of covariates for matching

First, a simple regression on ADHD as the outcome was fitted and used to estimate variance inflation factors (VIF), as a check for collinearity. All the VIFs for the 25 variables were small (between 1.0 and 1.5), except for F0PregSmoke (smoking during pregnancy) and F0SmokeLevel (the number of cigarettes smoked during pregnancy). Those two had VIFs of 7.3 and 7.4 respectively, which is under one recommended threshold of 10 (O'Brien, 2007) as an indicator of collinearity, but above another recommendation for a threshold of 2 (Allison, 2018). I decided to keep all 25 based on the higher threshold, and because the smoking indicator and level could have different effects.

Next, in order to simplify the analysis and focus on a smaller number of the most significant predictors or outcomes (Garrido et al., 2014), I used a variable selection procedure called vselect (Lindsey & Sheather, 2010), with the 'best' option. This option uses the leaps and bounds algorithm (Furnival & Wilson, 1974) to identify regression models with the best-fit subsets of variables using residual sums of squares. The procedure reports all the best-fit models with several fit indices, including adjusted R^2, AIC, and BIC[39], and the user can select the preferred variable set based on these. This procedure is proposed as more reliable than forward or backward stepwise regression, results of which can be biased by the order of variables (Lindsey & Sheather, 2010). Three separate vselect procedures were executed: one with ADHD as an outcome, and again with each of the other two outcomes of interest.

2.2.6.1.1 ADHD as the outcome (adhd_sg)

Stata syntax used for the procedure:

```
vselect adhd_sg sex10b F5DadEd F5MumMal F0SmokeLevel F5HomeLS
F0Unmarried F0LBW F5MumEd F5HomeUntidy F5PoorNbhd F0PregSmoke
F0PreTerm F0PostTerm F0Induced F5AuthCRV F0PreEcInd F5HVOCIQ
F5SEPVT F5Wheez F5HayF F5Ecz F5S1Pr1st6 F5MumSep1m F5BFed
F0PMisc, best
```

---

[39] BIC = Bayesian Information Criterion; AIC = Akaike Information Criterion

The output reported optimised variable subsets for between 1 and 25 variables. No model was optimal according to all five fit indices. The BIC was optimised with 5 variables, the AIC with 10, and R^2 with 12, and the difference between R^2 values for the 10 and 12 variable solutions was very small. BIC is often the preferred model selection index (Nylund et al., 2007), but BIC favours simpler models and makes more stringent adjustments for sample size (Heinze et al., 2018). As discussed previously, when identifying matching covariates, it is desirable to minimise the number of predictors to allow for probable matches but also to include all of the most important predictors. To balance between these opposing objectives, the middle option with optimal values for AIC, AICC, and Mallow's C was selected, i.e. the 10-covariate solution.

10 covariate solution (AIC = -2536.41, Adj $R^2 = 0.04$)

```
F5HVOCIQ sex10b F5MumMal F5HomeLS F5DadEd F0SmokeLevel F5SEPVT
F0PregSmoke F5Wheez F5PoorNbhd
```

For comparison, a backward stepwise regression was also fitted using p<0.20 as a significance threshold (Maldonado & Greenland, 1993). Stata syntax:

```
stepwise, pr(.2): logit adhd_sg sex10b F5DadEd F5MumMal
F0SmokeLevel F5HomeLS F0Unmarried F0LBW F5MumEd F5HomeUntidy
F5PoorNbhd F0PregSmoke F0PreTerm F0PostTerm F0Induced
F5AuthCRV F0PreEcInd F5HVOCIQ F5SEPVT F5Wheez F5HayF F5Ecz
F5SlPr1st6 F5MumSep1m F5BFed F0PMisc
```

The logit model identified the following set of 11 variables (N= 4,056, χ^2 =128.93, df = 11, p < .001, McFadden's $R^2 = 0.11$)

```
F5DadEd sex10b F5MumMal F0SmokeLevel F5HomeLS F5SEPVT F5HVOCIQ
F5HayF F5PoorNbhd F5Wheez F0PregSmoke
```

The variable list was the same except for the addition of the F5HayF (hayfever) variable. Because of the optimal AIC in the vselect procedure, and general recommendations against stepwise, I decided to use the 10 variables from the vselect procedure. For consistency, I used the vselect-based variable list with optimised AIC for the other two outcomes as well.

Next, the same vselect syntax was used for each of the two outcomes of interest, and variable lists were selected based on the optimised AIC.

2.2.6.1.2 Subjective wellbeing as the outcome (SWB)

11 covariate solution (AIC = 10821.95, Adj $R^2 = 0.02$):

```
F5SEPVT F5MumMal F0SmokeLevel F5HomeLS F5DadEd F0PregSmoke
F0Unmarried F5Wheez F5HVOCIQ F5Ecz F5MumSep1m
```

2.2.6.1.3 Education level as the outcome (OEEduc5)

15 covariate solution (AIC =18944.44, Adj R^2 = 0.18):

```
F5DadEd F5MumEd F5SEPVT sex10b F5BFed F5AuthCRV F5HomeUntidy
F5SlPr1st6 F5HVOCIQ F0PostTerm F5Ecz F5HomeLS F5Wheez
F0Unmarried F0SmokeLevel
```

Matching is recommended on covariates that impact BOTH membership in the treatment group AND the outcome of interest (Rosenbaum & Rubin, 1983; Stuart, 2010). So, lists were compared and de-duplicated, and the covariates that appeared in both the ADHD list and each of the two outcomes lists are summarised in Table 52.

ADHD and SWB	ADHD and OEEduc
F0PregSmoke	F0SmokeLevel
F0SmokeLevel	F5DadEd
F5DadEd	F5HomeLS
F5HomeLS	F5HVOCIQ
F5HVOCIQ	F5SEPVT
F5MumMal	F5Wheez
F5SEPVT	sex10b
F5Wheez	

Table 52. Covariates identified as optimal matching predictors based on combined relationship to ADHD and each separate outcome

The two lists were nearly the same. In the interest of parsimony, one set of nine variables covering the union of the two sets was used and matching was done once.

Variable name
F0PregSmoke
F0SmokeLevel
F5DadEd
F5HomeLS
F5HVOCIQ
F5MumMal
F5SEPVT
F5Wheez
Sex10b

Table 53. Optimal matching variables

2.2.7 Descriptive statistics summary for measures

In this section, a full summary of all measures used in Chapter 6 including distributions, missingness, and variable types is provided in Table 54. The table is split across two pages and does not follow APA formatting standards (i.e. gridlines are included) to support readability. Items 1-28 are in the first part of the table, and 29-56 in the second. There are two sets of variable names: one for Mplus and another for Stata[40].

[40] Mplus requires variable name to be 8 characters or less, whilst Stata does not. I had to rename most of the variables, so I developed a new convention to make the names as intuitive as possible.

No.	Mplus var	Stata var	Description	Age	N	M	SD	Min	Max	% Msng	Type
1	induced	F0Induced	Indicator mother's labour was induced	0	7967	0.26		0.00	1.00	6.48%	Bin
2	lbw	F0LBW	Low bith weight indicator (< 2500g)	0	7982	0.06		0.00	1.00	6.30%	Bin
3	mumage	F0MAge	Mother's age at birth	0	7948	25.94	5.36	14.00	50.00	6.70%	Scl
4	postterm	F0PostTerm	Indicator child was born post-term	0	6566	0.10		0.00	1.00	22.93%	Scl
5	preeclamp	F0PreEcInd	Indicator there were sympotms of pre-eclampisa during pregnancy/birth	0	7757	0.19		0.00	1.00	8.94%	Bin
6	preterm	F0PreTerm	Indicator child was born pre-term	0	6566	0.04		0.00	1.00	22.93%	Bin
7	prevmisc	F0PMisc	Indicator mother had previous miscarriages	0	7989	0.18	0.52	0.00	7.00	6.22%	Bin
8	smoke	F0PregSmoke	Indicator mother smoked during pregnancy	0	7952	0.40		0.00	1.00	6.66%	Bin
9	smokelvl	F0SmokeLevel	Mother's level of smoking during pregnancy	0	7952	0.85	1.13	0.00	3.00	6.66%	Cat
10	unmar	F0Unmarried	Mother unmarried at child's birth	0	7982	0.05		0.00	1.00	6.30%	Bin
11	backward	F5HVOCIQ	Child's development rated as slightly or definitely 'backward' by health visitor	5	6930	0.03		0.00	1.00	18.65%	Bin
12	bfed	F5BFed	Child was breastfed (mother recall)	5	7204	0.38		0.00	1.00	15.44%	Bin
13	daded	F5DadEd	Father's education level	5	6561	1.51	1.78	0.00	5.00	22.98%	Cat
14	ecz	F5Ecz	Indicator - mother-reported child eczema	5	6959	0.13		0.00	1.00	18.31%	Bin
15	epvt	F5SEPVT	English picture vocabulary test score	5	6808	0.07	0.97	-3.04	3.04	20.08%	Scl
16	hayfev	F5HayF	Indicator mother-reported hayfever	5	6929	0.04		0.00	1.00	18.66%	Bin
17	homelow	F5HomeLS	Child's home rated 'low standard' by health visitor	5	7118	0.03		0.00	1.00	16.45%	Bin
18	homeunti	F5HomeUntidy	Untidy home per health visitor	5	7116	0.06		0.00	1.00	16.47%	Bin
19	mumed	F5MumEd	Mother's education level	5	6928	1.00	1.34	0.00	5.00	18.68%	Cat
20	mummal	F5MumMal	Indicator mother had malaise/depression	5	7122	0.22		0.00	1.00	16.40%	Bin
21	poornbhd	F5PoorNbhd	Child's home neighborhood rated as poor by health visitor	5	7017	0.06		0.00	1.00	17.63%	Bin
22	sepfmum	F5MumSep1m	Child was speարated from mother for 1 month or more	5	7255	0.04		0.00	1.00	14.84%	Bin
23	slpoor6	F5SlPr1st6	Child had problems sleeping in 1st 6 months since birth (mother recall)	5	7162	0.14		0.00	1.00	15.93%	Bin
24	wheez	F5Wheez	Indicator - mother reported child wheezing	5	7176	0.20		0.00	1.00	15.76%	Bin
25	adhd	adhd_sg	ADHD subgroup indicator (0=non-ADHD, 1=ADHD)	10	8519	0.04		0.00	1.00	0.00%	Bin
26	adhdr	adhd_sgR	ADHD indicator, based on chapter 4, reversed (if needed for ORs)	10	8519	0.96		0.00	1.00	0.00%	Bin
27	adhdsev	ADHDness	Factor score of ADHD severity	10	8519	-0.12	0.89	-1.15	2.91	0.00%	Scl
28	authcrv	F5AuthCRV	Authoritarian child-rearing view (provided by parent)	10	7203	0.11		0.00	1.00	15.45%	Bin

No.	Mplus var	Stata var	Description	Age	N	M	SD	Min	Max	% Msng	Type
29	cstress	CSTot	Count of chronic stressors	10	8519	1.14	1.27	0.00	12.00	0.00%	Cou
30	estress	LETot	Count of stressful life events	10	8519	3.35	2.32	0.00	25.00	0.00%	Cou
31	loctheta	LoCtheta	Factor score of locus of control (adapted CARALOC scale)	10	8519	0.03	0.85	-2.15	2.03	0.00%	Scl
32	lowses	A10FSM	Indicator of childhood low SES, based on free school meals received in the last 12 months	10	8494	0.14		0.00	1.00	0.29%	Bin
33	maths	zBD3MATHS	z-score of maths test score	10	7821	0.07	0.98	-3.55	2.28	8.19%	Scl
34	setheta	SEtheta	Factor score of self-esteeem (adapted LAWSEQ)	10	8519	0.01	0.82	-2.48	1.30	0.00%	Scl
35	sex	sex10b	Sex (0=girls, 1-boys)	10	8519	0.49		0.00	1.00	0.00%	Bin
36	subtyp	dsmsubtype	DSM-5 ADHD subtype	10	369	2.08	0.92	1.00	3.00	95.67%	Cat
37	zext	z_tEScoreM	z-score of externalising behaviour rated by mother	10	8518	-0.06	0.93	-0.60	4.68	0.01%	Scl
38	zint	z_tIScoreM	z-score of internalising behaviour rated by mother	10	8517	-0.01	1.00	-0.77	3.57	0.02%	Scl
39	zplay	z_PScoreM	z-score of mother-rated engagement in play	10	8505	0.04	0.98	-3.65	3.10	0.16%	Scl
40	zread	zAge10RAge	z-score of child's reading age	10	6787	0.09	0.97	-3.24	1.96	20.33%	Scl
41	aced5	OEEduc5	Adult education level of academic qualifications (5 levels collapsed from 8 in data)	42/34	8519	2.12	1.67	0.00	5.00	0.00%	Cat
42	nvq	OEEDL	Adult education level on the 5-level NVQ scale	42/34	8516	2.72	1.46	0.00	5.00	0.04%	Cat
43	oalcgrp	OEAlcGrp	Indicator of problems with alcohol	42/34	6902	1.31	0.60	0.00	4.00	18.98%	Cat
44	odis	OEDis	Disability indicator	42/34	8468	0.14		0.00	1.00	0.60%	Bin
45	oedacvoc	OEEducAcVoc	Derived educational attainment level, using 5 academic levels plus vocational	42/34	8516	2.29	1.49	0.00	5.00	0.04%	Cat
46	oedacvocr	OEEducAcVocR	Adult education level including academic and vocational	42/34	8516	2.71	1.49	0.00	5.00	0.04%	Cat
47	oeduc	OEEduc	Adult education level on academic qualifications scale (8 levels reported in BCS70)	42/34	8519	3.44	2.78	0.00	8.00	0.00%	Cat
48	oeduc5r	OEEduc5R	Adult education level collapsed to 5 categories and reversed (if needed for ORs)	42/34	8519	2.88	1.67	0.00	5.00	0.00%	Cat
49	ohlth	OEGHlth	Adult self-rated general health	42/34	8501	2.33	1.05	1.00	5.00	0.21%	Ord
50	olwpart	OELwPart	Adult indicator they are living with a partner	42/34	8519	0.72		0.00	1.00	0.00%	Bin
51	osoc	OESocial	Adult social class of job (1970, broad categories)	42/34	7130	2.66	0.88	1.00	5.00	16.30%	Cat
52	swb	SWB	Factor score of subjective wellbeing (Warwick WB, Rutter Malaise, and life satisfcation)	42/34	8519	0.00	0.95	-4.63	2.47	0.00%	Scl
53	warwick	O42WarwickWB	Warwick-Edinburgh Mental Wellbeing Scale score	42	5981	49.26	8.26	14.00	70.00	29.79%	Scl
54	cem_m2	cem_matched_swb2	Indicator used for exact-matched sample	D	8519	0.08		0.00	1.00	0.00%	Bin
55	cem_w	cem_weights_swb	Weights used for matched weighted sample	D	8519	0.73	1.35	0.00	35.28	0.00%	Scl
56	idn	idcounter	Row id. Cannot use BCSID in Mplus because it is not numeric	D	8519	4260.00	2459.37	1.00	8519.00	0.00%	Seq

Table 54. Descriptive statistics of measures used in Chapter 6 (items 1-56)

SD is not reported for binary variables
Missing shading key if in colour/on screen: <2% green, 2-10% yellow, > 10% red
Variable types legend: Bin = binary, SCL = scale, or continuous, Cat = categorical, Cou = count, Ord = ordinal, Seq = sequence number

N.B. See chapter 6 appendix for two correlation tables: core variables, and a subset of significant variables

2.3 Identification of matched samples

In chapter 5 that the 'exact' option was used with Coarsened Exact Matching, which created a sample with an equal number of ADHD and control group observations that were most similar on the matching covariates. This resulted in dramatic data loss and reduced the sample size to N=546. A learning point from the pilot and further study of the literature indicated that this degree of data loss could actually increase bias, which is the opposite of the objective. The remediation plan was to repeat the matching using the weighted match option and utilise as much of the data available as possible.

A pre-matching analysis was done on the treatment group (ADHD) and controls to assess balance using the nine variables identified in the previous section. It was not possible to achieve good balance with the F5SEPVT variable (English Picture Vocabulary Test score from age 5) included. This indicated the pattern of responses for F5SEPVT was significantly different for ADHD vs. non-ADHD when stratified by the other variables. There was another very crude indicator of intelligence in the set: F5HVOCIQ (health visitor reported child's development as 'backward' at age 5), so, I dropped the F5SEPVT variable and relied on F5HVOCIQ as a rough intelligence stratum for matching purposes. The matched samples were derived using the eight remaining variables.

2.3.1 Weighted match

The procedure used to create a weighted matched sample was:

```
cem sex10b F0PregSmoke F0SmokeLevel F5DadEd F5HomeLS F5HVOCIQ
F5MumMal F5Wheez, treatment(adhd_sg)
```

This procedure used the default CEM behaviour of matching missings to missings. The resulting multivariate distance metric was ideal (L1 = 0.00). 739 strata were identified, and 143 strata were matched. 333 of the 369 in the ADHD subgroup were retained in the treatment group, and 5,874 weighted observations were retained in the control group (total N = 6,207). The weights variable was renamed cem_weights_swb (so it would not be overwritten by future procedures) and saved for use in MLR regression models.

2.4 Missingness, assumptions, and considerations with Mplus

2.4.1 Missingness

Most of the covariates selected for use in regression analyses had missing values. Hence, a missing data analysis was needed.

The covariates to be used in the SWB procedures include a set of variables related to my hypotheses about ADHD and stress, plus those from literature that were identified based on reviews and vselect procedures.

Two main predictors

ADHD and ADHD severity

Four variables based on the hypotheses that stress has a negative effect on outcomes and protective factors against stress have a positive one.

Life events, chronic stressors, locus of control, self-esteem

Five variables representing factors indicated in previous reviews as affecting outcomes for ADHD (Brassett-Grundy & Butler, 2004; Costello & Maughan, 2015)

Free school meals (low SES), maths, externalising problems, internalising problems, sex

All of the above were measured at age 10, and comprise the set of variables referred to in subsequent analyses as 'core'.

Next, 11 potential confounders were added from the ADHD predictors literature review, reduced to a smaller set expected to effect SWB using vselect (seven of these, underlined, were also used in the matching procedure)

Age 5: English picture vocabulary test, mother malaise, low standard home, father's education level, backward development, separated from mother for 1 month or more, wheezing problems, eczema problems

Age 0: mother's smoking level during pregnancy, mother smoked during pregnancy indicator, mother unmarried at birth

One additional confound for ADHD, based on the vselect procedure using ADHD as the outcome.

Age 5: poor neighbourhood

A total of 24 variables were identified as relevant to the SWB analysis. Ten of them had approximately 0% missing data. Missing data patterns were analysed in Stata for the remaining 14.

```
misstable patterns zBD3MATHS A10FSM F5SEPVT F5MumMal
F0SmokeLevel F5HomeLS F5DadEd F0PregSmoke F0Unmarried F5Wheez
F5HVOCIQ F5Ecz F5MumSep1m F5PoorNbhd
```

60% of observations had complete data. There were four patterns that represented >2% of observations missing. Sainani (2015) recommends that >2% missingness should be remedied using a more robust method than deletion or single mean imputation.

10% - all age 5 data missing (F5*)
6% - only father's education level from age 5 missing (F5DadEd)
4% - all missing except age 10 free school meals indicator (all age 0 and age 5 data missing)
3% - all data present except the age 5 English picture vocabulary test score (F5SEPVT)

These patterns are similar to the ones found in the smaller pilot dataset. Similarly, the 10% and 4% patterns most likely represented non-response at age 5, and immigrant children born in the BCS70 birth week who were recruited after age 5, respectively. This leaves missingness for father's education level and the age 5 picture vocabulary test to examine.

Given the rich data available in BCS70, it is likely that there are other variables that correlate with the missingness of father's education level and the vocabulary test score well enough to support an assumption of Missing at Random (MAR). I suspected that the missing data on father's education level could be related to low SES, because of absent fathers or fathers with a low education level that they didn't want reported. I suspected the missing vocabulary test data could also be related to low SES, through increased school absence. Variables were created to represent missingness in father's education level and the English Picture Vocabulary test, and evaluated in Mirador (Sabeti Lab at Harvard University et al., 2018) software to explore how other variables in my dataset related to the missingness variables. Mirador is a tool that calculates a similarity score to indicate the statistical relationship between all variables of all types (i.e. continuous or categorical) in a dataset, and ranks them in order by size of the relationship (Andres, 2014). The most related variables to father's education level missingness in childhood were financial problems and total stressful life events count. The most related variables to vocabulary test missingness were total stressful life events count (age 10), and low standard home (from age 5). All of these variables were already in the set of covariates I am evaluating for SWB (financial problems were measured using the free school meals/low SES indicator). Thus, the variables already included in the analysis should support an assumption of MAR and FIML computations for the moderate to large patterns of missingness.

2.4.2 Assumptions

In the action plan based on the chapter 5 pilot, it was noted that analyses should be moved to Mplus to accommodate the complexity of the data. However, before moving to Mplus, I evaluated assumptions in Stata. The full set of covariates was tested for heteroscedasticity using `regress` and the `estat` and `hettest` postestimation commands. All χ^2 values were significant (p < 0.01), indicating heteroskedasticity was a problem in this set of covariates. Thus, standard errors, test statistics and confidence intervals using SEM could be expected to be biased (Williams, 2015a).

Multivariate normality was tested on the full set of potential covariates using `mvtest`. The default Doornik-Hansen test indicated the null hypothesis should be rejected, i.e. the covariates were not multivariate normal ($\chi^2(32) = 783{,}000$, $p < 0.001$).

Stata procedures do not support this combination of missingness with violations of homoscedasticity and multivariate normality assumptions. Therefore, the analyses were moved to Mplus 8.2 (Muthen & Muthen, 2017), which can handle complex data like these. The relevant variables for subsequent analyses were identified and exported to Mplus data format using a user-written package (UCLA: Statistical Consulting Group, n.d.), as follows:

```
stata2mplus idcounter sex10b zBD3MATHS zAge10RAge adhd_sg
dsmsubtype ADHDness z_PScoreM LoCtheta SEtheta z_tEScoreM
z_tIScoreM A10FSM CSTot LETot O42WarwickWB F0Unmarried
F0PregSmoke F0PreTerm F0PostTerm F0Induced F0LBW F0SmokeLevel
F0PreEcInd F0PMisc F0MAge F5MumEd F5DadEd F5HomeLS
F5HomeUntidy F5PoorNbhd F5MumMal F5AuthCRV F5HVOCIQ F5BFed
F5MumSep1m F5SlPr1st6 F5Ecz F5HayF F5Wheez F5SEPVT OEEduc
OESocial OEGHlth OELwPart OEDis OEAlcGrp SWB cem_weights_swb
cem_matched_swb2 OEEDL OEEduc5 adhd_sgR OEEducAcVoc OEEduc5R
OEEducAcVocR using
"/Users/cottonjm/Documents/Cambridge/PhD/Thesis/Mplus/Take3/MO
AF2"
```

2.4.3 Considerations with Mplus

Moving to Mplus complicated the analysis process. An additional copy of the dataset was needed in the Mplus format (see above). Each model had to be defined in a separate input file, and each file had to include the full file structure definition. This was more cumbersome to manage than Stata do-files and increased the likelihood of copy-and-paste, text input, or other human errors. Also, Mplus does not provide a data browsing facility, so it was not possible to visually sense-check the data without transferring the data back to Stata (or another software package). However, these risks were all accepted because Mplus has robust estimation methods that can handle complex missing data patterns and violations of multivariate normality [41].

---

[41] Another option would have been to move the analyses to R, which can also accommodate the complex characteristics of the data and would have the added benefits of allowing code consolidation into a smaller number of files, no-need for repeated file definition, and a built-in data browser. However, R has a substantial learning curve, so this was not possible within my time constraints. If analysis on this data is continued post-doc, , use of R will be re-considered.

2.4.3.1 Estimation

Mplus documentation recommends the use of MLR or BAYES estimators for analysis of datasets where there is missingness and a mix of continuous and categorical covariates (like we have here), and advises that the BAYES estimator may be more robust, particularly when there are non-normal covariates with missingness and/or small sample sizes (Muthén & Schultzberg, 2017; Muthén & Muthén, 2017). However, the BAYES estimator does not support the use of weights, which were required for use of the matched sample. Several models were fitted with both MLR and BAYES estimators and compared, but only the MLR estimator results were reported in the results section. Some BAYES models on unweighted samples were included as robustness checks in the appendices for chapters 7 and 8.

2.4.3.2 MLR and model fit

In Mplus, if the dependent variable is continuous, MLR estimates linear regression, and if it is categorical, MLR estimates logistic regression (Muthén & Muthén, 2017). SWB is continuous, whilst EDL was modelled here as categorical.

BIC was reported and used as a comparative fit index when there were multiple models with the same dependent variable. A smaller BIC indicates a better fit. AIC was not reported with models because BIC is similar and has the advantage of being adjusted for sample-size. In several models the 'absolute' fit indices were perfect (RMSEA and SRMR = 0; CFI and TLI = 1) [42]. Per the Mplus discussion forum (Muthen & Muthen, 2007), if this happens when a model is just-identified, i.e. degrees of freedom (df) and chi-square ($\chi 2$) are both zero, the indices cannot be used to evaluate fit. The authors also note this result is not unusual when models only estimate a regression (or path model with no measurement model) using the Mplus SEM framework and the MLR estimator (Muthen & Muthen, 2007). Also, a non-significant chi-square is not a reliable indicator of fit for most of the models because the sample sizes are too large (> 400; Kenny, 2015). Where model fit could be evaluated, cut-offs for acceptable fit were applied based on widely-referenced simulation studies as follows: RMSEA <= 0.06, SRMR <= 0.08, CFI and TLI >= 0.95 (Hu & Bentler, 1999).

[42] BIC = Bayesian Information Criterion; AIC = Akaike Information Criterion; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardised Root Mean Square Residual; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index

2.4.3.3 File management and variable name changes

Since numerous (200+) Mplus files were created for and generated by these analyses, a file-naming convention was defined to help keep them organised. The convention is documented in the appendix for chapter 6.

Mplus has an eight-character limit for variable names. Stata does not have this limit, so several variable names in Stata were longer than 8 characters and needed to be renamed for Mplus. For example, z_tEScoreM was the Stata variable name for the z-score of externalising items as rated by the mother (at age 10) and was renamed to zext. The full list of variables from Stata mapped to new names in Mplus is included with the variable descriptions (Table 54) in the Measures section. A text comparison tool called UltraCompare (IDM Computer Solutions, 2019) was used to check a random selection of files to ensure the file definition text was the same across files (to minimise risk or error from copy, paste, and edit).

3 Summary and conclusion

To conclude, chapter 6 documented the pilot-inspired methodological improvements to the sample, measures, and statistical procedures that were used to analyse the relationship between ADHD and long-term outcomes in chapters 7 and 8. The improvements included a more rigorous literature review to identify ADHD predictors, extension of the analysis sample by adding outcomes data from the age 34 BCS70 sweep, comparison of the sample to the UK 2011 census as an external reference point, refinement of wellbeing and educational attainment measures, derivation of new measures for stressors and protective factors, and a software switch from Stata to Mplus to better accommodate complex missing data and violated multivariate normality.

Chapter 7 Childhood ADHD, stress, and adult subjective wellbeing

1 Introduction

Chapter 7 starts by examining the relationship between the measures of ADHD derived in chapter 4, and the measures of stress and protective factors against stress developed in chapter 6. The analysis is a test for evidence for or against State Regulation theory.

Next, the main section of the chapter answers research questions about the relationship between childhood ADHD and adult subjective wellbeing (SWB). A comparison is made between four approaches to estimating effects of ADHD: a naïve effect, controlled regression, matched sample effect, and matched sample effect controlling for additional factors, including stressors, protective factors, and a carefully selected set of potential confounds.

1.1 Brief recap of learning points incorporated from pilot

Learning from the pilot study in chapter 5 recommended several changes to methods and analysis procedures, some of which were described fully in chapter 6. Some further procedure changes were implemented in chapter 7. To recap briefly, these included:

- Use of robust regression estimation procedures in Mplus to address missing data and violated assumptions (MLR and FIML);
- Comparison between controlled regression and matching, using the same set of covariates;
- Separate models fitted where appropriate for girls and boys;
- A robustness check of the SWB measure by comparing selected outcome findings for SWB and the previously validated Warwick Edinburgh Mental Wellbeing Scale (WEMWBS) for cohort members with both measures available. Results are noted at the end of chapter 7 and analysis included in the appendix.

1.2 Research questions for chapter 7

The following research questions defined in chapter 3 are answered in chapter 7:

> RQ2: How do chronic stressors, life event stressors, locus of control, and self-esteem relate to ADHD and ADHD severity, all as measured at age 10? Does the relationship provide evidence to support state regulation theory?
>
> RQ3: What is the effect of childhood ADHD on adult wellbeing using different methods and covariate sets to estimate treatment effects? Do the results support state regulation theory?

Some discussion is included with results of each section, and a broader summary and discussion of results can be found after the full set of analyses.

2 ADHD, stress, and protective factors

I proposed a hypothesis based on SR theory that stressors should relate positively to ADHD, and protective factors against stress, defined in this study as locus of control, self-esteem and engagement in leisure activities, should have a negative relationship to ADHD.

Since all of these measures were taken from the age 10 sample, and item-response models were used to estimate theta values for those with missing item values, there is minimal missing data in this set of variables. Thus, the analysis was straightforward, and handled in Stata.

2.1 ADHD subgroup indicator

Logistic regression was used to predict the binary ADHD indicator with chronic stressors count (CSTRESS), life event stressors count (ESTRESS), locus of control score (LOCTHETA), self-esteem score (SETHETA), and engagement in leisure score (ZLEIS). I started with the full age-10 sample (N = 11,426) for this model, since all the variables were measured at age 10. 24 were missing data on , but this was a small percentage for a single variable, so the observations were deleted leaving N = 11,402. All the predictors except ZLEIS were significant at a 99.9% confidence interval, and in the direction expected. Using a recent guideline for interpreting odds ratios (ORs) as effect sizes (ES)[43], the effects of CSTRESS (*OR* = 1.556***) and SETHETA (Inv OR[44] = 1.377***) were small but appreciable, LOCTHETA (Inv OR = 1.315) fell just short of the threshold for small, and ESTRESS had no effect.

ADHD

Covariate	OR	SE	t	p		95% CI	Inv OR	ES
CSTRESS	1.556	0.040	17.350	<0.001	1.480	1.636		S
ESTRESS	1.100	0.018	5.860	<0.001	1.065	1.135		ne
LOCTHETA	0.760	0.044	-4.740	<0.001	0.679	0.852	1.315	ne/S
SETHETA	0.726	0.042	-5.550	<0.001	0.648	0.813	1.377	S
ZLEIS	0.953	0.042	-1.100	0.273	0.873	1.039	1.050	ne
_cons	0.016	0.001	-44.650	<0.001	0.013	0.019		

[43] OR 1.32 = small, 2.38 = medium, 4,79 = large, for non-rare outcome events (Olivier et al., 2017)
[44] Inverse ORs were reported for ORs < 1, because they are easier to interpret and compare to effect size guidelines

N	11,402
LR $chi^2(4)$	577.700
Prob > chi^2	< 0.001
Pseudo R^2	0.124

Table 55. Logistic regression of ADHD indicator on stressors and protective factors (Stata)

2.2 ADHD severity

ADHD severity is a continuous outcome variable, so OLS regression was used. Here again, all the predictors except ZLEIS were significant at a 99.9% confidence interval and were in the direction expected to support (i.e. not to refute) state regulation theory. The largest coefficient was on CSTRESS (b = 0.208***), indicating that a 1 unit increase in CSTRESS corresponded to a 0.208 unit increase in ADHD severity[45], next largest was LOCTHETA (b = -0.164***). Variance explained was 19.5%, indicating a small, but heading towards medium, effect size[46] for the model as a whole.

ADHDSEV

Covariate	Coef	SE	t	p	95% CI	
CSTRESS	0.208	0.006	34.300	<0.001	0.196	0.219
ESTRESS	0.039	0.003	11.910	<0.001	0.033	0.045
LOCTHETA	-0.164	0.010	-16.540	<0.001	-0.183	-0.144
SETHETA	-0.120	0.010	-11.830	<0.001	-0.140	-0.100
ZLEIS	-0.007	0.008	-0.930	0.351	-0.023	0.008
_cons	-0.444	0.015	-30.220	<0.001	-0.472	-0.415

N	11,402
F (5, 11,396)	554.060
p-value	< 0.001
R^2	0.195

Table 56. Linear regression of ADHD indicator on stressors and protective factors (Stata)

[45] The coefficients are unstandardized.
[46] Recommended R^2 effect size interpretation for social sciences: > 0.04 = small, > 0.25 = medium, > 0.64 = large (Ferguson, 2009)

After the regression the VIF command was used to test the collinearity between the predictor variables. All values were close to one, indicating collinearity was not a problem in this model (Field, 2009).

Variable	VIF	1/VIF
LOCTHETA	1.200	0.830
SETHETA	1.180	0.848
CSTRESS	1.100	0.912
ESTRESS	1.040	0.966
ZLEIS	1.030	0.968
Mean VIF	1.110	

Table 57. Variance inflation factors for regression of ADHD on stressors and protective factors

Overall, the data and models here indicate that the chronic stressors, locus of control, and self-esteem measured at age 10 have a small but significant and potentially important relationship with ADHD and/or ADHD severity measured at the same age. The relationship provides evidence supporting a state regulation theory of ADHD. The measures were taken at the same time and the data is observational, so it is not strong evidence for a causal theory. However, it does suggest that educational approaches for ADHD aimed at reducing chronic stressors, and increasing locus of control and self-esteem, may be a fruitful topic to explore further.

3 ADHD, stress, and subjective wellbeing (SWB)

Before endeavouring to answer RQ3, some nuances in the relationship between the ADHD subgroup indicator and ADHD severity measure are discussed because they are important background to the definition of regression models.

3.1 The ADHD indicator and severity as predictors of wellbeing

In chapter 5, it was noted that the size, direction and significance of the coefficient estimated for the ADHD subgroup indicator variable changed when ADHDSEV was added to a model. This behaviour stems in part from collinearity between the two measures, the relative rarity of ADHD subgroup membership in the data (4.3% in the unmatched sample), and also from a nonlinear relationship between ADHD and ADHDSEV, which was explored here further.

A two-way ANOVA showed that an interaction term for ADHD * ADHDSEV (severity) was not significant when predicting subjective wellbeing ($F = 0.05$, $p = 0.818$). However, univariate regressions of ADHDSEV predicting SWB for the ADHD and non-ADHD groups separately showed that ADHDSEV was a significant predictor in the non-ADHD group ($F(1, 8{,}148) = 124.10$, $p < 0.001$), but not in the ADHD group ($F(1, 367) = 1.36$, $p = 0.243$). The slopes and intercepts of SWB regressed on ADHDSEV were different in the two groups (ADHD: $b = -0.175$, $c = 0.041$; non-ADHD: $b = -0.143$, $c = 0.017$). In the ADHD group, the standard error of the estimate was large, and the confidence interval crossed zero, so it was inconclusive whether the effect of ADHDSEV was positive or negative within the ADHD group.

The non-linear relationship can be explained by the difference in methods used to derive the two measures (see chapter 4 for details). For a brief recap, the application of the DSM-5 criteria to create the ADHD subgroup weighted all symptoms the same, except for two necessary conditions: at least three symptoms from both parent and teacher raters, plus a parent rating of problematic behaviour. The ADHDSEV score weighted each item differently based on the relative rarity within the data, and the two conditions used in diagnosis were not given special status.

For example, since items were weighted for the ADHDSEV score and not for the ADHD indicator, there were children in the ADHD subgroup who met the criteria based on items/symptoms with low weights, and thus had a relatively low ADHDSEV score. Conversely, there were children with high ADHDSEV scores who were not in the ADHD subgroup, because the parent-rated Rutter score did not indicate moderate to severe behaviour problems, which was a necessary condition for membership in the ADHD subgroup. Thus, the two measures, whilst closely related and based on the same 16 symptoms, do not reconcile perfectly to each other. This is a common criticism of DSM-based indicators and there is work ongoing (mentioned in chapter 3) to change the measures so clinical categorical indicators and underlying continuous construct measures relate to each other statistically (Kotov et al., 2017).

A plot of a simple symptom count against the ADHD severity score grouped by non-ADHD and ADHD (Figure 30) shows there is a significant overlap in both symptom counts and severity scores across the two groups. Note that there is a large proportion of children in this sample who have a degree of ADHD symptomatology, but do not meet the diagnostic criteria as defined in chapter 4. For example, 73% of children at age 10 had at least one symptom, and about 28% of them had five or more. This supports the definition of ADHD as a partially continuous construct.

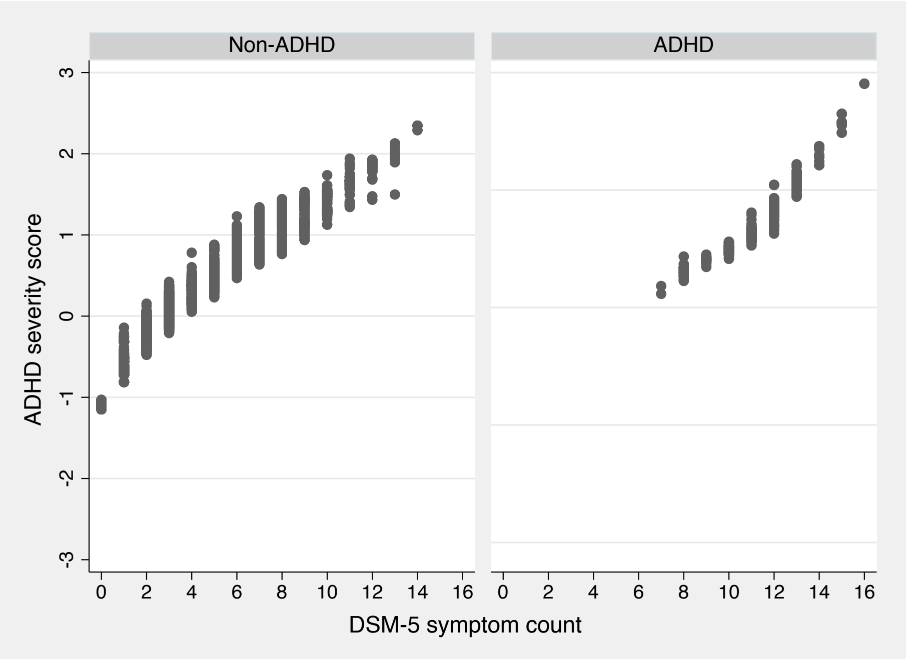


Figure 30. ADHD severity score vs. count of DSM-5 symptoms by ADHD subgroup
(used to derive ADHD subgroup indicator),

Also, scatterplots (Figure 31) of the relationship between ADHD severity and SWB in the non-ADHD group and ADHD subgroup show generally higher ADHD scores in the ADHD group, the overall patterns shape looks similar, but a regression fit line indicates the relationship is between ADHD severity and SWB is stronger in the non-ADHD group.

Figure 31. Scatterplot of ADHD severity vs. subjective wellbeing by group
(non-ADHD n = 8,150 vs. ADHD n = 369)

I intend to explore this further in future research and may alter the methods and measures slightly as a result[47]. However, for the purposes of this thesis, I used the two ADHD measures as they were reported in chapter 4 and noted the nuanced relationship.

In the present study, I created separate univariate models for ADHD and ADHDSEV as predictors of SWB and included the two factors together in some of the subsequent multivariate models, noting their effect on each other as a limitation.

Next, ADHD and other factors expected to predict or covary with adult subjective wellbeing were analysed using three samples:

6) A full, unmatched sample (N=8,519);
7) a matched sample pruned and weighted using coarsened exact matching to improve the balance between ADHD and 'controls' on a key set of confounds (N~6,207)[48]. In effect this was a relatively socio-economically disadvantaged sample; and
8) an ADHD subgroup sample (n=369).

3.2 Analysis of wellbeing outcome in the unmatched sample

The unmatched sample was the full sample extended in chapter 6 (N=8,519, n_{girls} = 4,387, n_{boys}=4,132) which included SWB outcome scores informed by data from age 34, where age 42 data was not available. A series of regression models was fitted on the unmatched sample, with girls and boys reported separately. Finally, a preferred multivariate model was selected, for comparison to the matched sample. Multiple models were fitted to identify an optimal compromise between parsimony and completeness. Adjustments were not made for multiple comparisons.

3.2.1 List of models – unmatched sample

Model 1 - univariate SWB on ADHD

Model 2 - univariate SWB on ADHDSEV

Model 3 – multivariate SWB on ADHD controlling for matching covariates

Model 4 –SWB on ADHD and all identified covariates

Model 5 –SWB on ADHD and core[49] plus significant model 4 covariates

Model 6 –SWB on ADHD plus covariates from model 5, plus wellbeing-related factors measured in adulthood

---

[47] The most likely alteration will be to remove the mother-rated behaviour problems indicator (per Rutter score) as a necessary condition.

[48] The ~ symbol is used because the matched sample includes observations weighted at both less than and greater than one, but not those weighted at zero.

N.B. Mplus includes even zero-weighted observations in counts, whilst Stata does for some procedures and does not for others. The output reported varies depending on whether Mplus or Stata was used.

[49] Core covariates = cstress, estress, loctheta, setheta, lowses, maths, zext, zint

3.2.2 Models – unmatched sample

Model 1 - univariate SWB on ADHD

Model 1 reported in Table 58 suggested that ADHD alone was a significant predictor of wellbeing (b_{girls} = -0.312, b_{boys} = -0.280). In robustness checks, Bayes and Stata SEM models produced similar results. However, R^2 values indicated almost zero variance was explained, and other fit indices (RMSEA, etc.) were not interpretable because the model was just identified.

Girls	Coef	SE	Z	p	Sig
ADHD	-0.312	0.091	-3.415	0.001	**
R^2	0.003			N	4,387
BIC	23280.268				
WBMGS1U[50]		X^2	0.000	DF	0.000

Boys	Coef	SE	Z	p	Sig
ADHD	-0.280	0.068	-4.107	< 0.001	***
R^2	0.005			N	4,132
BIC	23280.268				

Table 58. Regression of SWB on ADHD indicator for girls and boys: MLR estimator in Mplus

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

There was no missing data in the ADHD or SWB variables, and almost exactly the same results were produced using OLS in Stata: N_{girls} = 4,387, b = -0.315, p < 0.001, R^2 = 0.003; N_{boys} = 4,132, b = -0.280, p < 0.001, R^2 = 0.005. Results were also similar using the BAYES estimator in Mplus (b_{girls} = -0.317, b_{boys} = -0.281), see chapter 7 appendix.

Model 2 - univariate SWB on ADHDSEV

ADHD and ADHDSEV (ADHD severity) were not modelled here as joint predictors of SWB, because of the nuanced relationship between the two variables discussed at the beginning of chapter 7. The models reported in Table 59 showed that ADHDSEV alone was also a significant predictor of wellbeing. The coefficients were again negative, indicating that an

[50] WBMGS1U = file name for the Mplus input and output.

increase in ADHDSEV corresponded to a decrease in SWB. The coefficients were smaller than those for ADHD. R^2 values were still small but indicated more variance was explained for girls ($R^2 = 0.026$) than boys ($R^2 = 0.014$), and both values were larger than in the ADHD models. BIC was slightly smaller for ADHDSEV compared to ADHD, indicating a marginally better fit. Other fit indices (RMSEA, etc.) were not interpretable because the model was just identified.

Girls	Coef	SE	Z	p	Sig
ADHDSEV	-0.189	0.017	-10.860	< 0.001	***
R^2	0.026			N	4,387
BIC	23137.743				

Boys	Coef	SE	Z	p	Sig
ADHDSEV	-0.121	0.016	-7.480	< 0.001	***
R^2	0.014			N	4,132
BIC	23137.743				

Table 59. Regression of SWB on ADHDSEV score for girls and boys: MLR estimator in Mplus

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

Model 3 – multivariate SWB on ADHD controlling for matching covariates

Next, SWB was regressed on ADHD in a multivariate model controlling for the subset of covariates that were used in the Coarsened Exact Matching (CEM) procedure to create the matched sample. In model 3, the ADHD coefficient was still negative for both girls and boys but was only significant for boys. R^2 was larger for girls (3.2%) than boys (1.4%). For boys no more variance was explained here controlling for matching variables than in the previous model with ADHDSEV as a univariate predictor (both $R^2 = 0.014$). RMSEA = 0.047 and SRMR = 0.014 indicated acceptable fit, but CFI and TLI values were not useful because the RMSEA for the null model here = 0.03[51], which is better than the proposed model, and less than a rule of thumb cut-off of 0.158, indicating some of the variables have zero or very low correlations with each other (Kenny, 2015). The coefficient of ADHD (in this case without ADHDSEV in the model) is compared to the treatment effect derived using a matched sample, in the subsequent matched sample section.

51 $RMSEA = \frac{\sqrt{X^2 - df}}{\sqrt{df(N-1)}}$; χ^2 for baseline model = 160.092, $df = 16$, $N = 8{,}519$

Girls	Coef	SE	Z	p	Sig
ADHD	-0.143	0.088	-1.627	0.104	
BACKWARD	-0.045	0.110	-0.414	0.679	
DADED	0.045	0.009	4.854	< 0.001	***
HOMELOW	-0.442	0.120	-3.674	< 0.001	***
MUMMAL	-0.167	0.040	-4.180	< 0.001	***
SMOKE	0.168	0.079	2.123	0.034	*
SMOKELVL	-0.108	0.036	-3.001	0.003	**
WHEEZ	-0.067	0.041	-1.605	0.108	
R^2	0.032			N	4,387
RMSEA	0.047			CFI	0.487
SRMR	0.014			TLI	-0.025
BIC	71151.774				
WBMGS3U_adhdonly		X^2	81.866	DF	8.000
		p	< 0.001		

Boys	Coef	SE	Z	p	Sig
ADHD	-0.217	0.070	-3.109	0.002	**
BACKWARD	-0.119	0.099	-1.199	0.231	
DADED	0.030	0.009	3.237	0.001	**
HOMELOW	-0.025	0.103	-0.246	0.805	
MUMMAL	-0.087	0.039	-2.228	0.026	*
SMOKE	0.018	0.081	0.225	0.822	
SMOKELVL	-0.042	0.036	-1.187	0.235	
WHEEZ	-0.002	0.037	-0.048	0.962	
R^2	0.014			N	4,132
RMSEA	0.047			CFI	0.487
SRMR	0.014			TLI	-0.025
BIC	71151.774				

Table 60. Regression of SWB on ADHD and matching covariates for girls and boys: MLR estimator in Mplus

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

Model 4 –SWB on ADHD and all identified covariates

For the next models, all covariates suggested by the literature review, vselect procedure, and hypotheses about stress were included. R^2 was larger at 6.4% for girls and 4.6% for boys, but

BIC was also much higher, indicating relatively worse fit than the models with more parameters, although BIC inherently has a large penalty for complexity (Muthén & Schultzberg, 2017; Schwarz, 1978). SRMR = 0.019 indicated an acceptable fit, RMSEA = 0.099 did not, and the other indices were not interpretable. For girls, five 'non-core' variables were significant: EPVT, HOMELOW, SEPFMUM, SMOKE, and SMOKELVL (English picture vocabulary test, low standard home, separated from mother, mother smoked, and mother's smoking level during pregnancy). The coefficient on SMOKE switched signs and became positive, probably caused by the correlation between the two predictors SMOKE and SMOKELVL. For boys, the only variables that achieved significance were from the 'core' set. SETHETA and ZLEIS were the only predictors that were consistently significant and positive for both boys and girls across these MLR models as well as the BAYES models used as robustness checks (see chapter 7 appendix). The result indicates that self-esteem and engagement in leisure at age 10 may lead to better subjective wellbeing in adulthood, based on the unmatched sample, but model fit was not good enough to rely upon. Also, the unmatched sample was not well-balanced on confounds expected to relate to both ADHD and SWB.

Girls	Coef	SE	Z	p	Sig
ADHD	0.182	0.097	1.868	0.062	
ADHDSEV	-0.068	0.023	-2.995	0.003	**
BACKWARD	0.151	0.109	1.387	0.165	
CSTRESS	-0.028	0.014	-2.034	0.042	*
DADED	0.017	0.010	1.689	0.091	~
ECZ	0.023	0.049	0.475	0.635	
EPVT	0.038	0.019	1.996	0.046	*
ESTRESS	0.003	0.007	0.518	0.604	
HOMELOW	-0.250	0.116	-2.155	0.031	*
LOCTHETA	0.034	0.020	1.711	0.087	~
LOWSES	-0.085	0.048	-1.769	0.077	~
MATHS	0.032	0.020	1.621	0.105	~
MUMMAL	-0.073	0.040	-1.807	0.071	~
POORNBHD	-0.086	0.070	-1.230	0.219	
SEPFMUM	-0.207	0.081	-2.561	0.010	*
SETHETA	0.052	0.019	2.770	0.006	**
SMOKE	0.155	0.077	2.023	0.043	*
SMOKELVL	-0.087	0.035	-2.473	0.013	*
UNMAR	0.002	0.065	0.024	0.981	
WHEEZ	-0.048	0.041	-1.158	0.247	

ZEXT	-0.063	0.021	-2.937	0.003	**
ZINT	-0.020	0.015	-1.314	0.189	~
ZLEIS	0.038	0.015	2.522	0.012	*
R^2	0.064			N	8,519
RMSEA	0.099			CFI	0.000
SRMR	0.019			TLI	-3.862
BIC	303408.574				
WBMGS4U		X^2	983.452	DF	23.000
		P	< 0.001		

Boys	Coef	SE	Z	p	Sig
ADHD	-0.003	0.075	-0.040	0.968	
ADHDSEV	-0.005	0.021	-0.236	0.813	
BACKWARD	0.027	0.102	0.261	0.794	
CSTRESS	-0.045	0.013	-3.479	0.001	**
DADED	0.008	0.010	0.774	0.439	
ECZ	0.032	0.048	0.671	0.502	
EPVT	0.019	0.018	1.061	0.289	
ESTRESS	0.004	0.007	0.570	0.569	
HOMELOW	0.127	0.104	1.218	0.223	
LOCTHETA	0.043	0.020	2.166	0.030	*
LOWSES	-0.062	0.051	-1.232	0.218	
MATHS	0.039	0.019	2.033	0.042	*
MUMMAL	-0.028	0.039	-0.724	0.469	
POORNBHD	-0.110	0.075	-1.475	0.140	~
SEPFMUM	0.012	0.080	0.149	0.882	
SETHETA	0.064	0.020	3.220	0.001	**
SMOKE	0.019	0.080	0.239	0.811	
SMOKELVL	-0.031	0.035	-0.883	0.377	
UNMAR	-0.045	0.077	-0.591	0.554	
WHEEZ	0.009	0.037	0.249	0.803	
ZEXT	-0.038	0.017	-2.182	0.029	*
ZINT	-0.058	0.016	-3.512	< 0.001	***
ZLEIS	0.032	0.016	2.009	0.045	*
R^2	0.046			N	8,519
BIC	303408.574				

Table 61. Regression of SWB on all covariates indicated by literature review, vselect, and theory, for girls and boys, using the MLR estimator in Mplus

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

Model 5 –SWB on ADHD and core plus significant model 4 covariates

In an attempt to improve model fit, only core predictors and significant covariates from the previous step were included in model 5. Two variations were fitted (rather than using a grouping function) because there were different significant sets of predictors for girls and boys from model 4. These models (Table 62) produced a minimal reduction in R^2, and dramatic reduction in BIC, indicating better fit. Both models were just-identified, so other indices were not interpretable. MATHS, LOCTHETA, SETHETA, and ZLEIS (maths score, locus of control, self-esteem, and engagement in leisure) were significant and positive for both girls and boys. EPVT (English picture vocabulary test) was additionally protective for girls. ADHDSEV (severity) was negative for both but only significant for girls. Results using a BAYES estimator were similar.

Girls	Coef	SE	Z	p	Sig
ADHD	0.194	0.097	2.007	0.045	*
ADHDSEV	-0.073	0.023	-3.234	0.001	**
CSTRESS	-0.025	0.013	-1.825	0.068	~
EPVT	0.041	0.019	2.172	0.030	*
ESTRESS	0.003	0.007	0.514	0.607	
HOMELOW	-0.246	0.115	-2.140	0.032	*
LOCTHETA	0.039	0.019	1.985	0.047	*
LOWSES	-0.100	0.048	-2.095	0.036	*
MATHS	0.035	0.019	1.826	0.068	~
SEPFMUM	-0.215	0.081	-2.643	0.008	**
SETHETA	0.053	0.019	2.829	0.005	**
SMOKE	0.148	0.077	1.919	0.055	~
SMOKELVL	-0.090	0.035	-2.583	0.010	*
ZEXT	-0.068	0.021	-3.182	0.001	**
ZINT	-0.024	0.015	-1.591	0.112	~
ZLEIS	0.041	0.015	2.703	0.007	**
R^2	0.060			N	4,387
BIC	134445.198				
WBM0S5U		X^2	0.000	DF	0.000

Boys	Coef	SE	Z	p	Sig
ADHD	-0.004	0.074	-0.053	0.957	
ADHDSEV	-0.007	0.021	-0.325	0.745	
CSTRESS	-0.043	0.013	-3.380	0.001	**
ESTRESS	0.002	0.007	0.293	0.770	
LOCTHETA	0.046	0.020	2.363	0.018	*
LOWSES	-0.073	0.050	-1.467	0.143	
MATHS	0.051	0.018	2.828	0.005	**
SETHETA	0.066	0.020	3.331	0.001	**
ZEXT	-0.041	0.017	-2.404	0.016	*
ZINT	-0.057	0.016	-3.491	< 0.001	***
ZLEIS	0.036	0.016	2.290	0.022	*
R^2	0.043			N	4,132
BIC	115898.534				

Table 62. Regression of SWB on significant covariates per step 4, plus core covariates, for girls and boys: MLR estimator in Mplus

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

The R^2 values were still fairly small at 0.060 and 0.043 for girls and boys respectively. Given that the variables are all measured in childhood and outcomes in adulthood, a very high R^2 is not necessarily expected. So, in the next model, variables measured in adulthood and expected to correlate to SWB were added to see if a model could be fitted explaining a larger portion of the variance in SWB.

Model 6 –SWB on ADHD plus covariates from model 5, plus wellbeing-related factors measured in adulthood

This final model added covariates from adulthood that were expected to correlate to wellbeing (based on ONS-defined dimensions of wellbeing; see literature review in chapter 3), and thus I expected would explain more variance. The R^2 values were indeed much higher (0.208/0.232; girls/boys), and BIC was larger, but not nearly as large as the model including all childhood covariates. The models were just-identified so other indices were not interpretable. Three predictors from childhood were still significant for girls: ZEXT, ZINT, and ZLEIS (externalising problems, internalising problems, and engagement in leisure). Two from childhood survived significance for boys: SETHETA and ZINT (self-esteem and internalising problems). The largest coefficients for both boys and girls were for ODIS, OHLTH, and OLWPART (disabled indicator, general health rating, and living with partner indicator).

Girls	Coef	SE	Z	p	Sig
ACED5	0.026	0.010	2.732	0.006	**
ADHD	0.142	0.088	1.609	0.108	
ADHDSEV	-0.033	0.021	-1.583	0.113	
CSTRESS	-0.004	0.012	-0.304	0.761	
EPVT	0.018	0.017	1.068	0.286	
ESTRESS	0.011	0.006	1.936	0.053	
HOMELOW	-0.148	0.096	-1.536	0.125	
LOCTHETA	0.014	0.018	0.798	0.425	
LOWSES	-0.054	0.043	-1.272	0.203	
MATHS	0.001	0.019	0.031	0.975	
OALCGRP	-0.157	0.033	-4.766	< 0.001	***
ODIS	-0.229	0.048	-4.735	< 0.001	***
OHLTH	-0.265	0.015	-17.093	< 0.001	***
OLWPART	0.280	0.030	9.186	< 0.001	***
OSOC	-0.035	0.020	-1.769	0.077	
SEPFMUM	-0.125	0.075	-1.653	0.098	
SETHETA	0.021	0.017	1.202	0.229	
SMOKE	0.095	0.070	1.356	0.175	
SMOKELVL	-0.051	0.032	-1.630	0.103	
ZEXT	-0.047	0.019	-2.420	0.016	*
ZINT	-0.036	0.014	-2.572	0.010	*
ZLEIS	0.031	0.014	2.304	0.021	*
R^2	0.208			N	4187
BIC	182705.554				
WBM0S6U		X^2	0.000	DF	0.000

Boys	Coef	SE	Z	p	Sig
ACED5	0.003	0.009	0.375	0.708	
ADHD	-0.052	0.066	-0.783	0.433	
ADHDSEV	0.026	0.019	1.386	0.166	
CSTRESS	-0.016	0.011	-1.395	0.163	
ESTRESS	0.003	0.006	0.419	0.675	
LOCTHETA	0.019	0.018	1.078	0.281	
LOWSES	0.014	0.044	0.325	0.745	
MATHS	0.009	0.017	0.509	0.611	
OALCGRP	-0.059	0.020	-2.972	0.003	**

ODIS	-0.240	0.052	-4.611	< 0.001	***
OHLTH	-0.314	0.016	-20.154	< 0.001	***
OLWPART	0.327	0.031	10.465	< 0.001	***
OSOC	-0.045	0.019	-2.323	0.020	*
SETHETA	0.043	0.018	2.449	0.014	*
ZEXT	-0.025	0.016	-1.573	0.116	
ZINT	-0.052	0.014	-3.620	< 0.001	***
ZLEIS	0.021	0.015	1.464	0.143	
R^2	0.232			N	4,132
BIC	164450.222				

Table 63. Regression of SWB on all covariates from step 5, plus covariates from adulthood expected to correlate with SWB. For girls and boys: MLR estimator in Mplus

3.2.3 Selection of a preferred multivariate model in the unmatched sample

In order to compare models across samples, I selected a preferred model based on model fit and relevance. The last model fit best, but highlighted variables from adulthood, and my objective here is to focus on factors from childhood that affect adult outcomes. The next-best fit was model 5, with core predictors plus significant covariates from model 4. Therefore, it was selected as the preferred model for cross-sample comparisons.

3.3 Analysis of wellbeing in the matched sample

Next, the relationship between childhood ADHD, stress, and subjective wellbeing (SWB) in adulthood was evaluated using the weighted matched sample, which was created using the coarsened exact matching (CEM) procedure. The seven matching variables were: SMOKE, SMOKELVL, DADED, HOMELOW, BACKWARD, MUMMAL, and WHEEZ (mother smoked during pregnancy, mother's smoking level, father's education level, low standard home, backward development, mother malaise, and problems with wheezing, all measured at age 0 or 5). Full details on the matching process were reported in chapter 6. The models were fitted in Mplus using MLR estimators. Models with the BAYES estimator and SEM Stata SEM/MLMV were run selectively as robustness checks; some of the results are reported in the main text, and the rest in the chapter 7 appendices.

Similar to the procedure for the unmatched sample, a series of regression models was fitted in the weighted matched sample. An exception is that there is no model of SWB on ADHD and the covariates used for matching, which was step 3 in the unmatched sample section. This is because the weights already control for the matching variables.

3.3.1 List of models – matched sample

Model 1 – univariate SWB on ADHD

Model 2 – univariate SWB on ADHDSEV

Model 3 – multivariate SWB on ADHD and all relevant covariates

Model 4 –SWB on ADHD and core plus significant covariates from model 3

Model 5 –SWB on ADHD and model 4 covariates plus covariates from adulthood

3.3.2 Models – matched sample

Model 1 – univariate SWB on ADHD

In model 1, ADHD regressed on SWB in the weighted matched sample had a negative coefficient for girls and boys, but only significant for boys. This is a 'treatment effect', of ADHD, determined using CEM as method for estimating treatment effects. The pattern of the negative relationship and significance only for boys is similar to the finding in model 3 of the previous section (regression modelled on the full unmatched sample with matching variables as controls). The coefficients, or treatment effects in this sample are smaller (by 0.04) than those from OLS regression controlling for the same variables, i.e. the effect of ADHD on SWB was smaller in the matched sample. Also, the coefficients for boys based on both OLS and CEM were significant at $p < 0.05$, but would lose significance in a multiple tests scenario, or with more stringent criteria that is often used with large samples like this one (e.g. $p < 0.01$). R^2 values were nearly nil, indicating very little variance was explained. Models were just identified, so other indices were not interpretable.

	OLS	CEM	Difference
Girls	-0.143	-0.100	-0.043
Boys	-0.217*	-0.173*	-0.044

Table 64. Difference between treatment effect of ADHD using OLS

(with matching variables as covariates: step 3 on the unmatched sample) vs. CEM (step 1 on the matched sample)

Girls	Coef	SE	Z	p	Sig
ADHD	-0.100	0.095	-1.057	0.291	
R^2	0.000			N	4387
BIC	24383.704				
WBMGS1W		X^2	0.000	DF	0.000

Boys	Coef	SE	Z	p	Sig
ADHD	-0.173	0.077	-2.246	0.025	*

R^2	0.002	N	4132
BIC	24383.704		

Table 65. Regression of SWB on ADHD subgroup indicator, for girls and boys: MLR estimator in Mplus

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

Model 2 – univariate SWB on ADHDSEV

In a univariate model, ADHD severity had a significant and negative effect on SWB for both girls and boys, but little variance was explained (2.6% and 1.5% for girls and boys respectively).

Girls	Est	SE	t	p	sig
ADHDSEV	-0.182	0.020	-8.940	< 0.001	***
R^2	0.026			N	2923

Boys	Est	SE	t	p	sig
ADHDSEV	-0.125	0.018	-6.910	< 0.001	***
R^2	0.015			N	3284

Table 66. Regression of SWB on the ADHDSEV in the matched sample, for girls and boys: OLS estimator in Stata (no missing data)

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

Model 3 – multivariate SWB on ADHD and all relevant covariates

Model 3 included all covariates identified by the literature review, vselect procedure, and indicated by the hypotheses for the present study, i.e. stressors and protective factors against stress. RMSEA and SRMR were within acceptable limits, but CFI and TLI were not interpretable because RMSEA for the null model was less than 0.158 (Kenny, 2015). The R^2 at 10.3% for girls and 4.6% for boys indicated more variance was explained here by the multivariate set than by the previous models, which is to be expected given the larger number of predictors. Only two predictors were significant for girls, BACKWARD (b = 0.630) and ZEXT (b = -0.134). A few other variables had relatively large coefficients: ADHD (b = 0.238), WHEEZ (b = 0.181), and POORNBHD (b= - 0.137).

No predictors were significant for boys in model 3, but HOMELOW (b = 0.189) and POORNBHD (b = - 0.250) had the largest coefficients.

Girls	Coef	SE	Z	p	Sig
ADHD	0.238	0.131	1.822	0.069	
ADHDSEV	-0.014	0.049	-0.293	0.770	
BACKWARD	0.630	0.257	2.449	0.014	*
CSTRESS	-0.006	0.027	-0.203	0.839	
DADED	0.065	0.041	1.595	0.111	
ECZ	-0.104	0.139	-0.745	0.456	
EPVT	0.047	0.049	0.956	0.339	
ESTRESS	0.017	0.016	1.101	0.271	
HOMELOW	-0.054	0.207	-0.263	0.793	
LOCTHETA	0.010	0.045	0.234	0.815	
LOWSES	-0.089	0.112	-0.798	0.425	
MATHS	0.079	0.042	1.881	0.060	
MUMMAL	-0.076	0.084	-0.911	0.362	
POORNBHD	-0.137	0.170	-0.807	0.420	
SEPFMUM	-0.110	0.145	-0.761	0.446	
SETHETA	0.096	0.054	1.777	0.076	
SMOKE	-0.058	0.244	-0.238	0.812	
SMOKELVL	-0.005	0.101	-0.050	0.960	
UNMAR	0.018	0.125	0.146	0.884	
WHEEZ	0.181	0.111	1.633	0.103	
ZEXT	-0.134	0.055	-2.431	0.015	*
ZINT	0.002	0.029	0.070	0.944	
ZLEIS	0.071	0.039	1.813	0.070	
R^2	0.103			N	4,387
RMSEA	0.047			CFI	0.000
SRMR	0.025			TLI	-6.008
BIC	318788.275				
WBMGS4W		X^2	237.798	DF	23.000
	Null	X^2	107.304	DF	46.000
	RMSEA Null	0.013			

Boys	Coef	SE	Z	p	Sig
ADHD	0.066	0.101	0.652	0.514	
ADHDSEV	-0.038	0.050	-0.761	0.447	
BACKWARD	-0.032	0.263	-0.123	0.902	
CSTRESS	-0.042	0.023	-1.780	0.075	~

DADED	-0.003	0.024	-0.134	0.894	
ECZ	0.034	0.117	0.290	0.772	
EPVT	0.002	0.035	0.048	0.962	
ESTRESS	-0.008	0.015	-0.569	0.570	
HOMELOW	0.189	0.277	0.683	0.495	
LOCTHETA	0.064	0.046	1.390	0.165	~
LOWSES	0.085	0.138	0.619	0.536	
MATHS	0.053	0.041	1.293	0.196	~
MUMMAL	0.002	0.083	0.027	0.979	
POORNBHD	-0.250	0.137	-1.821	0.069	~
SEPFMUM	0.180	0.131	1.377	0.169	~
SETHETA	0.047	0.039	1.204	0.229	
SMOKE	-0.122	0.334	-0.365	0.715	
SMOKELVL	-0.010	0.130	-0.076	0.939	
UNMAR	-0.047	0.112	-0.418	0.676	
WHEEZ	-0.015	0.069	-0.223	0.823	
ZEXT	-0.028	0.036	-0.767	0.443	
ZINT	-0.022	0.030	-0.741	0.459	
ZLEIS	0.005	0.039	0.122	0.903	
R^2	0.046			N	4,132
RMSEA	0.047			CFI	0.000
SRMR	0.025			TLI	-6.008
BIC	318788.275				

Table 67. Regression of SWB on all covariates indicated by literature and present study hypotheses for girls and boys: MLR estimator in Mplus

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

Model 4 –SWB on ADHD and core plus significant covariates from model 3

Model 4 includes only core predictor covariates plus those that were significant at p > 0.05 in model 3. Core predictors included the two ADHD variables, five stress-related variables, and four found to be predictors of outcomes in ADHD in a review. They were: ADHD, ADHDSEV, ESTRESS, CSTRESS, LOCTHETA, SETHETA, ZLEIS, LOWSES, MATHS, ZEXT, and ZINT (ADHD indicator, ADHD severity, life events stressors, chronic stressors, locus of control, self-esteem, engagement in leisure, low SES (free school meals), maths, externalising problems and internalising problems).

R^2 was smaller than for model 3 at 8.3% for girls and 3.7% for boys. The models were just identified, so most fit indices were not useful, but BIC was much smaller than for model 3, indicating a relatively better fit.

For girls in the matched sample, MATHS (b = 0.101), ZEXT (b = -0.142), and ZLEIS (b = 0.091) survived as significant predictors of SWB. The positive effect of my IQ proxy (MATHS) and negative effect of behaviour problems (ZEXT, or externalising problems) in childhood was consistent with findings elsewhere about adult wellbeing (Clark et al., 2017; Costello & Maughan, 2015; Mensah & Hobcraft, 2008). The positive effect of ZLEIS (engagement in leisure) was a new and interesting finding that should be explored further in future research, as it suggests engagement in leisure may be a protective factor for adult wellbeing in a socio-economically disadvantaged sample.

The large and positive coefficient on ADHD has been discussed previously as a spurious interaction with ADHDSEV. The two coefficients taken together jointly were not significant for boys or girls. The large and positive coefficient on BACKWARD (child's development rated as 'backward' by a health visitor at age 5) was also unexpected and investigated further. Very few children in the preferred sample were rated as backward: (221, or 2.6% of N=8,519). However, 36 (9.8%) of our ADHD subgroup (n=369) were rated as such. A univariate regression of SWB on BACKWARD (just using regress in Stata and ignoring unmet assumptions) was significant and had a negative coefficient, though explained almost zero variance for girls with available data (F (1, 3,584) = 5.88, p = 0.015, Adj R^2 = 0.001). ANOVA analysis showed there was a significant interaction between ADHD and BACKWARD for girls (F = 4.85, p = 0.028) when predicting SWB. Therefore, the positive and significant coefficient for BACKWARD in the girls' model was dismissed as probably spurious, or at least not reliable.

For boys, only MATHS (b = 0.080, p = 0.035) was significant, and CSTRESS was fairly close to significance (b = -0.041, p = 0.105). The positive effect of IQ (measured by proxy) in childhood on adult wellbeing again provides support for previous findings in literature. The closeness of CSTRESS to significance provides some evidence to suggest further investigation of the relationship between chronic stress in childhood and adult wellbeing may be worthwhile.

Girls	Coef	SE	Z	p	Sig
ADHD	0.303	0.146	2.070	0.038	*
ADHDSEV	-0.031	0.050	-0.613	0.540	
BACKWARD	0.523	0.249	2.101	0.036	*
CSTRESS	0.000	0.028	0.011	0.992	

ESTRESS	0.013	0.015	0.844	0.398	
LOCTHETA	0.044	0.044	1.012	0.311	
LOWSES	-0.133	0.103	-1.284	0.199	
MATHS	0.101	0.045	2.256	0.024	*
SETHETA	0.085	0.056	1.521	0.128	
ZEXT	-0.142	0.063	-2.255	0.024	*
ZINT	0.002	0.029	0.079	0.937	
ZLEIS	0.091	0.039	2.330	0.020	*
R^2	0.083			N	4,387
BIC	126426.981				
WBM0S5W		X^2	0.000	DF	0.000

Boys	Coef	SE	Z	p	Sig
ADHD	0.072	0.107	0.668	0.504	
ADHDSEV	-0.037	0.051	-0.725	0.469	
CSTRESS	-0.041	0.025	-1.622	0.105	~
ESTRESS	-0.009	0.014	-0.659	0.510	
LOCTHETA	0.055	0.046	1.216	0.224	
LOWSES	0.055	0.142	0.389	0.697	
MATHS	0.080	0.038	2.109	0.035	*
SETHETA	0.047	0.042	1.123	0.261	
ZEXT	-0.027	0.038	-0.720	0.472	
ZINT	-0.014	0.031	-0.445	0.656	
ZLEIS	0.009	0.040	0.226	0.821	
R^2	0.037			N	4,132
BIC	118815.705				

Table 68. Regressions of SWB on core covariates plus significant variables from step 4: MLR estimator in Mplus

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

Model 5 – SWB on ADHD and model 4 covariates plus covariates from adulthood

Model 5 included six covariates measured in adulthood that were expected to relate to wellbeing in adulthood. This model explained a much larger proportion of variance in SWB, at 29.6% for girls and 25.6% for boys, indicating a medium-sized effect. BIC was larger than for step 5, indicating relatively worse fit, but to be expected given the increased number of predictors and large associated increase in R^2. Both models were just identified, so other indices were not useful.

The only variable from childhood that survived significance was BACKWARD (again positive), for girls. Based on the analysis discussed with model 4, this effect was dismissed as spurious. ADHD also had a relatively large positive coefficient for girls, but it was not significant, and the standard error was large. I have assumed this was also spurious due to the relationship between ADHD and ADHDSEV.

In adulthood, ACED5 (academic education level), had a significant but small effect for girls only ($b = 0.060$), as did OALCGRP (alcohol problems indicator; $b = -0.167$). The largest effects on SWB for both girls and boys were explained by ODIS (disability indicator; $b_{girls} = -0.395$; $b_{boys} = -0.393$), OHLTH[52] (general health rating; $b_{girls} = -0.290$; $b_{boys} = -0.321$), and OLWPART (living with a partner indicator; $b_{girls} = 0.223$; $b_{boys} = 0.444$), all measured in adulthood. The positive effect of living with a partner was almost twice as large for boys compared to girls.

Girls	Coef	SE	Z	p	Sig
ACED5	0.060	0.020	3.078	0.002	**
ADHD	0.193	0.120	1.604	0.109	
ADHDSEV	-0.017	0.044	-0.376	0.707	
BACKWARD	0.550	0.214	2.571	0.010	*
CSTRESS	0.015	0.023	0.660	0.509	
ESTRESS	0.019	0.012	1.567	0.117	
LOCTHETA	0.029	0.037	0.779	0.436	
LOWSES	-0.026	0.078	-0.336	0.737	
MATHS	0.031	0.038	0.806	0.420	
OALCGRP	-0.167	0.075	-2.238	0.025	*
ODIS	-0.395	0.100	-3.958	< 0.001	***
OHLTH	-0.290	0.031	-9.258	< 0.001	***
OLWPART	0.223	0.060	3.748	< 0.001	***
OSOC	-0.051	0.044	-1.153	0.249	
SETHETA	0.022	0.038	0.573	0.567	
ZEXT	-0.074	0.044	-1.684	0.092	
ZINT	0.009	0.025	0.348	0.728	
ZLEIS	0.055	0.029	1.880	0.060	
R^2	0.296			N	4,387
BIC	174242.114				

[52] OHLTH was coded 1-5, where 1=excellent and 5=poor health, which is why the coefficient is negative.

WBM0S6W		X^2	0.000	DF	0.000

Boys	Coef	SE	Z	p	Sig
ACED5	-0.002	0.019	-0.118	0.906	
ADHD	-0.008	0.091	-0.086	0.932	
ADHDSEV	-0.016	0.042	-0.377	0.706	
CSTRESS	-0.008	0.020	-0.421	0.674	
ESTRESS	-0.003	0.012	-0.249	0.804	
LOCTHETA	0.047	0.038	1.246	0.213	
LOWSES	0.164	0.117	1.400	0.162	
MATHS	0.026	0.037	0.702	0.483	
OALCGRP	-0.074	0.041	-1.782	0.075	
ODIS	-0.393	0.105	-3.742	< 0.001	***
OHLTH	-0.321	0.028	-11.512	< 0.001	***
OLWPART	0.444	0.064	6.971	< 0.001	***
OSOC	0.009	0.037	0.256	0.798	
SETHETA	0.029	0.035	0.834	0.404	
ZEXT	0.001	0.032	0.029	0.977	
ZINT	-0.002	0.026	-0.068	0.946	
ZLEIS	0.010	0.034	0.286	0.775	
R^2	0.256			N	4,132
BIC	166952.830				
WBM1S6W		X^2	0.000	DF	0.000

Table 69. Regression of SWB on covariates from step 5, plus adult variables that correlate with SWB: MLR estimator in Mplus

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

3.3.3 Selection of a preferred multivariate model in the matched sample

The step 4 model had the best fit amongst models relevant to childhood factors in adult outcomes, and thus was used in cross-sample comparisons.

3.4 Analysis of wellbeing in the ADHD subgroup sample

Finally, an overall objective of my PhD was to add to our collective understanding of what leads to optimal outcomes specifically for people with ADHD. The theory proposed to explain ADHD outcomes is State Regulation theory. To recap, the overall hypothesis is that stress in childhood relates to worse outcomes, and factors that protect against stress relate to better outcomes.

The ADHD subsample is much smaller than the others, and thus has reduced power to detect separate effects estimates for boys and girls or to accommodate a large number of covariates. Thus, I sought to create models that would maximise power and reliability of results. First, I combined girls and boys, and added sex as a covariate. I also added SUBTYP (ADHD subtype: combined, primarily hyperactive, primarily inattentive) because this variable is uniquely relevant to the ADHD subgroup. Then I started analysis with core variables from previous literature findings and my hypotheses about stress and ADHD: (from the literature; ADHDSEV, SUBTYP, SEX, MATHS, ZEXT, ZINT, LOWSES, and from my hypotheses: CSTRESS, ESTRESS, LOCTHETA, SETHETA, ZLEIS. The approach identified a total of 12 covariates[53].

Initially I fitted a model with these 12 covariates, and noted that the MATHS variable, which was a significant predictor of SWB in other models using the unmatched and matched samples, was not significant. I found this surprising. I thought perhaps a proxy IQ measure other than MATHS might be important for ADHD. ZREAD (z-score of reading 'age') was also available from the age 10 data. MATHS and ZREAD were highly correlated in this sub-sample (r = 0.733, p < 0.001), but not exactly, so I thought it would be interesting to see how MATHS and ZREAD behaved within the ADHD group. Thus, I added ZREAD to the model, bringing the total to 13 covariates.

Model 1 – multivariate SWB on core covariates plus sex, ADHD subtype and ZREAD in the ADHD subgroup sample, girls and boys combined

Model 1 on the ADHD sample (n=369) explained approximately 9.9% of variance in SWB for the group. The model was just identified, so other indices were not interpretable.

As Table 70 indicates, significant predictors of SWB were CSTRESS (b = -0.068*), ZEXT (b = -0.110*), and ZREAD (b = 0.190*), (chronic stressors, externalising problems, and reading ability, all measured at age 10). SUBTYP (b = -0.165; ADHD subtype) was close to significance. SUBTYP was coded so that 1 = combined, 2 = hyperactive, and 3 = inattentive, so the negative coefficient here indicates the inattentive type was associated with the lowest SWB. ADHDSEV (b = -0.301; ADHD severity score) had the largest coefficient but was not near significance because the standard error of the estimate was large.

Girls + Boys	Coef	SE	Z	p	Sig

[53] I also tried this model with a pared-down set of the matching covariates: DADED, SMOKELVL, MUMMAL, HOMELOW, and WHEEZ. None were close to significance in predicting SWB in the ADHD-only sample.

ADHDSEV	-0.301	0.245	-1.228	0.219	
CSTRESS	-0.068	0.030	-2.281	0.023	*
ESTRESS	0.035	0.022	1.569	0.117	~
LOCTHETA	0.058	0.070	0.831	0.406	
LOWSES	-0.060	0.128	-0.470	0.639	
MATHS	-0.101	0.095	-1.058	0.290	
SETHETA	0.082	0.063	1.300	0.193	~
SEX	-0.003	0.110	-0.023	0.981	
SUBTYP	-0.165	0.094	-1.758	0.079	~
ZEXT	-0.110	0.045	-2.428	0.015	*
ZINT	-0.039	0.045	-0.878	0.380	
ZLEIS	0.002	0.054	0.036	0.971	
ZREAD	0.190	0.093	2.053	0.040	*
R^2	0.099			N	369
BIC	12781.831				
WBMBS8A					

Table 70. SWB on ADHD and core covariates with girls and boys combined, ADHD-only sample (*N*=369)

**** p<0.001, ** p<0.01, * p<0.05, ~ p > 0.05 but < 0.20*

This model used MLR and FIML to estimate a linear regression, and thus provided no test statistic or significance value. If it were an OLS regression using an 'F' test, the model would be well-powered: $(1-\beta) = 0.98$ where $R^2 = 0.09$, $\alpha = 0.05$, n = 369, and covariates = 13; Faul, Erdfelder, Lang, & Buchner, 2009). However, the estimator in use here to handle missing data and violated assumptions is MLR in Mplus. MLR (maximum likelihood with robust standard errors) generally requires larger sample sizes, due to its iterative nature. A closer approximation is the power needed to calculate χ^2 with df = 119, $(1-\beta) = 0.80$, $\alpha = 0.05$, to detect an effect size between small and medium (the MLR model in Table 69 above had df = 119), which was estimated by G*Power at n = 701 (Faul et al., 2009).

In the interest of increasing power, I used the G*Power formula for a χ^2 test to evaluate a number of other model variations with fewer covariates. The only one that was sufficiently powered ($(1-\beta) = 0.80$) to detect its effect was a 3-covariate model including CSTRESS, ZEXT, and ZREAD, i.e. the significant covariates from Model 1.

Model 1a – SWB on significant covariates from model 1 on ADHD subgroup

Model 1a (Table 71) explained 7.1% of variance in the ADHD subgroup, large enough to be considered practically significant (Ferguson, 2009). I standardised the regression coefficients here because in this model (unlike any of the others so far), all the predictors and the outcome were modelled as continuous variables. Thus, they are straightforward to interpret: for example, a 1 SD increase in CSTRESS correlated to a 0.108 SD decrease in SWB. All three were significant.

Girls + Boys	ß	SE	Z	p	Sig
CSTRESS	-0.108	0.053	-2.035	0.042	*
ZEXT	-0.115	0.054	-2.137	0.033	*
ZREAD	0.166	0.058	2.871	0.004	**
R^2	0.071			N	369
BIC	4516.866				

WBMBS10A_top3

Table 71. SWB on ADHD and top 3 covariates with girls and boys combined, ADHD-only sample (*N*=369)

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

Since Model 1a was the only model tested that was sufficiently powered, it was interpreted in the discussion as the preferred model.

3.4.1 ADHD severity in the matched sample vs. the ADHD subgroup

To evaluate the effect of ADHDSEV (severity score) as a predictor of SWB, I fitted univariate models of SWB on ADHDSEV in the matched sample, for girls, boys, and girls and boys combined. As summarised in Table 72, the effects were negative and highly significant for all three variations and explained 1.3% of variance in SWB for boys and the combined groups and 1.7% for girls. According to Ferguson, (2009) these are too small to meet a minimum threshold of 4% for effects to be considered practically important in the social sciences.

Within the ADHD subgroup, the effect of ADHDSEV on SWB was not significant and R^2 was near zero, in all three groups. This finding does not support reports elsewhere that within an ADHD sample, severity is a key factor in outcomes (e.g. Costello & Maughan, 2015).

ADHDSEV						
	Matched		R^2	ADHD		R^2
Girls	-0.150	***	0.017	0.048	ns	0.000
Boys	-0.126	***	0.013	-0.295	ns	0.010

Both	-0.129	***	0.013	-0.175	ns	0.003

Table 72. ADHDSEV as a predictor of SWB in the matched sample and ADHD subgroup

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

4 Chapter 7 discussion

Chapter 7 sought to answer the following research questions about the relationship between ADHD and stress in childhood, and between ADHD, stress, and subjective wellbeing (SWB) in adulthood:

> RQ2: How do chronic stressors, life event stressors, locus of control, self-esteem, and engagement in leisure relate to ADHD and ADHD severity, all as measured at age 10? Does the relationship provide evidence to support state regulation theory?
>
> RQ3: What is the effect of childhood ADHD on adult wellbeing using different methods and covariate sets to estimate treatment effects? Do the results support state regulation theory? How do they compare to findings from other research?

4.1 Answering RQ2: stress, protective factors, and ADHD

A logistic model showed that chronic stressors (*OR* = 1.556***), locus of control (Inv *OR* = 1.315***), and self-esteem (Inv *OR* = 1.377***) had small but potentially important effects on membership in the ADHD subgroup, with a pseudo R^2 = 0.124. A linear regression model showed that chronic stressors (*b* = 0.208***), life event stressors (*b* = 0.039***), locus of control (*b* = -0.164***), and self-esteem (*b* = -0.120***) had highly significant effects on ADHD severity score. Although the high significance level may be a by-product of the large sample size (*N* = 11,402), the model explained 19.5% of variance in ADHD severity, indicating a small but important[54] effect for the model.

Thus overall, the BCS70 age 10 data and the models reported here indicate that chronic stressors, locus of control, and self-esteem have a small but potentially important relationship with ADHD, and the relationship provides some evidence for the hypothesis I used to operationalise state regulation theory. The effects were very small for life event stressors, and

[54] Recap: R^2 effect size interpretations in use: > 0.04 = small, > 0.25 = moderate, >0.64 = large (Ferguson, 2009)

not significant for the engagement in leisure measure. The implication is that interventions or educational approaches for ADHD aimed at reducing chronic stressors, and increasing locus of control and self-esteem, may be a fruitful topic to explore further.

4.2 Answering RQ3: ADHD, stress, and subjective wellbeing

Four different variations of methods and models were used to test effects of ADHD on wellbeing. They were: 1) naïve regression, 2) controlled regression (on matching covariates), 3) matched sample regression, and 4) controlled matched sample regression.

4.2.1 Method comparison

A four-way comparison was made to show the difference between effects found for each method variation (Table 73).

The (unstandardized) coefficient sizes and significance levels became progressively smaller as the method moved from least controlled to most controlled. The first method indicates a relatively large and highly significant negative effect of ADHD on SWB, and the last, most controlled method indicates no significant effect. The comparison supports the treatment effects literature arguing that matching is a stronger method than controlled regression (e.g. Rosenbaum & Rubin, 1983).

The most controlled method was used to answer the research question, because it was the best informed via relevant covariates.

	Least controlled		→		Most controlled
	1±	2	3	4-Girls	4-Boys
b	-0.284***	-0.205***	-0.149*	NA±±/ns	NA/ns
R^2	0.003	0.020	0.001	0.083	0.037

Table 73. Treatment effect of ADHD on SWB from least to most controlled methods

**** p<0.001, ** p<0.01, * p<0.05*
Reference category = level 0: A levels or higher (or vocational equivalent)
±1 – unmatched sample naïve regression
2 - unmatched sample, regression controlled for matching covariates
3 – matched sample
4 – matched sample plus controlled regression for girls and boys separately
±± Coefficient is NA (not applicable) because ADHD + ADHDSEV effects were evaluated jointly in these models.

4.2.2 Most controlled method

The matched sample models did not provide satisfactory evidence that ADHD, or the variables defined as stressors and protective factors in childhood, had important effects on adult subjective wellbeing, in the matched sample. Maths, as a proxy for IQ, was important and positive, and externalising behaviours were negative for girls only. The small and positive effect of leisure activity on SWB for girls should be interpreted with caution, but perhaps explored further. ZLEIS was not necessarily a robust measure; it was newly derived for this study and not validated in other samples, so may be better explained by collinearity or unobserved confounds. For example, although LOWSES was factored into the models, a finer measure of relative socio-economic advantage was not, and girls from particularly advantaged backgrounds may have had more opportunities to engage in leisure activities. Such a background could be a better explanation of the small lift in adult SWB.

4.2.3 Comparisons to other research

As described above, I found in the matched sample, using measures derived in chapter 4, that ADHD did not have an effect large enough to be considered important (Ferguson, 2009) on adult SWB. Comparing this to other findings in literature was not straightforward, for a few reasons. First, my sample was non-clinical, whilst most other ADHD samples with long-term outcomes reported were clinical. Also, other studies measured specific mental health outcomes like depression, rather than a broader construct like wellbeing, perhaps because of their clinical nature. However, elsewhere the lack of depression (or lack of malaise) has been used as a measure of wellbeing (e.g. Layard, Clark, Cornaglia, Powdthavee, & Vernoit, 2014), so a depression outcome can serve as a basis for a broad comparison to (low) wellbeing. Related findings from five studies in the literature as reported in chapter 3 are re-summarised here for reference:

Odds ratio effects of ADHD on a depression outcome from literature review:

- Study 1: the previous study of ADHD in BCS70 reported a significant independent effect of ADHD on both depression (malaise) and life satisfaction (related to wellbeing), measured at age 30 and controlling for a large number of biological, social and economic factors. However when I converted their reported statistics to effect sizes, a small effect was indicated for boys (1.32) and an effect that was not practically significant (1.19) for girls (Brassett-Grundy & Butler, 2004).
- Study 2: systematic review of 98 studies on ADHD outcomes found a medium effect per a pooled odds ratio (2.31; Erskine et al., 2016).
- Study 3: 16-year follow-up of the multimodal treatment study of ADHD (MTA, N=717)) found a small effect (OR = 1.43; Hechtman et al., 2016)
- Study 4: 33-year follow-up at age 41 (N=271) found a small effect (OR = 1.55; Klein et al., 2012)
- Study 5: age 38 follow-up with the Dunedin cohort (*N*~956) found a nonsignificant effect (OR = 1.20; Moffitt et al., 2015)

My findings in the present study initially appeared to disagree with the effects reported in study 1. However, when I converted their statistics to effect sizes and used the same size thresholds as the present study, the effects were the same for girls (no effect) but not boys; their effect was small and practically important, whilst mine was not.

My findings (no effect) do agree with study 5 (Moffitt et al., 2015), and do not agree with findings of a small-medium effect as reported (or calculated post-hoc by me) in studies 2, 3, and 4 (Erskine et al., 2016; Hechtman et al., 2016; Klein et al., 2012).

A pair of studies evaluated effects of childhood factors on adult life-satisfaction as a proxy for wellbeing, using BCS70 data (Clark et al., 2017; Layard et al., 2014), though not specifically for ADHD. They found that emotional health in childhood was the strongest predictor of adult life-satisfaction, based on multivariate models including cognitive abilities, behaviour problems, and family background as covariates. Their measure of emotional health did not map directly to any of the constructs used in my chapter 7 study. Items were included from three ages (5, 10, and 16), but a complete list was not provided. Here is the description that was provided:

> *"Questions on children's emotional health are more internal, and relate to worry, unhappiness, sleeplessness, eating disorder, bedwetting, fearfulness, school avoidance, tiredness and psychosomatic pains." (Layard et al., 2014 p. F725)*

Based on this description, their emotional health construct had content that either approximately mapped to my internalising problems or chronic stressors measures or was not included in any of my measures. The strongest childhood effects on adult wellbeing in my matched (i.e. relatively socio-economically disadvantaged) sample controlling for key confounds were from maths (as IQ proxy) and externalising problems (behaviour problems) measured at age 10. My maths and externalising problems measures did not overlap with the Layard et al., (2014) measure of emotional health. So, broadly, my findings do not support theirs. My use of predictors only from age 10, and socio-economically disadvantaged sample are the most likely explanations for the different findings.

4.2.4 Wellbeing in the ADHD subgroup

The results of the preferred model on the ADHD subgroup (R^2 = 7.1%) were within the realm of practically significant (R^2 > 4%; Ferguson, 2009), and more thought-provoking in terms of support for state regulation theory, with significant effects of CSTRESS (*ß* = -0.108*), ZEXT (*ß* = -0.115*), and ZREAD (*ß* = 0.166**).

The CSTRESS variable (chronic stressors), was a simple count based on 22 items indicating problems with health, school, and relationships, identified as be stressful based on content of other stressful item questionnaires for children (see chapter 6). Of the 22 items, eight were rated by a teacher, seven by a parent, and seven by a medical professional. The use of items from multiple independent raters makes this measure less likely to be affected by bias that can be a concern with parent-only rated measures. However, it has not been validated in other samples or against other chronic stress measures.

The significance of ZEXT, i.e. externalising problems, supports evidence from elsewhere that comorbidity is important in long-term outcomes for ADHD (Costello & Maughan, 2015). Additionally, the relatively large and significant effect of ZREAD in this group, combined with a lack of effect for MATHS, was remarkable and warrants further investigation. The implication is that reading ability may be more important for long-term wellbeing than maths ability within the ADHD group. In the other samples (unmatched and matched), either both reading and maths abilities were important, or only maths. The significance of ZREAD however does support findings elsewhere in literature more generally that IQ may be an important factor in adult outcomes for ADHD (Costello & Maughan, 2015).

On reflection, it was interesting that the ZINT variable, representing internalising problems, was not a significant predictor of adult subjective wellbeing for the ADHD group. As described in chapter 6, internalising problems are broad indicator of 'within self' problems experienced in childhood, which include symptoms of anxiety, depression, and somatization. Subjective wellbeing has been measured in other studies of BCS70 (e.g. Goodman et al., 2015) as a low score on the Rutter Malaise[55] Inventory (Butler et al., 1997; Rutter et al., 1970). The Malaise Inventory items are similar in meaning/content to the SDQ internalising subscale used to produce the ZINT variable, so one might expect that (high) ZINT measured in childhood would predict (low) wellbeing in adulthood. However, I re-evaluated the relationship within the ADHD subgroup with a univariate linear regression, and the lack of effect held: childhood ZINT did not predict adult SWB (N=360, $F(1, 367) = 1.56$, $p = 0.213$, $Adj\ R^2 = 0.001$). This evidence indicates that childhood internalising problems do not necessarily lead to adult internalising problems for those with childhood ADHD. Other studies have also drawn the conclusion that comorbid anxiety with ADHD does not increase the likelihood of negative life outcomes, and may relate to a potential protective influence of anxiety on reduced impulsive behaviours and improved effectiveness of treatment (Jensen et al., 2001; Kessler et al., 2005; Obsuth et al., 2020; Schatz & Rostain, 2006).

[55] Refers to negative affect, often equated with depression and/or anxiety.

5 Robustness check of SWB factor score as a measure of wellbeing

Two of the regression models were replicated using the Warwick Edinburg Mental Wellbeing scale (WEMWBS), as a robustness check of the SWB factor score measure. The detail of the analysis is included in the chapter 7 appendix.

Sample differences and interactions were reasonable explanations for the minor differences between the results for the two outcome measures, and thus provided some evidence that WEMWBS and SWB measured the same construct.

Chapter 8 Childhood ADHD, stress, and adult educational attainment

1 Introduction

Chapter 8 reports a collection of analyses of the relationships between ADHD and stress measured in childhood at age 10, and educational attainment outcomes, measured in adulthood at age 42 or 34. Educational attainment, although an outcome in its own right, was selected as a proxy for socioeconomic status, and indicator of objective wellbeing.

Reflecting back to the pilot study reported in chapter 5, education level was measured there using an 8-level scale of academic qualifications reported at age 42. Subsequent to the pilot, in chapter 6, the measurement was refined. Data from age 34 was extracted if possible where it was not available at age 42, and the 8-level scale was recoded to four variations which were shorter, more balanced scales. The four variations were:

1) 0-4 level academic qualifications (OEDUC4)
2) 0-4 level NVQ[56] (academic + vocational qualifications: NVQ4)
3) 0-1 level high/low academic qualifications (OEDUC2)
4) 0-1 level high/low NVQ (academic + vocational qualifications: NVQ2)

Three samples were used for analyses (the same were used in chapter 7):

9) The full, unmatched sample (N=8,519);
10) the matched sample pruned and weighted using coarsened exact matching to improve the balance between ADHD and 'controls' on a key set of confounds (N~6,207)[57]. In effect this was a relatively socio-economically disadvantaged sample; and
11) the ADHD subgroup sample (n=369).

2 Method

2.1 Further exploration of the educational attainment measures

I began the analyses with exploration of how the observations were distributed by level for different variations of the outcome measure.

First, observations were cross-tabulated to reveal how the they were spread across the 0-4 level academic levels in three samples of interest: unmatched, weighted matched, and ADHD subgroup only (Table 74).

[56] National Vocational Qualifications
[57] The ~ symbol is used because the matched sample includes observations weighted at both less than and greater than one, but not those weighted at zero.

0-4 level academic		Unmatched		Weighted matched		ADHD subgroup	
0	No qualifications	28%	2421	34%	2896.44	46%	170
1	GCSE D-E	7%	604	9%	771.58	13%	47
2	GCSE A-C	26%	2221	27%	2307.37	23%	86
3	A-levels	6%	492	5%	412.34	5%	20
4	Degree	33%	2781	25%	2131.28	12%	46
			8519		8519.00		369

Table 74. Distribution of observations for the 0-4 level academic EDL measure

N.B. qualification level names in this table represent the main academic qualification. They also include similar qualifications (e.g. O-levels, CSEs, HE diploma…).

Overall in the unmatched and weighted matched samples, levels 1 (GCSE D-E) and 3 (A-levels only) accounted for smaller proportions of observations than the other three levels. This was not surprising because both categories represent a generally unplanned academic trajectory: cohort members who opted to take O-levels but achieved poor results, and those who opted to take A-levels but then did not proceed to university. Since each category was modelled separately (as nominal) against a reference category, these smaller groups had less power, i.e. it was more difficult to detect smaller effects.

Based just on percentages, there were marked differences between the ADHD subgroup and the larger samples, unmatched and matched. It was much more likely for the ADHD group to have no qualifications or level 1 (GCSE D-E), and much less likely for the ADHD group to have a degree. The differences were smaller between the weighted matched group and the ADHD group, indicating the matching variables were important for balancing the sample. These comparisons in Table 74 were made between the whole samples (including ADHD) and the ADHD-only group. Differences between ADHD and non-ADHD samples are explored next with relative risk ratios.

2.1.1 Relative risk ratios

Next, I examined the difference between the spread of observations for ADHD vs. non-ADHD in the unmatched sample. I chose to use relative risk ratios (RRRs) because they are preferable over Odds Ratios (ORs) when the prevalence of the event (e.g. 0 - no qualifications) is not rare in the data; (ORs) are most useful when the prevalence of an event is low (i.e. < 10%) (Chen et al., 2010). RRRs also have the advantage of being straightforward to calculate and interpret. For example, from Table 75 below I can conclude that the risk of being in the 0 (no-qualifications) level for the academic-only measure is 67% higher for a cohort member (CM) in the ADHD subgroup than a CM in the non-ADHD group. Similarly, it is 11% less likely for an ADHD person to be in level 2 (GCSE A-C) than non-ADHD. I applied the method twice: once using the academic qualifications only outcome, and a second time using the NVQ outcome, which included both academic and vocational qualifications, for comparison.

0 – 4 Academic levels

		Non-ADHD		ADHD		RRR
0	No qualifications	2,251	27.62	170	46.07	1.67
1	GCSE D-E	557	6.83	47	12.74	1.87
2	GCSE A-C	2,135	26.2	86	23.31	0.89
3	A-levels	472	5.79	20	5.42	0.94
4	Degree	2,735	33.56	46	12.47	0.37
		8,150		369		

0-4 NVQ levels (academic + vocational)

		Non-ADHD		ADHD		RRR
0	No qualifications	881	10.81	71	19.24	1.78
1	NVQ1	628	7.71	46	12.47	1.62
2	NVQ2	2,070	25.41	120	32.52	1.28
3	NVQ3	1,213	14.89	60	16.26	1.09
4	NVQ4	3,355	41.18	72	19.51	0.47
		8,147*		369		

Table 75. ADHD vs non-ADHD Relative Risk Ratios by educational attainment level (EDL) in the unmatched sample

ADHD RRR = ADHD% / non-ADHD%
** There were 3 missing values for NVQ. All 3 of the observations were categorised as level 0 within the academic measure*

As expected, the relative risk of having one of the lower two education levels was higher for the ADHD than the non-ADHD group with both EDL measures. However, unexpectedly, the risk of no qualifications was higher when vocational qualifications were included. Using the academic-only measure, the ADHD group was less likely than the non-ADHD group to have attained any of the top three education levels: GCSE A-C, A-levels, or a degree. When vocational qualifications were included, however, the ADHD group was more likely to be in the middle levels (NVQ 2 and 3, or GCSE A-C and A-levels or vocational equivalents). Using either measure, the ADHD group was less likely to be in the top level (4 - degree or similar). Since there is a slightly different pattern of outcomes for the ADHD subgroup, the NVQ measure of EDL is used for some regression model comparisons in the next section.

0-1 level high/low academic qualifications

Next, to obtain a broader view on outcomes, I conducted a similar analysis on the unmatched and weighted matched samples using a binary outcome for academic qualification, and examined the distribution by sex and ADHD vs. non-ADHD (Table 76).

Unmatched sample

		Girls	Boys	Total	Pct	RRR
Non-ADHD	Low EDL	2,525	2,418	4,943	60.65%	0.74
	High EDL	1,735	1,472	3,207	39.35%	2.20
	Total	4,260	3,890	8,150		
ADHD	Low EDL	102	201	303	82.11%	1.35
	High EDL	25	41	66	17.89%	0.45
	Total	127	242	369		

Weighted matched sample

		Girls	Boys	Total*	Pct	RRR
Non-ADHD	Low EDL	1395.87	2683.83	4079.70	69.45%	0.84
	High EDL	597.41	1196.89	1794.30	30.55%	1.72
	Total	1993.28	3880.72	5874.00		
ADHD	Low EDL	89	185	274	82.28%	1.18
	High EDL	24	35	59	17.72%	0.58
	Total	113	220	333		

Table 76. Tabulation of 0-1 level high/low academic EDL by sample, ADHD subgroup, and sex, with RRR

Non-ADHD RRR = Non-ADHD% / ADHD%
ADHD RRR = ADHD% / non-ADHD%
** These cross-tabs were done in Stata, and Stata does not include observations in counts that have a weight of 0.*

This tabulation shows it is 1.35 times more likely for an ADHD subgroup member to be in the low education group (regardless of sex) in the unmatched sample, and 1.18 times more likely in the weighted matched sample. Interestingly the difference between girls and boys in the ADHD High EDL group is much less than it is in the non-ADHD group. Guidelines for relative risk ratio (RRR) effect sizes have been suggested at 1.22, 1.86, and 3.00 for small, medium, and large respectively (Olivier et al., 2017). Using these guidelines, the effect in the unmatched sample was classed as small, whilst in the matched sample, there was no appreciable effect. Thus, using a matched sample produces a reduced effect of ADHD on educational attainment.

The difference in absolute risk is an important factor that is often overlooked in the interpretation of relative risk (Noordzij et al., 2017). The risk of low education in the weighted matched sample was 82.28% for the ADHD group, and 69.45% in the non-ADHD group, yielding an absolute risk difference of 12.83%. 12.83% is enough of a difference to warrant further investigation of the robustness of the effect size. Also, RRR is a simple ratio and does not provide a standard error or confidence interval, so logistic regression was used to evaluate the effect with more precision.

2.2 Predictors and covariates of educational attainment (EDL)

Before proceeding to the regression models, the full set of relevant covariates is identified for EDL. The set includes core variables measuring ADHD, stress, and known ADHD outcome predictors, plus a list of covariates from the literature found to best predict both ADHD and EDL based on the vselect procedure (see chapter 6). Variables included in the matching procedure are <u>underlined</u>.

12 core covariates:

> `ADHD ADHDSEV`
> (ADHD and severity)
>
> `ESTRESS CSTRESS LOCTHETA SETHETA ZLEIS`
> (life event stressors, chronic stressors, locus of control, self-esteem, engagement in leisure)
>
> <u>`SEX`</u> `LOWSES MATHS ZEXT ZINT`
> (sex, low SES/free school meals, maths, externalising and internalising problems)

14 covariates from the ADHD predictors literature review which also predicted EDL per vselect.

> <u>`DADED`</u> `MUMED EPVT BFED AUTHCRV HOMEUNTI SLPOOR6` <u>`BACKWARD`</u>
> `POSTTERM ECZ` <u>`HOMELOW`</u> <u>`WHEEZ`</u> `UNMAR` <u>`SMOKELVL`</u>
> (Father's education, mother's education, picture vocabulary test, breastfed, authoritarian child rearing views, home untidy, poor sleep 1st 6 months, backward development, post-term, eczema, home low standard, wheezing, unmarried at birth, mother's pregnancy smoking level)

3 additional confounds based on literature and the vselect procedure, using ADHD as the outcome.

> <u>`MUMMAL`</u> <u>`SMOKE`</u> `POORNBHD`
> (Mother malaise, mother smoked during pregnancy, poor neighbourhood)

At this stage, 29 variables were identified as relevant for EDL. Next missingness was evaluated for this set of covariates, since it was different to the set used for SWB.

2.3 Missingness

Patterns of missingness affecting >2% of observations for this set of covariates were similar to those identified for SWB, in that most related to data missing on all variables from the age 0 and/or 5 sweeps. However, 12% of missingness was on mother's education level at age 5 (MUMED), and 3% on the post-term pregnancy indicator at age 0 (POSTTERM), and these were new patterns. I created missing indicator variables for these, and used Mirador (Sabeti Lab at Harvard University et al., 2018) to explore which variables were most related to the missingness. For MUMED it was SMOKE and PRETERM, and for POSTTERM it was UNMAR and SMOKE. Only PRETERM (pre-term birth indicator) was not already in the list of covariates above, so PRETERM was added, to support the FIML estimation.

Thus, a total of 30 variables were analysed as predictors and covariates for education level (the eight matching variables are underlined).

ADHD ADHDSEV
ESTRESS CSTRESS LOCTHETA SETHETA ZLEIS
<u>SEX</u> LOWSES MATHS ZEXT ZINT
<u>DADED</u> MUMED EPVT BFED AUTHCRV HOMEUNTI SLPOOR6 <u>BACKWARD</u>
POSTTERM ECZ <u>HOMELOW</u> <u>WHEEZ</u> UNMAR <u>SMOKELVL</u>
<u>MUMMAL</u> <u>SMOKE</u> POORNBHD
PRETERM

2.4 Estimation approach

In the chapter 5 pilot study, OLS (linear) regression was used to estimate EDL models, but it was noted this was not ideal given the distribution of the data, and adjustments for a more appropriate modelling approach were deferred. Thus, for the chapter 8 analysis, the proportional odds assumption was re-tested in the updated sample with a user-written Stata package called omodel, and a logit regression of EDL on ADHD subgroup and ADHD severity. Violation of proportional odds was confirmed (χ^2 was significant at 95% CI) for both of the updated measures of EDL that could potentially be modelled as ordered categorical (OEDUC4 and NVQ4). Thus, an alternative approach should still be taken.

If the proportional odds assumption is not met for an ordered categorical outcome, (as is the case here) Mplus documentation recommends modelling the outcome as nominal, or unordered categorical (Muthén & Muthén, 2017). This approach significantly increases complexity, since separate coefficients are estimated for each 'category' of education level and each covariate. To test the limits of complexity I could model with this type of outcome, I attempted to estimate a regression of a nominal EDL on the full set of 30 covariates identified as relevant, and the model failed to estimate. The output recommended using the

MONTECARLO integration option, which did then allow the model to estimate successfully[58].

2.5 Regression models

Given after chapter 7 it was clear that only a limited set of models were useful, here in chapter 8 I only used the unmatched sample to build models for comparing treatment effect sizes between methods, and all other models were based on the matched sample or ADHD subgroup sample. For clarity nonetheless the sample (matched or unmatched) is noted under the caption of the table.

Models were fitted in a progression, similar to the one used to analyse SWB. Simple univariate regressions of EDL on ADHD were modelled first, tested with each of the four different measures for comparison. Next, EDL measures were regressed on the continuous ADHD severity variable. A third set of models used the unmatched sample, and regressed EDL on ADHD controlling for the matching covariates. This allowed for a comparison between regression and matching methods for estimating effects, controlling for the same variables. Fourth, EDL was regressed on all 30 relevant covariates. Fifth, models were fitted regressing EDL on all the core variables plus those relevant from the all-covariates model (4). Finally, models were fitted on the ADHD subgroup sample. Multiple models were fitted to identify an optimal compromise between parsimony and completeness. Adjustments were not made for multiple comparisons.

2.5.1 Notes on logistic model reporting

Odds ratios (ORs) were reported when an effect size was needed (i.e. preferred models). Odds ratios < 1 were inverted to allow for ease of interpretation and comparison to effect size thresholds. Thresholds used to evaluate odds ratios for outcomes with > 10% prevalence were: 1.32 = small, 2.38 = medium, 4.79 = large (Olivier et al., 2017).

Unstandardized (log-odds) coefficients, which only provide direction and significance, were reported for the multivariate models where effect size was not needed. Positive coefficients indicated that the probability was higher for membership in the group of interest compared to the reference group, for higher values of X. (Muthén & Schultzberg, 2017) In all of the

---

[58] Initially the montecarlo model took over four hours to run. Mplus discussion boards advised that performance could be improved by reducing the number of integration points (to 500 or 1,000), or by using a BAYES estimator. It is not possible to use sample weights with BAYES in Mplus, so I updated subsequent models to use MLR with MONTECARLO integration with 500 points. To ensure this was sufficient integration I confirmed there were no negative ABS changes between loglikelihoods in the TECH8 output (Muthén, 2016).

Chapter 8/EDL models, the reference group was coded to be the highest educational attainment level.

Mplus provides the McKelvey & Zavoina (MZ) pseudo R^2 (McKelvey & Zavoina, 1975) for categorical outcome models (Muthen & Muthen, 2004). The MZ-R^2 is an approximation of relative proportion of variance explained, and assumes there is a continuous latent construct underlying the categorical indicator (McKelvey & Zavoina, 1975). Simulations have shown M-ZR^2 to be the most similar to the R^2 calculated by OLS regression amongst the most commonly used pseudo-R^2 variations, and least impacted by the prevalence of success (i.e. outcome variable = 1) in the data (Veall & Zimmerman, 1996; Windmeijer, 1995). However, MZ-R^2 should not be interpreted for a single model in isolation (for example, to calculate effect size), and cannot be assumed synonymous with OLS variance explained (Bo et al., 2006). MZ-R^2 can be compared between models on the same outcome to indicate which explains relatively more or less variance (Bo et al., 2006), and viewed as supplementary to variable-level estimates (Peng et al., 2002).

The models described in the subsequent paragraphs do not include the same fit statistics that were reported with the SWB outcome (RMSEA, etc.). Those statistics are based on means, variances, and covariances, which are not interpretable with nominal or categorical outcomes (Muthen & Muthen, 2015). For the models predicting binary outcomes, MZ-R^2 was reported (discussed above). For most of the nominal outcome models, global measures of fit were indeterminable because there were zero degrees of freedom, which is common when SEM is used to estimate single path models (Muthen & Muthen, 2007). If appropriate, BIC was used to make comparisons between models that estimated the same dependent/outcome variable.

2.5.2 List of models

Model 1 – 0-4 level academic EDL on ADHD
Model 1a – 0-4 level NVQ EDL on ADHD
Model 1b – 0-1 level high/low academic EDL on ADHD
Model 2 – 0-4 level academic EDL on ADHD severity
Model 2a – 0-1 level high/low academic EDL on ADHD severity
Model 3 – 0-4 level academic EDL on ADHD controlling for matching covariates in the unmatched sample
Model 3a – 0-1 level high/low academic EDL on ADHD controlling for matching covariates in the unmatched sample
Model 4 – 0-4 level academic EDL on all 30 relevant covariates

Model 5 – 0-1 level high/low academic EDL on core predictors plus significant covariates (at any level) from model 4
Model 5a - model 5 with 0-1 level high/low NVQ, EDL by sex
Model 5b - model 5 with ADHD variable removed and ADHDSEV added
Model 6 – 0-1 level academic EDL on selected core set of variables, in the ADHD subgroup
Model 6a – 0-1 level academic EDL on three significant variables from model 6, in the ADHD subgroup

2.5.3 Models

Model 1 – 0-4 level academic EDL on ADHD

Odds were higher for achieving lower educational attainment levels in the ADHD group compared to the degree-level reference category (Table 77). Effect sizes were medium for levels 0 and 1, small for level 2 and not significant for level 3.

Level	OR	p-value	sig
0 - no qualifications	2.842	< 0.001	***
1 - GCSE D-E	3.054	0.005	**
2 - GCSE A-C	1.842	0.023	*
3 - Any A levels	2.350	0.055	~

Table 77. 0-4 level academic EDL on ADHD

**** p<0.001, ** p<0.01, * p<0.05, ~ p > 0.05 but < 0.20*
Matched sample; reference category = level 4: he diploma, degree or higher degree

Model 1a – 0-4 level NVQ EDL on ADHD

When vocational qualifications were included, the results were still significant and again indicated ADHD subgroup cohort members (CMs) were more likely to have a lower EDL than the highest (degree or similar) level. All of the OR effects sizes here were small (i.e. between 1.32 and 2.38), and all were significant. Thus, when vocational qualifications were included in the EDL outcome variable, it discriminated less between ADHD and non-ADHD cohort members.

Level	OR	p-value	sig
0 - no qualifications	2.180	0.005	**
1 - NVQ1	2.284	0.010	*
2 - NVQ2	1.948	0.004	**
3 - NVQ3	1.914	0.013	*

Table 78. 0-4 level NVQ EDL on ADHD

**** p<0.001, ** p<0.01, * p<0.05, ~ p > 0.05 but < 0.20*
Matched sample; reference category = level 4: degree or vocational equivalent

Model 1b – 0-1 level high/low academic EDL on ADHD

The model reported in Table 79 predicted a two-level academic-only outcome, where low = levels 0-2 and high = levels 3-4. Again, this model indicated ADHD cohort members were significantly more likely to be in the lower education group (GCSEs/O-levels or below) than non-ADHD. The effect size here was small, and since this was modelled as a binary (vs. nominal) outcome, Mplus provided an MZ-R^2 statistic (0.008).

OR	p-value	sig
2.043	0.001	**
MZ-R^2	0.008	
BIC	10361.741	

Table 79. 0-1 level high/low academic EDL on ADHD

Matched sample; reference category = level 0: A levels/similar or above

N.B. In the previous section (1.2), the RRR for ADHD vs. non-ADHD in the matched sample was 1.18 for the low education group, i.e. the risk was 18% higher of being in the low education category for the ADHD group. The OR of 2.043 is a much larger number. Relative risk ratios and odds ratios are often used in similar contexts but are calculated on different scales (probability vs. odds): Risk ratio = % with low education in treatment (ADHD) group / % with low education in control (non-ADHD) group The disparity between the two statistics is greater when the prevalence of the outcome event in the sample is > 10% (Chen et al., 2010; Davies et al., 1998; Olivier et al., 2017; Windmeijer, 1995)*, and in this case the prevalence for low education is ~70% in the matched sample.*

Model 2 – 0-4 level academic EDL on ADHD severity

Next, the 0-4-level academic EDL outcome was regressed on the continuous ADHD severity score in the matched sample. The direction and significance of effects were similar to the comparable model on the ADHD subgroup indicator (Model 1). The relationship reversed direction for level 3, although the OR was not significant. This could be caused by the relatively small sample size in level 3, and/or suggest a nonlinear relationship between ADHD severity and educational attainment. Box plots were produced to examine visually, and there did appear to be a slightly nonlinear pattern (Figure 32). I do not plan to explore the pattern further here, but it could be an interesting topic for future research.

Level	OR	p-value	sig
0 - no qualifications	2.005	< 0.001	***
1 - GCSE D-E	2.402	< 0.001	***
2 - GCSE A-C	1.495	< 0.001	***

3 - Any A levels	0.892	0.320	*ns*

Table 80. Multinomial regression of 0-4 level academic EDL

**** p<0.001, ** p<0.01, * p<0.05, ~ p > 0.05 but < 0.20*
Matched sample; reference category = level 4: he diploma, degree or higher degree

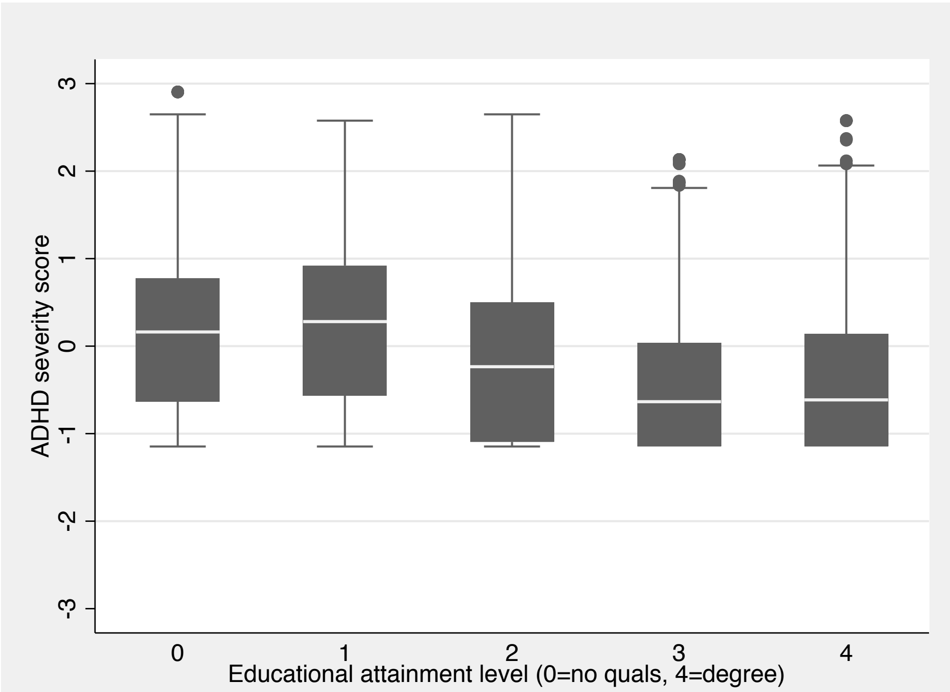


Figure 32. Box plot of 0-4 level academic EDL vs. ADHD severity

Model 2a – 0-1 level high/low academic EDL on ADHD severity

Another variation of the model tested the effect of ADHD severity (ADHDSEV) on the 0-1 level high/low academic EDL outcome, for girls and boys combined. The effect was significant and small in size (*OR* = 1.858***). The MZ-R^2 here (0.087) was much larger than for model 1b (EDL on ADHD, MZ-R^2 = 0.008), indicating ADHDSEV explained more variance than ADHD, which is to be expected given ADHDSEV is a continuous variable. The BIC was also smaller for model 2a vs. model 1b, indicating a relatively better fit.

Odds Ratio	p-value	sig
1.858	< 0.001	***
MZ-R^2	0.087	
BIC	9907.047	

Table 81. 0-1 level high/low academic EDL on ADHD

Matched sample; reference category = level 0: A-levels or similar and above

The next two models calculated treatment effects in the unmatched sample controlling for the variables used to create the matched sample. They provide a comparison between a regression methods and matching method (Model 1) to calculate treatment effects. I returned to using the ADHD binary indicator variable, because it is more widely reported and thus comparable to other literature.

Model 3 – 0-4 level academic EDL on ADHD controlling for matching covariates in the unmatched sample

The treatment effect sizes for ADHD calculated in the unmatched sample in model 3 were larger than those observed in model 1, demonstrating that use of the matched sample produces different estimates than regression on a full sample with matching variables as covariates. However, the effect sizes, direction and significance pattern were similar.

Model 3 (regression)				Model 1 (matching)		
Level	OR	p-value	sig	OR	p-value	sig
0 - no qualifications	3.148	< 0.001	***	2.842	< 0.001	***
1 - GCSE D-E	3.567	0.001	**	3.054	0.005	**
2 - GCSE A-C	1.950	0.009	**	1.842	0.023	*
3 - Any A levels	2.470	0.030	*	2.350	0.055	~

Table 82. Model 3 compared to Model 1 (controlled regression vs. matching methods, controlling for the same covariates, 0-4 level academic EDL)

**** p<0.001, ** p<0.01, * p<0.05, ~ p > 0.05 but < 0.20*
Unmatched sample; reference category = level 4: he diploma, degree or higher degree
N.B. estimates for matching covariates are not included in the table for simplicity of presentation.

Model 3a – 0-1 level high/low academic EDL on ADHD controlling for matching covariates in the unmatched sample

The treatment effect size for ADHD in the unmatched sample using a binary EDL outcome and controlling for matching covariates was still small (OR = 2.185), but slightly larger than the effect calculated using the matched sample (Model 1b: OR = 2.043). The MZ-R^2 for the full model was 0.147; much larger than model 1b on the matched sample with ADHD as the only predictor (0.008).

OR	p-value	sig
2.185	< 0.001	***

Table 83. 0-1 level high/low academic EDL on ADHD, controlling for matching covariates

**** p<0.001, ** p<0.01, * p<0.05*
Unmatched sample; reference category = level 0: A levels/similar or above
N.B. estimates for matching covariates are not included in the table for simplicity.

Model 4 – 0-4 level academic EDL on all 30 relevant covariates

Next, the 0-4 level academic EDL was regressed on all 30 relevant covariates to reveal significant predictors for use in subsequent models. Unstandardized log-odds coefficients were reported, and no MZ-R^2 stats were available.

ADHD and ADHDSEV did not jointly have significant effects at any level. The switching between positive and negative signs for the coefficients for the two variables together in the same model suggested spurious effects from collinearity, as has been observed previously with these two variables in the same model. As expected, higher parental education level (DADED, MUMED), and maths and reading scores at age 10 (MATHS, ZREAD) had a negative and significant association with lower educational attainment. Higher locus of control at age 10 (LOCTHETA) was significantly associated with lower likelihood of membership in one of the lower three categories, and higher self-esteem (SETHETA; *b* = 0.256*) was associated with a lower probability of membership in category 1 (GCSE D-E).

	0:no quals		1:GCSE D-E		2:GCSE A-C		3:Any A levels	
Covariate	Est	sig	Est	sig	Est	sig	Est	sig
ADHD	0.005		-0.245		0.010		1.631	***
ADHDSEV	0.206	*	0.538	***	0.180	~	-0.259	~
AUTHCRV	-0.684	*	0.133		-0.206		-0.528	~
BACKWARD	0.350		0.926	~	0.143		-2.735	**
BFED	-0.265	~	-0.028		-0.234	~	0.302	
CSTRESS	-0.075	~	-0.067		-0.183	**	-0.112	
DADED	-0.291	***	-0.328	***	-0.254	***	-0.021	
ECZ	-0.020		0.731	*	0.263	~	0.184	
EPVT	-0.121	~	-0.104		-0.089		-0.298	*
ESTRESS	0.046	~	-0.040		-0.010		-0.007	
HOMELOW	0.356		0.715		-0.582		-1.722	*
HOMEUNTI	0.204		-0.020		0.248		0.603	
LOCTHETA	-0.209	*	-0.390	**	-0.236	**	-0.045	
LOWSES	0.437	*	-0.081		0.130		0.252	
MATHS	-0.499	***	-0.418	**	-0.263	*	-0.089	
MUMED	-0.243	**	-0.723	***	-0.088	~	-0.057	
MUMMAL	0.120		-0.128		0.116		0.238	
POORNBHD	-0.395	~	-1.205	**	-0.396	~	-0.548	~
POSTTERM	-0.037		-0.472	~	0.010		-0.090	
PRETERM	-0.032		0.228		0.095		0.627	
SETHETA	0.041		0.256	*	0.021		0.242	~
SEX	0.241	~	0.110		0.104		-0.154	

SLPOOR6	-0.186		-0.108		-0.140		0.044	
SMOKE	0.076		0.261		0.673	~	0.205	
SMOKELVL	0.024		-0.077		-0.192		0.096	
UNMAR	0.598	*	0.448		0.515	~	-0.151	
WHEEZ	-0.423	*	-0.344		0.199		0.143	
ZEXT	0.140	*	-0.069		0.086		-0.025	
ZINT	-0.042		0.087		0.008		0.203	*
ZLEIS	0.084	~	0.069		0.017		0.146	~
ZREAD	-0.427	***	-0.610	***	-0.193	~	0.370	*

BIC: 368387.344

Table 84. 0-4 level academic EDL regressed on all 30 relevant predictors and covariates

**** p<0.001, ** p<0.01, * p<0.05, ~ p > 0.05 but < 0.20*
Matched sample; reference category = level 4: HE diploma, degree or higher degree

We see here some unexpected behaviour for the chronic stressors count (CSTRESS) variable. The model estimate had a significant negative coefficient (b = -0.183**) for level 2, which indicated that higher levels of CSTRESS were associated with higher probability of membership in the 4 (degree) group, compared to the 2 (GCSE A-C group). CSTRESS was a count based on 22 chronic stressor items measured at age 10, and most items were problems experienced at/with school, or ongoing medical problems. I expected these types of stressors to have a negative, not positive, impact on EDL. The effect observed here could be a genuine effect, or could be spurious due to collinearity with other variables. To test this, I ran a simple regression of OEDUC4 (0-4 level academic EDL) on CSTRESS. This model indicated a cohort member was slightly and significantly ($p < .001$) more likely to be in one of the lower two groups (0 or 1 vs. the degree group) for higher levels of chronic stressors at age 10, and there were no significant effects for groups 2 or 3. Thus the positive effect in model 4 (suggested link with higher education) of CSTRESS at level 2 does appear spurious and suggests that collinearity with one or more other variables is an issue.

The coefficients and p-values at each educational attainment level contain potentially interesting information about the direction and significance of specific factors at specific levels; for example, the likelihood of stopping education at A-levels. However, my research questions were not posed at that level. Thus, the remaining models focused exclusively on the 0-1 high/low academic educational attainment outcome (OEDUC2r), where 0 = level 3 (A-levels) or higher, and 1 = level 2 (GCSE A-C) or below. The key findings were similar regardless of how the outcome was modelled, and the simpler binary outcome also provides the MZ-R^2 statistic, which allows for comparison of variance explained between models.

Since the coefficient for SEX was fairly large and bordered on significance for the no qualifications category (b = 0.241, p = 0.07), and there are statistics available from other research on ADHD in BCS70 reported separately for girls and boys that can be used as a comparison (Brassett-Grundy & Butler, 2004), subsequent models were fitted separately for girls and boys. Finally, ADHDSEV was removed from the model, to focus on the effects of membership in the ADHD subgroup alone without the spurious effects from collinearity.

Model 5 – 0-1 level high/low academic EDL on core predictors plus significant covariates[59] from model 4

For this model logistic odds ratios (and inverse odds ratios[60]) were reported instead of coefficients, because they are interpretable as effect sizes.

Without ADHDSEV in the model, the ADHD indicator was not significant for boys or girls. DADED, MATHS, and ZREAD were significant with small effect sizes for both sexes. AUTHCRV (authoritarian child-rearing views; Inv OR = 1.883**), MUMED (Inv OR = 1.379***) and LOCTHETA (locus of control; Inv OR = 1.399***) were also significant and small for girls. Higher AUTHCRV, an authoritative parenting views indicator, was associated with higher educational attainment for girls, which was unexpected. Usually authoritarian parenting is associated with lower attainment (Pinquart, 2016). However, the effect held in univariate regressions. MUMED and LOCTHETA had p-values < 0.05 for boys, but did not reach practical significance for non-rare events (OR > 1.32; Olivier et al., 2017).

Additionally, for girls, ZLEIS (engagement in leisure; OR = 1.217*) was associated with lower attainment and POORNBHD (Inv OR = 1.626***) was associated with higher attainment, which were both the opposite direction expected. Univariate regressions showed both of these variables had a relationship with EDL in the expected/reverse direction, so the findings in this model (5) were probably spurious due to collinearity.

Finally, for boys, ZEXT (OR = 1.213*) significantly corresponded to lower educational attainment based on the p-value, but the OR was not large enough to be practically significant.

Girls

Covariate	OR	SE	Est-1/SE	p-value	sig	Inv OR
ADHD	0.698	0.218	-1.385	0.166	~	1.433

[59] Significant at any EDL from 0-4
[60] Inverse odds ratios allow ORs < 1 to be compared to OR effect size thresholds

AUTHCRV	0.531	0.170	-2.764	0.006	**	1.883
BACKWARD	2.131	2.152	0.525	0.599		
CSTRESS	0.930	0.056	-1.242	0.214		1.075
DADED	0.739	0.096	-2.707	0.007	**	1.353
ECZ	1.290	0.369	0.786	0.432		
EPVT	0.835	0.085	-1.945	0.052	~	1.198
ESTRESS	0.974	0.033	-0.787	0.431		1.027
HOMELOW	2.286	1.905	0.675	0.500		
LOCTHETA	0.715	0.069	-4.163	< 0.001	***	1.399
LOWSES	1.267	0.294	0.908	0.364		
MATHS	0.709	0.087	-3.348	0.001	**	1.410
MUMED	0.725	0.069	-3.986	< 0.001	***	1.379
POORNBHD	0.615	0.178	-2.163	0.031	*	1.626
SETHETA	1.068	0.135	0.508	0.612		
UNMAR	1.515	0.398	1.295	0.195	~	
WHEEZ	0.825	0.229	-0.764	0.445		1.212
ZEXT	1.119	0.101	1.180	0.238		
ZINT	1.094	0.089	1.063	0.288		
ZLEIS	1.217	0.093	2.335	0.020	*	
ZREAD	0.669	0.089	-3.703	< 0.001	***	1.495
R^2	0.359					
BIC	149161.12					
EDM0S5W_oeduc2r_nosev						

Boys

Covariate	OR	SE	Est-1/SE	p-value	sig	Inv OR
ADHD	1.254	0.318	0.796	0.426		
AUTHCRV	0.797	0.240	-0.846	0.397		1.255
BACKWARD	1.344	0.858	0.401	0.688		
CSTRESS	0.932	0.050	-1.361	0.174	~	1.073
DADED	0.756	0.043	-5.704	< 0.001	***	1.323
ECZ	1.104	0.238	0.436	0.663		
EPVT	0.974	0.085	-0.310	0.757		1.027
ESTRESS	1.030	0.030	0.978	0.328		
HOMELOW	0.687	0.672	-0.467	0.641		1.456
LOCTHETA	0.833	0.081	-2.049	0.040	*	1.200
LOWSES	1.326	0.379	0.862	0.389		
MATHS	0.664	0.073	-4.607	< 0.001	***	1.506

MUMED	0.855	0.060	-2.413	0.016	*	1.170
POORNBHD	0.701	0.254	-1.178	0.239		1.427
SETHETA	0.965	0.079	-0.448	0.654		1.036
UNMAR	2.096	0.786	1.394	0.163	~	
WHEEZ	0.936	0.158	-0.408	0.684		1.068
ZEXT	1.213	0.089	2.384	0.017	*	
ZINT	0.909	0.067	-1.355	0.175	~	1.100
ZLEIS	0.939	0.064	-0.955	0.340		1.065
ZREAD	0.637	0.074	-4.880	< 0.001	***	1.570

R^2 0.329

BIC 140351.29

EDM1S5W_oeduc2r_nosev

Table 85. 0-1 level high/low academic EDL regressed on core plus significant variables from model 4, by sex

**** p<0.001, ** p<0.01, * p<0.05, ~ p > 0.05 but < 0.20*
Matched sample; reference category = level 0: A levels or higher

Next, the 0-1 level high/low NVQ EDL outcome was evaluated for comparison to Model 5.

Model 5a - model 5 with 0-1 level high/low NVQ, by sex

The ADHD indicator was not a significant predictor of the NVQ-based educational attainment outcome for boys or girls.

For girls, mother's education level (MUMED) was highly significant (b = -0.476,***), but father's (DADED), and ZREAD were not. LOCTHETA (b = -0.194*), MATHS (b = -0.285*), and ZLEIS (b = 0.184) were also significant, and in the same direction as they were in the model using the academic-only outcome.

For boys, father's education level (DADED; b = -0.261***) achieved significance but not mother's. For boys in this model, some new variables appeared as influential: ECZ (b = 0.477*), HOMELOW (b = -1.308*), ZINT (b = -0.180**), and ZLEIS (b = -0.180**). MATHS (b = -0.453***) and ZREAD (b = -0.300**) were significant also, as they have been in other models using other outcome measures. ZINT, or internalising problems, and HOMELOW (low standard home at age 5) had effects in the opposite direction expected, i.e. higher scores were associated with a higher education level which may have been due to collinearity.

Girls

Covariate	Est	SE	Est/SE	p-value	sig
ADHD	-0.201	0.285	-0.707	0.480	

AUTHCRV	-0.318	0.294	-1.081	0.280	
BACKWARD	1.575	0.875	1.799	0.072	~
CSTRESS	0.085	0.057	1.487	0.137	~
DADED	-0.158	0.121	-1.310	0.190	~
ECZ	-0.007	0.275	-0.024	0.981	
EPVT	-0.149	0.094	-1.592	0.111	~
ESTRESS	-0.047	0.033	-1.448	0.147	~
HOMELOW	0.974	0.535	1.822	0.068	~
LOCTHETA	-0.194	0.091	-2.118	0.034	*
LOWSES	0.145	0.208	0.698	0.485	
MATHS	-0.285	0.111	-2.575	0.010	*
MUMED	-0.476	0.094	-5.038	< 0.001	***
POORNBHD	-0.244	0.283	-0.862	0.389	
SETHETA	0.096	0.102	0.936	0.350	
UNMAR	-0.014	0.262	-0.054	0.957	
WHEEZ	-0.119	0.238	-0.502	0.616	
ZEXT	0.129	0.079	1.636	0.102	~
ZINT	0.098	0.069	1.417	0.156	~
ZLEIS	0.184	0.068	2.704	0.007	**
ZREAD	-0.164	0.121	-1.356	0.175	~
MZ-R^2	0.321				
BIC	149909.656				

EDM0S5W_onvq2r_nosev

Boys

Covariate	Est	SE	Est/SE	p-value	sig
ADHD	0.042	0.210	0.202	0.840	
AUTHCRV	-0.272	0.255	-1.068	0.286	
BACKWARD	0.568	0.457	1.243	0.214	
CSTRESS	-0.084	0.053	-1.570	0.116	~
DADED	-0.261	0.054	-4.798	< 0.001	***
ECZ	0.477	0.188	2.535	0.011	*
EPVT	-0.123	0.081	-1.520	0.128	~
ESTRESS	0.037	0.028	1.309	0.191	~
HOMELOW	-1.308	0.647	-2.021	0.043	*
LOCTHETA	-0.101	0.095	-1.067	0.286	
LOWSES	-0.003	0.255	-0.013	0.990	
MATHS	-0.453	0.100	-4.527	< 0.001	***

MUMED	-0.047	0.070	-0.672	0.502	
POORNBHD	-0.444	0.313	-1.419	0.156	~
SETHETA	0.072	0.080	0.903	0.367	
UNMAR	-0.101	0.290	-0.350	0.727	
WHEEZ	-0.276	0.158	-1.743	0.081	~
ZEXT	0.139	0.068	2.036	0.042	*
ZINT	-0.180	0.068	-2.652	0.008	**
ZLEIS	-0.180	0.068	-2.638	0.008	**
ZREAD	-0.300	0.105	-2.857	0.004	**
MZ-R^2	0.263				
BIC	141122.422				
EDM1S5W_onvq2r_nosev					

Table 86. Multinomial regression of binary high/low NVQ education level cut at level 3 (NVQ 3): core plus significant covariates from full set

**** p<0.001, ** p<0.01, * p<0.05, ~ p > 0.05 but < 0.20*
Reference category = level 0: NVQ level 3 or higher

As discussed in the outcome measures section of chapter 6, educational attainment measures including vocational equivalents are controversial, and thus results of this model should be interpreted with caution. However here there were some interesting differences. For example, with the NVQ outcome, the parental education influence appeared to be gender-specific, with mothers influencing girls and fathers influencing boys. Also, for girls, engagement in leisure (ZLEIS) was associated with lower qualification levels, whilst for boys it was the opposite. Maths and reading ability at age 10 were both significant here for boys like they were in other models, but for girls, maths was significant and reading was not.

Model 5b - model 5 with ADHD variable removed and ADHDSEV added

Finally, ADHDSEV was evaluated separately, to show the effect of ADHD symptomatology regardless of whether it has met a threshold for the clinical disorder. Here, ADHDSEV was significant for boys, and on the borderline (p = 0.05) for girls. Log-odds coefficients were reported (Table 87).

DADED, LOCTHETA, MUMED, and ZREAD were significant for both girls and boys, though the significance level varied some. ZLEIS was significant only for girls, appearing to be a negative influence on education. ZEXT, which was significant in the previous model, became insignificant for boys with ADHDSEV in the model, indicating collinearity.

Girls

Covariate	Est	SE	Est/SE	p-value	sig

ADHDSEV	0.230	0.117	1.960	0.050	~
AUTHCRV	-0.593	0.315	-1.885	0.059	~
BACKWARD	0.778	1.013	0.768	0.443	
CSTRESS	-0.108	0.062	-1.725	0.084	~
DADED	-0.297	0.135	-2.194	0.028	*
ECZ	0.262	0.278	0.942	0.346	
EPVT	-0.183	0.102	-1.797	0.072	~
ESTRESS	-0.029	0.033	-0.867	0.386	
HOMELOW	0.963	0.837	1.150	0.250	
LOCTHETA	-0.325	0.097	-3.337	0.001	**
LOWSES	0.216	0.235	0.916	0.359	
MATHS	-0.310	0.122	-2.532	0.011	*
MUMED	-0.313	0.095	-3.277	0.001	**
POORNBHD	-0.448	0.293	-1.529	0.126	~
SETHETA	0.086	0.127	0.674	0.500	
UNMAR	0.426	0.264	1.611	0.107	~
WHEEZ	-0.177	0.280	-0.632	0.527	
ZEXT	0.010	0.095	0.108	0.914	
ZINT	0.074	0.082	0.901	0.368	
ZLEIS	0.195	0.078	2.516	0.012	*
ZREAD	-0.369	0.132	-2.788	0.005	**
MZ-R^2	0.365				
BIC	159694.629				

EDM0S5W_oeduc2r_noad

Boys

Covariate	Est	SE	Est/SE	p-value	sig
ADHDSEV	0.220	0.085	2.574	0.010	*
AUTHCRV	-0.200	0.294	-0.678	0.498	
BACKWARD	0.303	0.631	0.480	0.631	
CSTRESS	-0.094	0.055	-1.694	0.090	~
DADED	-0.280	0.057	-4.924	< 0.001	***
ECZ	0.106	0.219	0.483	0.629	
EPVT	-0.019	0.087	-0.223	0.823	
ESTRESS	0.029	0.030	0.994	0.320	
HOMELOW	-0.294	1.015	-0.289	0.772	
LOCTHETA	-0.192	0.096	-1.995	0.046	*
LOWSES	0.268	0.274	0.978	0.328	

MATHS	-0.383	0.110	-3.494	< 0.001	***
MUMED	-0.155	0.071	-2.196	0.028	*
POORNBHD	-0.344	0.364	-0.944	0.345	
SETHETA	-0.009	0.082	-0.106	0.915	
UNMAR	0.758	0.387	1.956	0.050	~
WHEEZ	-0.059	0.168	-0.354	0.724	
ZEXT	0.136	0.073	1.876	0.061	~
ZINT	-0.108	0.074	-1.471	0.141	~
ZLEIS	-0.055	0.067	-0.825	0.409	
ZREAD	-0.440	0.118	-3.723	< 0.001	***
MZ-R^2	0.333				
BIC	150606.929				
EDM1S5W_oeduc2r_noad					

Table 87. 0-1 level high/low academic EDL regressed on model 5 variables plus ADHDSEV minus ADHD, by sex

**** p<0.001, ** p<0.01, * p<0.05, ~ p > 0.05 but < 0.20*
Matched sample; reference category = level 0: A levels or higher

2.5.4 Selection of a preferred multivariate model for the matched sample

For a preferred model on the matched sample, I selected model 5. This model measured educational attainment as 0-1 level high/low academic qualifications only, with girls and boys reported separately and ADHDSEV removed from the model. This model was preferred because academic (non-vocational) qualifications are less controversial, there was evidence of different effects by gender, and collinearity caused by including ADHDSEV would have obscured the effect of belonging to the ADHD subgroup, which was needed to make comparisons to other research.

2.5.5 Educational attainment (EDL) in the ADHD subgroup (n = 369)

As discussed with the analysis of the wellbeing outcome in chapter 7, the ADHD subgroup sample was much smaller than the unmatched and matched samples, and power to detect effects and control for large numbers of independent variables was limited. A recent simulation study recommended a minimum sample size of 500 for logistic regressions in large observational datasets (Bujang et al., 2018). I have noted that at n=369, the sample is probably underpowered for any logistic model[61], so I sought to simplify and minimise the number of independent variables.

I used the simpler two-level education outcome, and combined girls and boys keep the sample size at 369. To select the list of covariates, I started with ten core predictors that have been prioritised throughout chapters 7 and 8: ADHDSEV, MATHS, LOWSES, ZEXT, ZINT, LOCTHETA, SETHETA, ZLEIS, CSTRESS and ESTRESS. I added SUBTYP, ZREAD, and SEX, as these are of particular interest within the ADHD group, and DADED because it has been a key predictor of EDL in the other models. The purpose of this 14-covariate overfitted model was to identify significant predictors, and build another model only using those, and increase power.

Model 6 – 0-1 level academic EDL on selected core set of variables, in the ADHD subgroup
ORs for Model 1 (Table 88) were reported, and if they were < 1, they were inverted (column: Inv OR) so they could be more easily interpreted as effect sizes. Only DADED (Inv *OR* = 1.742***), ZEXT (Inv *OR* = 1.637***), and ZREAD (Inv *OR* = 2.309***) were significant predictors, and they all could be interpreted as small appreciable effect sizes[62]. DADED and ZREAD were in the direction expected, i.e. higher levels correlated to higher educational attainment. ZEXT (externalising problems) however, was not in the direction expected, as higher levels also correlated to higher EDL. This is likely to be spurious due to collinearity, possibly due to correlations with ADHDSEV and/or SUBTYP. The next model (1a), did not contain either of those variables, so the effect was revisited there.

Covariate	OR	Inv OR	SE	Est-1/SE	p-value	sig
ADHDSEV	1.847		1.471	0.576	0.565	
CSTRESS	0.882	1.134	0.102	-1.159	0.246	
DADED	0.574	1.742	0.079	-5.425	< 0.001	***
ESTRESS	1.078		0.093	0.835	0.404	
LOCTHETA	0.801	1.248	0.220	-0.908	0.364	
LOWSES	0.944	1.059	0.578	-0.097	0.923	
MATHS	1.027		0.283	0.096	0.923	
SETHETA	1.229		0.338	0.677	0.498	
SEX	1.878		0.772	1.136	0.256	
SUBTYP	0.832	1.202	0.253	-0.662	0.508	

[61] It may be possible to achieve a sample size of 500 or more by attempting to retrieve educational attainment data on ADHD subgroup members identified at age 10 from other adult BCS70 sweeps (for example at age 30 and 38). I will investigate post-doc.
[62] Based on Inv OR > 1.32 = small, > 2.38 = medium, > 4.79 = large (Olivier et al., 2017)

ZEXT	0.611	1.637	0.089	-4.391	< 0.001	***
ZINT	0.925	1.081	0.159	-0.473	0.636	
ZLEIS	1.083		0.210	0.396	0.692	
ZREAD	0.433	2.309	0.137	-4.143	< 0.001	***
BIC	8264.023					
MZ-R^2	0.325					
EDMB_ADHDM1b_core3						

Table 88. 0-1 level high/low EDL regressed on a subset of core predictors in the ADHD subgroup only (*n*=369)

**** p<0.001, ** p<0.01, * p<0.05, ~ p > 0.05 but < 0.20*
ADHD subsample; reference category = level 0: A levels or higher

Model 6a – 0-1 level academic EDL on three significant variables from model 6, in the ADHD subgroup

In model 6a (Table 89), all three variables were again significant, and in the same direction, i.e. the positive effect for ZEXT remained[63]. However, the effect size for ZEXT (Inv *OR* = 1.309) was smaller here, and per the 1.32 OR threshold (Olivier et al., 2017) I have been using, was not quite large enough to be considered a small important effect. DADED and ZREAD did meet the threshold. The BIC here was smaller than for model 1, indicating a better fit. The MZ-R^2 estimate of variance explained was also smaller (0.247 here, vs. 0.325 in model 1), but not as much smaller as one might expect given the drop from 14 to 3 predictors.

Covariate	OR	Inv OR	SE	Est-1/SE	p-value	sig
DADED	0.590	1.695	0.075	-5.444	< 0.001	***
ZEXT	0.764	1.309	0.090	-2.619	0.009	**
ZREAD	0.496	2.016	0.107	-4.723	< 0.001	***
BIC	3179.237					
MZ-R^2	0.247					
EDMB_ADHDM1b_core4						

Table 89. 0-1 level high/low EDL regressed on significant predictors from model 6,

**** p<0.001, ** p<0.01, * p<0.05, ~ p > 0.05 but < 0.20*
ADHD subsample; reference category = level 0: A levels or higher

The final model evaluated the effect of ADHD severity on EDL, just within the ADHD subgroup.

[63] ZEXT was highly correlated with ADHDSEV, so I also tested this model with DADED, ZREAD and ADHDSEV. ADHDSEV was in the direction expected, but like ZEXT, was also not significant (*OR* = 1.761, *p* = 0.373) *File: EDMB_adhdM1_core5.*

Model 6b – 0-1 level high/low academic on ADHD severity in the ADHD subgroup

ADHD severity OR was sizable but was not significant. This was due to wide variation (95% CI: 0.974 – 4.832).

Odds Ratio	p-value	sig
2.169	0.187	ns
MZ-R^2	0.023	
BIC	348.323	

Table 90. 0-1 high/low academic EDL regressed on ADHD

ADHD subsample; reference category = level 0: A-levels or similar and above

3 Chapter 8 discussion

Chapter 8 answered the following research question about the relationship between childhood ADHD, stress, and adult educational attainment (EDL):

> RQ4: What is the effect of childhood ADHD on adult educational attainment using different methods and covariate sets to estimate treatment effects? Do the results support state regulation theory?

Four different variations of methods and models were used to test effects of ADHD on educational attainment. They were: 1) naïve regression, 2) controlled regression (on matching covariates), 3) matched sample regression, and 4) controlled matched sample regression. I reported on binary education outcomes, because the measures with more detail required more complex models, and the complexity was not required to answer the research questions.

3.1 Method comparison

Table 91 shows a four-way comparison between methods, three for girls and boys combined, one with them separate. The first three methods indicate a significant negative 'treatment effect' of ADHD on educational attainment level. The largest effect was found using naïve regression method (*OR* = 2.979***), indicating the odds for the ADHD group of being in the lower educational attainment group were 2.979 times the odds for non-ADHD. This is a medium effect size according to the recommended effect size guidelines for odds ratios for non-rare outcome events (Olivier et al., 2017). The effect size based on the other two methods was classed as small: for the controlled regression (*OR* = 2.185***), and in the matched sample (*OR* = 2.043**). The pseudo R^2 was larger in the unmatched univariate model (MZ-R^2 = 0.015) than it was in the matched univariate model (MZ-R^2 = 0.008), indicating that ADHD accounted for more variance in the unmatched model. This is as expected because in the unmatched sample, controls were dissimilar to the ADHD group on key confounds, so ADHD appeared to explain more variance.

Least controlled		→	Most controlled	
1±	2	3	4-Girls	4-Boys
2.979***	2.185***	2.043**	0.698(+)ns	1.254(-)ns

Table 91. Odds ratio treatment effects of ADHD on EDL from least to most controlled methods

**** p<0.001, ** p<0.01, * p<0.05*
Reference category = level 0: A levels or higher (or vocational equivalent)
±1 – unmatched sample naïve regression
2 - unmatched sample, controlled regression
3 – matched sample
4 – matched sample plus controlled regression, academic only outcome, for girls and boys separately

3.1.1 Most controlled method

The fourth variation shown in Table 91 shows the effect of ADHD in the matched sample, controlled for additional relevant confounds, with girls and boys reported separately. The 0-1 level high/low academic measure of educational attainment was modelled as the outcome, where high education was defined as A-levels or higher, and low as GCSEs A-C (or similar) or below. This model represents the most precise approximation of the ADHD effect taking into account all relevant[64] confounds based on literature, the data available in BCS70, and quasi-experimental matching methods.

In the preferred matched and controlled models, father's education level, maths, and reading ability had small but important effects on educational attainment for both boys and girls. For girls only, mother's education level, authoritarian parenting views, and locus of control had small and important effects.

3.1.1.1 Maths, reading, ADHD, and educational attainment: problematic collinearity?

Age 10 maths and reading scores predicted educational attainment, but ADHD did not, which contradicts findings from other studies of a strong effect of ADHD on educational attainment. I attribute this contradiction to improved balance between ADHD and controls in the matched sample. However, a possible concern here is that maths and reading may be so correlated with either ADHD or educational attainment that it may not be appropriate to control for

[64] Recap: vselect was the Stata procedure used in chapter 6 to select sets of covariates that best predicted ADHD and the EDL outcome, based on several fit indices.

them as confounds. The concern arises in part because psychology studies conventionally use maths and reading scores as measures of educational attainment, particularly in the temporal context of primary or secondary school (Feinstein & Symons, 1997; Gathercole et al., 2003). Also, in this cohort, a large proportion (probably > 50%) had completed their education by age 16, only six years after the age 10 assessment. One might infer there could be partial or complete separation in models containing these perhaps too-similar constructs. Separation occurs when one vector in a logistic regression accurately predicts allocation into another vector, with no overlap (Albert & Anderson, 1984; Webb et al., 2004). For example, if all maths z-scores at zero or below corresponded to the low education group, and all scores above zero corresponded to the high education group.

I used MLR in Mplus to build these models, which is a maximum likelihood procedure robust to moderate collinearity, and if there was partial or complete separation, the procedure should have failed to estimate. However, to address this concern fully and further clarify the nature of the relationships, I analysed additional descriptive and inferential statistics. First, pairwise Pearson's correlations are shown in Table 92. The correlation is highest between maths and reading ($r = 0.733$), and lowest between ADHD severity and adult educational attainment level ($r = -0.090$). The correlations were all significant, but only the relationship between maths and reading was strong enough to indicate a moderate or strong effect (e.g. > 0.50; Cohen, 1988; Ferguson, 2009).

	Maths	**Reading**	**EDL**	**ADHD**
Maths	1.000			
Reading	0.733	1.000		
EDL	0.373	0.391	1.000	
ADHD	-0.178	-0.172	-0.090	1.000

Table 92. Correlation table: maths, reading, EDL (two-level scale), and ADHD (indicator)

All correlations significant at $p < 0.001$

Although the mean maths and reading scores were higher for the higher educational attainment group and the non-ADHD group, there was extensive overlap across the categories, and this is further confirmation that separation is not a problem within the four constructs. I.e. cohort members with low age 10 maths scores and ADHD belonged to the high education group, and others with high maths scores and non-ADHD were in the low education group (Figure 33; Figure 34).

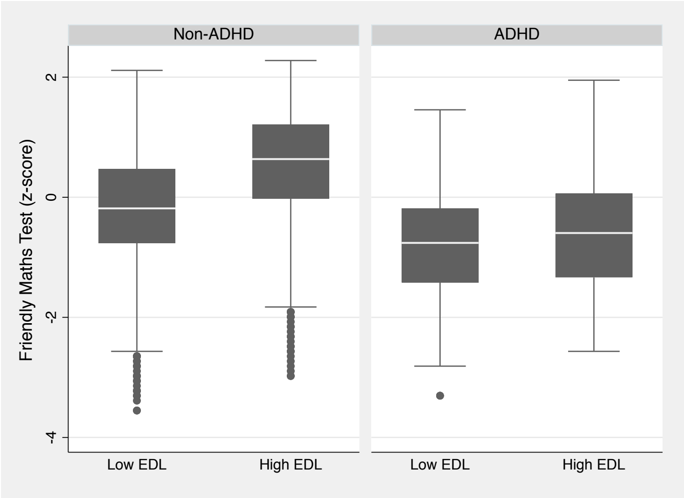


Figure 33. Maths by EDL and ADHD groups

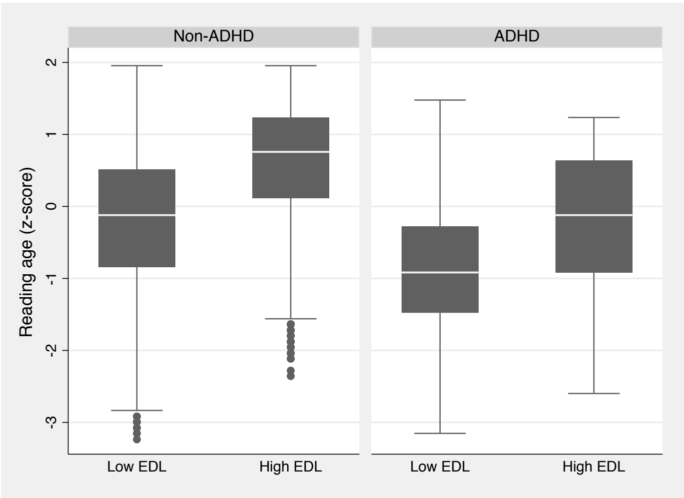


Figure 34. Reading by EDL and ADHD groups

Next, to provide additional assurance, I ran some inferential statistics. First, a simple linear regression of the age 10 maths score predicting the dichotomous (low/high) education outcome (N=7,821, $F(1, 7819) = 1604.29$, $p < 0.001$, $Adj\ R^2 = 0.17$), and the dichotomous ADHD outcome (N=7,821, F(1, 7819) = 255.76, p < 0.001, $Adj\ R^2 = 0.31$). Postestimation variance inflation factors (VIFs) were 2.25 or less, indicating moderate collinearity, and values above 2 were for maths and reading. Finally, standardised residuals were estimated, and no observations indicated excessive collinearity based on a threshold of +/- 2 (Field, 2009).

Based on these additional analyses, I confirmed that including the age 10 maths and reading scores (as proxy measures of IQ), the ADHD indicator, and the adult educational attainment group (as a proxy measure of SES and indicator of objective wellbeing) in the same models was a sound choice and did not result in separation or problematic collinearity.

3.1.1.2 Locus of control

The small and positive effects observed for locus of control in girls were encouraging and provide some support for my hypothesised operationalisation of state regulation theory. The effect observed could translate to opportunities to improve educational attainment by using approaches or interventions targeting increased locus of control in children with certain socio-economic disadvantages (as identified by the matching variables). However, the finding is based on observational data, which cannot prove causality. Also, locus of control is not usually measured in current psychology research, because it has been succeeded by other constructs (e.g. self-efficacy, or growth mindset. Also, there could be omitted variable bias, and/or the variable could be interacting with a correlated variable, such as maths ability, and/or self-esteem. Thus, to add strength to this finding it should be tested using a more current related construct (e.g. self-efficacy, or growth mindset), mediation models, and ideally verified with experimental designs.

3.1.2 Comparisons to other research

To compare the findings here to evidence reported elsewhere, I returned to data on ADHD outcomes from the literature reviewed in chapter 3. For convenience of the reader, the education effects are re-summarised here:

Odds ratio effects of ADHD on a low[65] education outcome from literature review:

- Study 1: the previous study of ADHD in BCS70 reported a significant independent effect of ADHD on educational attainment, measured at age 30 and controlling for a large number of biological, social and economic factors (Brassett-Grundy & Butler, 2004). However, when I converted their reported statistics to effect sizes, I found a small effect for boys (1.41) and non-significant effect (1.20) for girls.
- Study 2: systematic review of 98 studies on ADHD outcomes found a large effect per a pooled odds ratio (6.47; Erskine et al., 2016).
- Study 3: 16-year follow-up of the multimodal treatment study of ADHD (MTA, N=717)) found a medium effect (OR = 2.50; Hechtman et al., 2016)
- Study 4: 33-year follow-up at age 41 (N=271) found a large effect (OR = 7.04; Klein et al., 2012)
- Study 5: age 38 follow-up with the Dunedin cohort (*N*~956) found a medium effect (OR = 3.67; Moffitt et al., 2015)

My findings of effects for ADHD on EDL in the most controlled analysis were not large enough to be practically significant for girls or boys. This is similar to the findings from Study 1 for girls, but differed from the small effect found for boys (the previous study of ADHD in

[65] Low = no tertiary education.

BCS70; Brassett-Grundy & Butler, 2004). My findings do not agree with medium to large effects reported (or calculated post-hoc, by me) in the other four studies.

The most likely explanation for the substantive differences in findings is differences in sampling. Although study 1 used a different measure of ADHD and did not use a matching method to balance the sample, both study 1 and the present study used a non-clinical sample (based on BCS70) and controlled extensively for confounding. Thus, the findings are more similar. Study 2 (Erskine et al., 2016) was a review of other studies using mostly clinical samples, studies 3 – 5 (Hechtman et al., 2016; Klein et al., 2012; Moffitt et al., 2015) used clinically identified samples, and the extent confounding was controlled for varied, but was generally less extensive than the present study or study 1. Clinical samples have the advantage of a professional diagnosis, but they tend to include only the more severe, combined-type ADHD cases (Willcutt, 2012). My ADHD subgroup as identified in chapter 4 was classified as ~46% inattentive subtype and ~14% hyperactive/impulsive subtype. These groups have rarely been studied, so are unlikely to be well-represented in the studies used for comparison.

The positive effect of authoritarian child rearing views on educational attainment for girls was not expected. Authoritarian parenting is characterised by high levels of expectations/demands and low levels of support/warmth, whilst authoritative parenting has high levels of both expectations and support (Maccoby & Martin, 1983). Authoritarian parenting is usually associated with lower educational attainment, and authoritative with higher, but there are overlapping dimensions across the two constructs (Dornbusch et al., 1987; Pinquart, 2016). The measure used here from the age 5 sweep was a binary indicator set to '1' if items from a questionnaire on parenting attitudes were in the 90th percentile or above on an authoritarian scale (Institute of Child Health, 1975). Attitudes do not necessarily equal parenting practices, as was also noted in Flouri (2007), so the indicator could be misleading in that sense. However, those who scored in the 90th percentile or above on this scale were likely to have had a high level of expectations for their children, which is a sub-domain of authoritarian parenting that is associated with higher levels of attainment, compared to children with parents who have low expectations (Pinquart, 2016). Flouri (2007) found that although authoritarian parenting measured at age 5 in BCS70 had a negative influence on educational attainment at age 26 for the whole available sample, the effect disappeared when the sample was restricted to children from a socio-economically disadvantaged background (Flouri, 2007). Her definition of socio-economically disadvantaged included low parental education level and mother depression (malaise). Variables representing both of those factors were included in the matching procedure I used to create the weighted matched sample for my analyses. My matched sample had a significantly lower mean for father's education level, and higher mean for mother malaise than the unmatched sample, so was relatively socio-economically disadvantaged on those factors, like Flouri's. However, whilst Flouri (2007) found that the negative effect of authoritarian parenting was reduced to no effect, I found that it appeared to change direction and have a positive effect for girls, using outcome data from ages 42 and 34. My sample was matched on additional factors of disadvantage, including low standard home, mother smoking during pregnancy, backward development, and wheezing, so was probably more disadvantaged than Flouri's (2007) sample, and was also separated for girls and boys. Both differences could have contributed to the slightly different findings.

3.2 Educational attainment in the ADHD subgroup

The sample size was relatively small for the ADHD subgroup (n=369), particularly for a logistic regression using FIML to handle missing data. The preferred model revealed small and practically significant effects for father's education level and reading ability on the 0-1 level high/low academic EDL, for girls and boys combined. ADHD severity was not a significant predictor of EDL in this group, which does not support findings in literature that severity is a key factor in outcomes (e.g. Costello & Maughan, 2015).

It was interesting that father's education level was still an important differentiator within this group, because (ignoring missing data) the mean level was already significantly lower than for the non-ADHD sample ($M_{ADHD} = 0.76$, $M_{Non\text{-}ADHD} = 1.54$, $t = 9.31$, $p < 0.001$). It was also notable that reading ability at age 10 was significant for this group whilst maths was not, given maths was consistently significant in the other samples. The implication is that educational approaches and/or specific interventions targeting reading ability at age 10 in ADHD children may be beneficial. This finding would need to be validated using other samples and/or experimental designs, and taking into account current socioeconomic and educational contexts.

Chapter 9 Summary and conclusion

1 Recap of thesis rationale

This thesis started with a curiosity about why some children with ADHD have better adult life outcomes than others. It has been estimated that 50% of children with ADHD do not have significant difficulties functioning as adults, and evidence indicates that ADHD severity, comorbidity, and IQ improve chances of better functioning (Costello & Maughan, 2015). However, like Costello & Maughan (2015) pointed out in their review, more knowledge is needed about psychological and social factors that may be open to influence. Also, studies of ADHD using longitudinal cohort data on non-clinical samples are lacking, and needed to mitigate methodological problems common in cross-sectional clinical samples (Caye, Swanson, et al., 2016).

Based on these gaps in knowledge, a broad objective was defined, to evaluate how psychosocial factors for those with ADHD in childhood relate to positively framed outcomes in settled adulthood. The objective was satisfactorily met by the research in the present thesis. Next, the transformation of the objective into research questions is discussed, followed by the answers to the questions from chapters 4, 7, and 8, and implications of the findings for future research. The chapter ends with reflections on learning from the analysis process, and a summary of strengths and limitations.

2 Developing an approach to study long-term ADHD outcomes

In chapter 2 the 1970 British Cohort Study (BCS70) was selected as the data source for this thesis because it has a large, non-clinical UK-population-based sample, data available in childhood that could be mapped to ADHD symptoms, a rich array of psychosocial data, and availability of outcomes data over age 30 (settled adulthood).

Chapter 3 provided background on ADHD, and reviewed the three most widely-discussed causal theories of ADHD: executive dysfunction (Barkley, 1997), dynamic developmental (Sagvolden et al., 2005), and state regulation. State regulation was selected as the basis for my research because it can account for all three sub-types of ADHD, the unusual heterogeneity of case profiles, intra-individual variability (IIV) in performance[66], and has not been refuted

[66] Recap: IIV is one of the only consistent findings throughout ADHD research (Kuntsi & Klein, 2011)

by other evidence. The theory was conceptually operationalized in Figure 3, and proposed that stress and protective factors against stress influence ADHD severity, which in turn influences long-term outcomes.

Chapter 3 next reviewed literature on long-term outcomes for ADHD, and revealed a previous working paper on age 30 outcomes for ADHD in BCS70 (Brassett-Grundy & Butler, 2004). The previous study of ADHD in BCS70 had retrospectively identified ADHD using age 10 behaviour data (Brassett-Grundy & Butler, 2004). However, they used a set of items/symptoms which included several that are not part of the current (DSM-5) definition of ADHD, validation was minimal, they did not derive an ADHD subtype, and did not derive or evaluate a continuous severity measure. The Brassett-Grundy & Butler (2004) study provided a baseline for my work and inspired several aspects of my methodology.

The outcomes review also led to selection of wellbeing and educational attainment as the outcomes of focus for this research. Finally, additional literature was consulted to inform the operationalization of stress and protective factors as: life event stressors, chronic stressors, locus of control, self-esteem, and engagement in leisure.

Building on chapter 3, I defined the first research question in chapter 4:

> RQ1: How can data science methods be used to retrospectively identify and validate robust categorical and continuous measures of DSM-5 ADHD in the BCS70?

RQ1 was answered using a data mining framework and extending methods used elsewhere to retrospectively measure psychological constructs in existing data (e.g. Brassett-Grundy & Butler, 2004; Garcia-Barrera et al., 2011; Goodman et al., 2015; Wall et al., 2015). Three measures were created: a binary ADHD indicator, ADHD subtype, and continuous ADHD severity score (Cotton & Baker, 2018a). All three measures were based on a mapping of DSM-5 ADHD criteria to BCS70 age 10 questionnaire items[67]. The mapping was reviewed by international ADHD experts, and the measures were validated using tests of psychometric properties, other mapped ADHD scales, and categorised prevalence data from epidemiology. The use of a zero-inflated mixture model (ZIMM) with an item response component was an advanced quantitative approach, novel for ADHD data, and a replication of the method reported in Wall et al., (2015). The ZIMM model specifically accommodated the rare event rate of ADHD in the population and the binary nature of the DSM criteria, which are ignored

---

[67] Recap: most of the items on the original BCS70 age 10 questionnaires were derived from Rutter (1967) and Conners (1969) child behaviour scales and were answered by both parents and teachers.

in parametric models (Finkelman et al., 2011; Wall et al., 2015). The model also had the advantage of weighting the importance of items based on probabilities from the data, which provided a more robust measure than a simple sum score.

For the reader's convenience, a summary of the distribution of the age 10 BCS70 data across the new measures is restated here:

				ADHD subgroup	ADHD severity	
Subtypes		% of ADHD	% of Total	% Boys	Mean	SD
Combined	238	40%		73%	2.15	0.24
Hyperactive	82	14%		68%	1.64	0.17
Inattentive	274	46%		678%	1.56	0.17
Total ADHD	594		5.20%	70%	1.81	0.35
Non-ADHD	10,832		94.80%	51%	-0.16	0.82
Total	11,426			51.54%	-0.06	0.91

Table 93. Summary of age 10 data by ADHD measures derived in chapter 4

The resulting sample from age 10 (*n*=594) was large enough to support complex modelling methods and proceed to evaluate relationships with other constructs in childhood and outcomes in adulthood (Cotton & Baker, 2018a). The derivation and use of an ADHD severity measure also supported a general trend in both biological and behavioural research to study ADHD as a continuous construct (e.g. Groen-Blokhuis et al., 2014; Heidbreder, 2015; Swanson et al., 2012).

3 Extending previous knowledge of factors in ADHD outcomes

Equipped with the three new measures of ADHD from chapter 4, methods were piloted in chapter 5 for evaluating long-term ADHD outcomes. Chapter 5 included literature reviews of quasi-experimental methods for estimating causal effects in observational data, matching methods in particular, and of risk factors for ADHD (i.e. potential confounding variables). For the pilot, age 10 data was linked to age 42 data, matching methods used to create an exact-matched sample of ADHD and controls (N=546), and effects of ADHD tested using linear regression on three age 42 outcomes: health and wellbeing, educational attainment, and social class. I found that ADHD had a significant negative effect on all three outcomes based on p-values. The effect size based on adjusted R^2 values was too small to be considered practically important for wellbeing but was small and practically important for educational attainment and social class.

Based on review and feedback, considerable improvements were made to the methods used in the pilot, which were implemented in chapters 6, 7, and 8. Chapter 6 included the definition of refined outcome measures, stress and protective factors, and available relevant covariates. At the end of chapter 6 a complete set of constructs was operationalised which allowed me to articulate the remaining research questions.

> RQ2: How do chronic stressors, life event stressors, locus of control, self-esteem, and engagement in leisure relate to ADHD and ADHD severity, all as measured at age 10? Does the relationship provide evidence to support state regulation theory?
>
> RQ3: What is the effect of childhood ADHD on adult subjective wellbeing using different methods and covariate sets to estimate treatment effects? Do the results support state regulation theory?
>
> RQ4: What is the effect of childhood ADHD on adult educational attainment using different methods and covariate sets to estimate treatment effects? Do the results support state regulation theory?

3.1.1 Chapter 7 provided the answers to research questions 2 and 3

3.1.1.1 RQ2: stress, protective factors, and ADHD

RQ2 was answered by evaluating the relationship between the ADHD measures developed in chapter 4 (Cotton & Baker, 2018a), and the measures of stressors and protective factors developed in chapter 6 using multivariate logistic and linear regression. Chronic stressors and self-esteem had small but practically significant effects on ADHD group membership. Chronic stressors, life event stressors, self-esteem, and locus of control were significant ($p<0.001$) in a linear model predicting ADHD severity, with a small and practically significant effect size. All the variables were measured at the same time (age 10), so the direction of causality cannot be inferred. For example, some of the chronic stressors (e.g. being bullied) could be caused by ADHD symptomatology, or part of a feedback loop. Several of the stressors were based on parent recall of events or problems experienced by the cohort member between ages 5 and 10, so also could have preceded the appearance of ADHD symptoms, but this could not be ascertained from the data.

These results provide some evidence supporting my hypothesised operationalisation of state regulation theory at the conceptual level (Figure 3). Analyses showed that higher stress related to increased ADHD severity and likelihood of membership in the ADHD subgroup, whilst higher protective factors against stress related to reduced ADHD severity and likelihood of membership in the ADHD subgroup.

Even without a causal direction, it is important to know there is a relationship between these constructs, and that this knowledge is taken into consideration in practice. If this is not well-understood, ADHD interventions and management practices could unintentionally have the undesired effect of increasing stress. For example, medication for ADHD may act as a protective factor against stress in the short term, but then become a source of additional stress (via side-effects) in the longer-term. This is something that could be evaluated by practitioners and used to adjust treatment plans. Also, if the concept of reducing stress and increasing locus of control and self-esteem is understood to be fundamental, this knowledge can guide the features of new educational interventions or approaches for ADHD. Therefore, this finding should be taken forward and replications and/or refinements attempted with other samples and other study designs. If the finding holds, it should be communicated widely, including through teacher and ADHD practitioner training.

3.1.1.2 RQ 3: ADHD, stress, and wellbeing

To answer RQ 3, a sequence of models and methods was tested that progressed from least to most restrictive in terms of controlling for confounding factors. The negative effect of ADHD was largest and most significant in the least restrictive, univariate regression on the unmatched sample. However even in this model the effect size of ADHD on subjective wellbeing was too small to be practically important. The statistical significance of the ADHD effect reduced and then disappeared as the models were more controlled for confounding.

The most controlled model used the matched sample[68] and controlled for stressors, protective factors, and other key confounds. Here none of the hypothesised predictors measured in childhood (ADHD, severity, stressors, or protective factors) were practically significant for wellbeing. The only small and practically significant factors were maths, externalising problems and engagement in leisure, the latter two for girls only. The findings differed from reports elsewhere of small-medium effects for childhood ADHD on adult depression (e.g.

[68] Recap: the matched sample included controls balanced with the ADHD group on sex, father's education level, mother malaise/depression, low standard home, mother's smoking habits during pregnancy, wheezing indicator, and backward development indicator. It had the effect of producing a relatively socially disadvantaged sample, with 53% boys.

Erskine et al., 2016; Hechtman et al., 2016). Wellbeing and depression are not identical outcomes, but are related, and have been used as proxy measures for each other in BCS70 (e.g. Schoon & Kneale, 2013). The key driver for differences is again likely to be the different composition of the non-clinical sample. The findings do not provide strong support for my operationalization of state regulation theory, but provide some, via the positive effect of engagement in leisure on wellbeing for girls. This finding is limited by weaknesses in the measurement of engagement in leisure. Maths ability could also be interpreted as a protective factor against stress, but it was not presented in that context here, so is only noted and will be investigated further in future analyses.

3.1.1.3 Wellbeing specifically within the ADHD subgroup

Within the ADHD subgroup (n = 369), ADHD severity in childhood did not have a significant effect on adult wellbeing in a univariate regression. This finding does not support my hypothesised operationalization of state regulation theory, and differs from reports elsewhere that severity is a key predictor of mental health outcomes for ADHD (Caye, Spadini, et al., 2016; Cherkasova et al., 2013; Costello & Maughan, 2015; Lara et al., 2009; Molina et al., 2009). The difference could be attributed to the higher proportion of girls and inattentive subtype represented in my subgroup, the age of outcome measurement, and/or differences between wellbeing and depression as outcomes.

In a multivariate model on the ADHD subgroup sample with girls and boys combined, chronic stressors, externalising problems and reading skills had small and practically significant effect on adult wellbeing. This finding supports other reports that IQ and externalising problems (comparable to comorbid conduct problems) are associated with adult outcomes in ADHD (Caye, Spadini, et al., 2016; Costello & Maughan, 2015).

The relationship in the ADHD subgroup between higher chronic stressors in childhood and lower wellbeing in adulthood provides some support for my hypothesised operationalisation of state regulation theory, although the effect did not appear to operate through ADHD severity. It was notable that reading but not maths ability measures were significant predictors of higher wellbeing within this group, where the reverse was true (maths, not reading) in other samples. The reading effect held up in a univariate regression. Therefore, addressing chronic stressors and cultivating reading ability at age 10 could be particularly important for the long-term wellbeing of children with ADHD symptomatology, and warrants further investigation.

3.1.2 Chapter 8 provided the answer to RQ4

RQ 4: ADHD, stress, and educational attainment

To answer RQ4, relationships were investigated between childhood ADHD, stressors, protective factors, and potential confounds in the matched and ADHD subgroup samples as they related to adult educational attainment. Different categorical measures were tested for educational attainment, but the preferred models used a binary high/low[69] academic EDL outcome.

Again, I tested a sequence of models and methods that progressed from least to most restrictive in terms of controlling for confounding. As with subjective wellbeing, the negative effect of ADHD on educational attainment was largest and most significant in the least restrictive, univariate regression on the unmatched sample. According to the most restrictive multivariate model on the matched sample controlling for effects of stressors, protective factors, and other confounders, the effect of childhood ADHD on educational attainment in settled adulthood was not practically significant for girls or boys.

Although ADHD was not a significant predictor of educational attainment, father's education level, and maths and reading scores from age 10 had small but practically significant effects for both girls and boys. These factors are generally widely accepted to be correlated to educational attainment. There were no further notable findings for boys.

Additionally, for girls, higher locus of control, mother's education level, and authoritarian child rearing views of parent were significantly associated with higher attainment, with small appreciable effect sizes. The positive effect of locus on control on attainment for girls agrees with prior evidence from the BCS70 discussed in chapter 8 (Flouri, 2006; Joshi et al., 2016), although those studies reported effects for girls and boys combined. This finding supports the hypothesised positive relationship between protective factors against stress (locus of control) and outcomes. The effect of mother's education level for girls but not boys implies a specific influence of the female parent on the female child. As concluded in the chapter 7 discussion, the positive effect of authoritarian child-rearing views was probably acting via a sub-dimension of high expectations, specifically within a socio-economically disadvantaged (matched) sample.

---

[69] Recap: A-levels and above = high, GCSE A-C and below = low

3.1.2.1 Educational attainment specifically within the ADHD subgroup

The preferred model on the ADHD subgroup sample (n = 369) indicated that father's education level and reading score at age 10 were significant and important with small effect sizes. It was again noteworthy that reading, but not maths, appeared to be the key proxy measure of IQ in this sample, which could be useful for the design of targeted ADHD educational approaches. The findings cannot be interpreted as causal, however, because they are from a single study on observational data, so would need to be confirmed with other samples and experimental designs.

4 Implications of findings for future research and practice

A significant contribution of my work is a well-documented and robust data mining method to retrospectively derive three measures of ADHD (indicator, subtype, severity score) in BCS70, and the measures themselves, for use in future research. The measures of wellbeing, life event and chronic stressors, and engagement in leisure developed in chapter 6 could also be useful to others. There is currently a large ongoing cross-organisation (ESRC, MRC, CLS, and others) research initiative to harmonise mental health measures in the British birth cohorts and maximise use of the data (CLOSER, 2019), and the work in this thesis may be able to contribute. Importantly, the data mining method can be applied in datasets internationally to retrospectively identify ADHD and/or other psychosocial constructs, where semantically similar data items are available. Global relevance was evidenced by publication of the method in an international journal (Cotton & Baker, 2018b).

Second, the finding of a relationship between increased chronic and life event stressors, decreased locus of control and self-esteem and ADHD severity provides further evidential support at the conceptual level for a state regulation theory of ADHD. It is important that approaches to engaging and educating ADHD children are developed based on the best causal explanation of severity and impairment. If this finding is confirmed in other samples, it should be incorporated into training of teachers and other practitioners who work with ADHD children.

Third, the differences in effects observed between controlled regression and use of a matched sample suggested that when control groups are selected for studies of ADHD children, they should be matched on specific factors related to socio-economic status. Many previous studies matched only on sex and a broad SES indicator. Important factors indicated in my research included father's education level, mother's malaise/depression, low standard home, and mother's smoking during pregnancy. These factors could also be indicators of a genetic pre-disposition for ADHD, i.e. of ADHD in the parents. Genetic data was not available to confirm in the BCS70, but it would be interesting to test this idea in another data set that has genetic data, such as ALSPAC[70].

Fourth, chronic stressors, locus of control, and reading ability at age 10 may be important factors in long-term wellbeing and educational attainment specifically for children with ADHD. Effects should be tested in other samples and considered in experimental designs of prospective studies.

Fifth, groundwork has been laid here for future work testing indirect relationships and using more complex structural/path models (see section 5.1), as well as testing for ADHD effects in other cohorts, such as the Millennium Cohort Study (MCS) or the National Child Development Study (NCDS). I will endeavour to coordinate such further work with the ongoing cross-organisation project (mentioned above) to harmonise mental health data across British birth cohort datasets.

4.1 International and intergenerational generalisability

The sample used for this thesis was a cohort born in 1970 in Great Britain. Key childhood measures were derived using data items collected between 1970 and 1980, whilst key adult measures were collected in 2004 and 2012. Intergenerational changes are unavoidable in studies of relationships between childhood factors and adult outcomes, and my findings are most directly comparable to other studies that span a similar timeframe, in a similar cultural context. Outside of a similar timeframe and context, demography and medicalisation of ADHD are important factors in generalisability.

4.1.1 Demography

The models selected for reporting key findings were based on the preferred (weighted) matched sample (N=6,207). Ethnicity for that sample was 96% English, etc (probably white), and 1.4% Indian. 98% were single births, 52.9% were male, 20.5% of homes had a primary earner with a partly skilled or unskilled job, and 28% with a managerial/technical or professional job. My findings should broadly generalise to other samples with similar attributes.

For the ADHD subgroup specifically, findings should broadly generalise to other non-clinical, undiagnosed, unmedicated samples identified using measures based on DSM-5 or DSM-IV criteria (they are very similar; American Psychiatric Association, 2000; Substance Abuse and

---

[70] Avon Longitudinal Study of Parents and Children

Mental Health Services Administration, 2016) with a similar composition by sex (70% boys) and subtype (46% inattentive, 14% hyperactive, and 40% combined).

4.1.2 Medicalisation of ADHD

Diagnosis and treatment of ADHD in the UK in 1980 (the year of the source data for my retrospective ADHD assessment) was rare, in fact virtually none of the children in my sample had data indicating a diagnosis related to ADHD, at least at the age of 10[71]. This an important and unusual attribute of the BCS70 data (for a discussion see chapter 4, section 1.1.1). Changes in expectations, perceptions, and behaviours of the child and others (doctors, parents, teachers, peers), reduction of symptoms, and experience of side-effects are possible sources of differences between undiagnosed/untreated and diagnosed/treated children. All of these factors could interfere with trajectories, outcomes, and comparisons between samples.

Diagnosis and treatment of ADHD in the UK has remained low over time, in a global context. The recently reported rates of 1.6% for diagnosis and about 0.7% for medication (NHS Digital, 2018) are much lower than the 11% diagnosis and 9% treatment rates reported for schoolchildren in the United States (Bergey et al., 2018). A recent qualitative review of practices across sixteen countries reported that key influences on ADHD medical practices were cultural attitudes towards medicalising child behaviour, the societal importance of child academic performance, and health care system structures and policies (Bergey et al., 2018). The countries in the review included the United States, United Kingdom, Ireland, Canada, Australia, New Zealand, Chile, Brazil, Argentina, France, Germany, Italy, Portugal, Taiwan, Japan, and Ghana (Bergey et al., 2018). The countries were categorised into two qualitative groups: those that tend to medicate as first line treatment, and those who minimise medication (Conrad & Singh, 2018). The first group included the United States, Canada, Australia, and Germany, and the second group included the UK, France, Brazil, and Japan (Conrad & Singh, 2018). Based on this review, my findings should generalise better to countries that minimise medication, i.e. in the second group.

5 Post hoc reflections on conceptual model and methods

Learning throughout the process of developing this thesis has been considerable, and there is not space to recount it all. However, I have summarised a few items of note I plan to take into consideration for subsequent analyses and preparation of manuscripts for publication.

---

[71] Only five children in the matched sample (N=6,207) had an ICD-9 code indicating a Hyperkinetic Disorder diagnoses (314-), which is less than 0.1%. All five of the children with an HKD/314 diagnosis also had at least one other ICD-9 diagnosis.

5.1 Hypothesised indirect effects

In chapter 3 (Figure 3) I hypothesised a conceptual model which portrayed stress and protective factors as having indirect effects on outcomes through ADHD severity and impairment. The models I actually tested in chapters 7 and 8 were more like Figure 35, below. Maximum likelihood procedures used were robust to correlations between the predictor variables, but indirect relationships were not modelled explicitly.

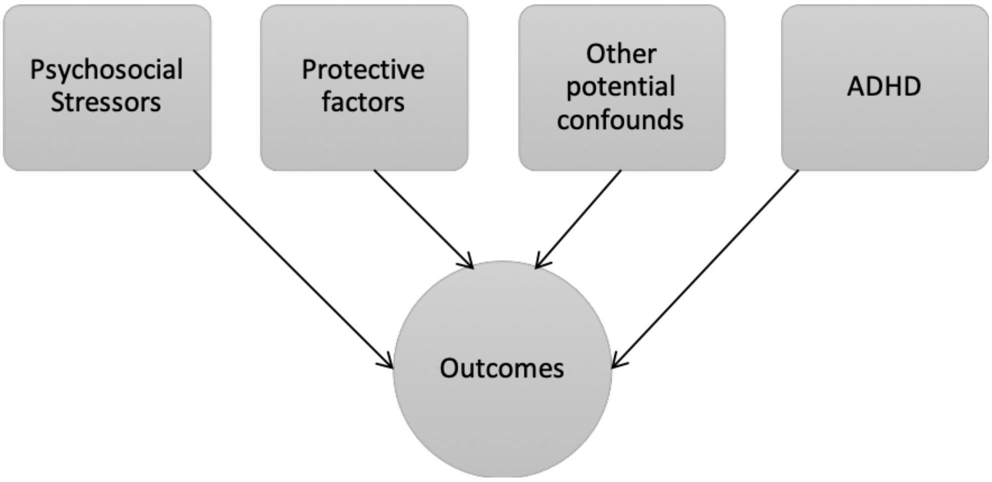


Figure 35. Conceptual model of direct effects actually tested in chapters 7 and 8

The process of developing and validating ADHD measures, analysing missingness and bias, implementing procedures robust to missingness and non-normality, piloting and implementing the matching procedure, and building multivariate models of direct relationships took more space in the thesis than expected. Thus, mediation and moderation models will be developed in my next phase of work. A draft of a revised conceptual model based on my learning from the direct relationships tested in my analyses and further reflection is proposed in Figure 36. It includes a moderator relationship, or interaction, between stressors and protective factors, as well as a mediator relationship between responses to stress (ADHD, internalising, and externalising) and outcomes, via a 'reflect, plan, and adapt' process, which should relate to executive functions (EF). The model also shows points where education could be expected to have an influence. At the point of defining covariates for the thesis (chapter 6) I did consider whether executive functions could be measured in BCS70, but did not find viable items available in the age 5 or 10 sweeps. Since I am proposing to test EF as a mediator, i.e. part of the causal chain, I will revisit this in future work, and will review age 16 and 26 data for candidate EF item mappings. Inclusion of data from these sweeps, if identified, will require further analysis of missing data, because both had specific and substantive missingness.

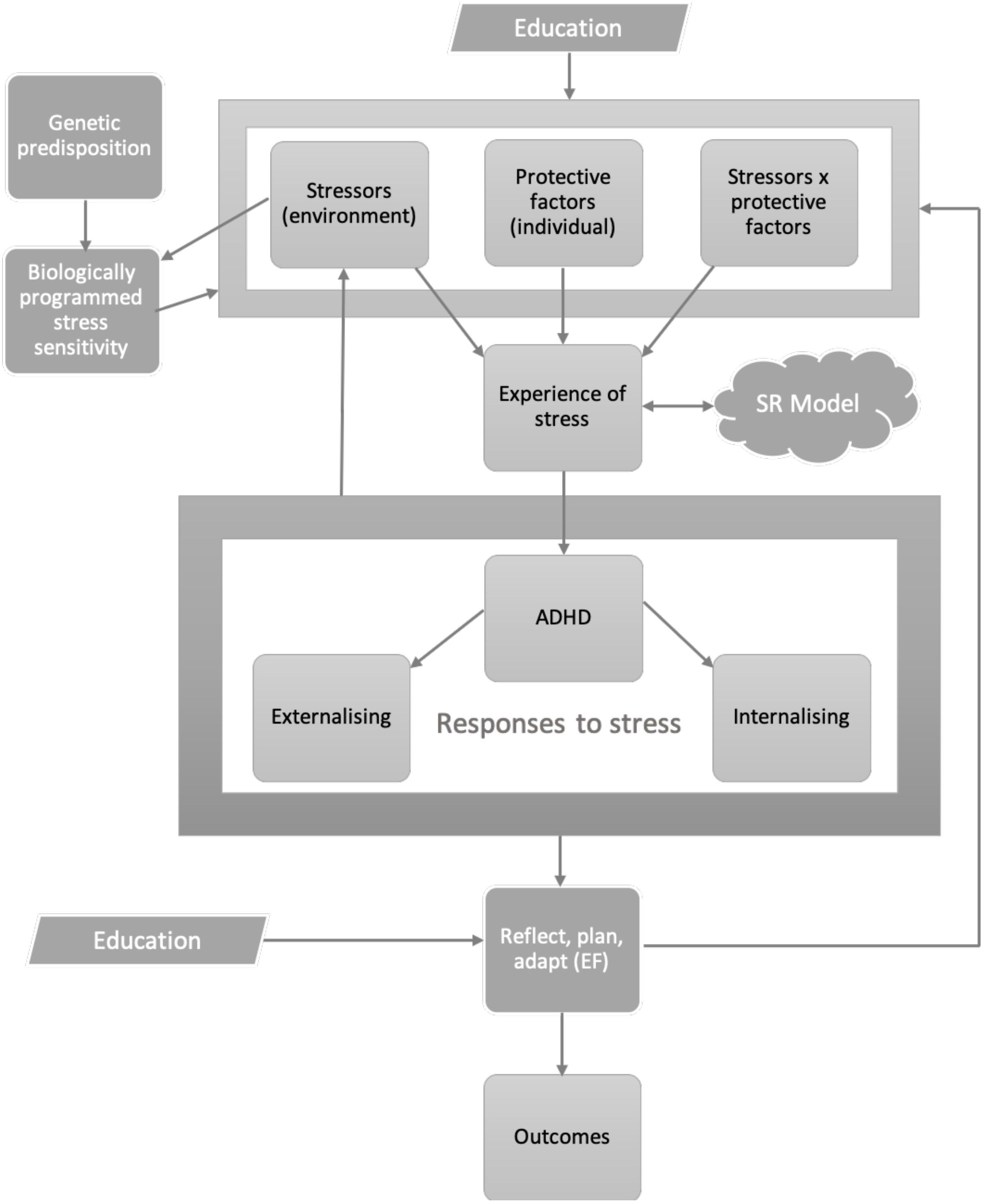


Figure 36. Revised more complex conceptual model, proposed for use in future work

5.2 Measuring ADHD

Through analysing the effects of ADHD and severity on both outcomes, I encountered anomalies caused by the different approaches used in chapter 4 to derive the two measures. To recap, the subgroup indicator was based on application of approximated DSM-5 criteria:

2/3rds of either or both subtype symptom lists, plus two necessary conditions[72]. The model-based calculation of the severity score did not afford special status for those two conditions. Also, in the application of the DSM-5 criteria to derive the indicator, all items were weighted equally, whilst in the severity model, items were weighted based on relative probabilities within the sample. The net result was that some members of the ADHD subgroup had relatively low severity scores, because the symptoms that met the criteria had low weights in the mixture model. Conversely, some cohort members had high severity scores, but were not in the ADHD subgroup because the necessary conditions were not met. I think both measures were still sound, but ideally the relationship between the two should be linear. This inconsistency would be present in any comparison of a DSM-based diagnosis to a statistically calculated severity score, has been noted elsewhere as problematic, and accordingly a large initiative is underway to reclassify disorders so categorical and continuous indicators are consistent with each other (Kotov et al., 2017). In future work with this data, I intend to consult again with ADHD experts on dropping the two necessary conditions from the ADHD subgroup derivation.

5.3 Measuring confounds

Wheezing, hayfever, and eczema indicators were evaluated as separate potential confounding variables in chapters 7 and 8, because of evidence from a large (N = 21,756) study conducted in Taiwan showing that atopic diseases were significant risk factors for ADHD (Chen et al., 2014). Under my hypothesis operationalising a state regulation theory of ADHD, however, any type of medical problem that increases stress should relate to increased severity of ADHD. Since atopic diseases are ongoing medical problems similar to other medical problems I counted as chronic stressors, I will test collapsing them into the chronic stressors count in future work with this data instead of controlling for them as separate confounds.

5.4 Implementing matching methods

I found that variable selection was the most difficult aspect of implementing matching methods. I used both an analyst-led iterative process (in the pilot, chapter 5) and automated (vselect) procedure (chapter 6) with guidance from the literature to select the matching variables. The automated procedure produced a set of items that included some problematic collinearity. On reflection, the best compromise is probably to use an automated process, then make adjustments if warranted by knowledge of the constructs.

---

[72] 1) Parent indicated moderate to severe problems, and 2) at least 3 items endorsed by both teacher and parent.

5.5 Challenge of balancing/bridging terminology, methods, and conventions across disciplines

Studies that were influential on my thesis were drawn from literature across the sub-fields of education economics, longitudinal life-course studies, psychology, and neuroscience. The use of theory, terminology, methods, and conventional practices across these disciplines varies significantly. For example, analysts in education economics and longitudinal life-course studies tend to use Stata and MPlus, whist those in psychology and neuroscience tend to use SPSS and/or R. The types of procedures available and terminology used varies somewhat across the software packages. Also there is a tendency in educational economics and longitudinal life-course studies to examine models including numerous and sometimes broad indicators or proxies for constructs (e.g. intelligence, or using adult educational attainment as a proxy measure of SES), whilst in psychology and neuroscience the tendency is to use only rigorously validated measures and examine fewer, more precisely defined constructs (e.g. working memory). Thus, it can be difficult to conduct analysis and write for audiences across these disciplines. For example, in chapter 6 when I selected a measure of IQ, I chose a maths test score as a simple proxy instead of computing a subscale or composite score from the BAS items. From an education economics (and scope minimisation) point of view this was a reasonable choice, but from a psychology point of view perhaps not. On reflection, and after re-working the section describing the measures of IQ, I expect that if I submit a paper based on this work to a psychology journal, I may need to complete some additional work to calculate/use the available BAS scores.

6 Strengths and limitations

6.1 Strengths

First, this thesis presented a collection of literature reviews and analyses on the topic of long-term outcomes in ADHD. The ADHD literature (as discussed in section 5.5) is cross-disciplinary. I consulted with experts and attended events across disciplines, which was facilitated by my membership in the Cambridge Neuroscience network, and role as a policy fellow for the ESRC Educated Brain Seminar Series. Thus, a broad-level strength is that the research questions and methods used to answer them were informed by cross-disciplinary engagement, which is a key element in producing research that can improve practice.

Second, the use of secondary longitudinal birth cohort data from the BCS70 made use of existing research resources and allowed me to study the impact of childhood factors on long-term (settled adulthood) outcomes, which would not otherwise have been possible within the resource constraints of a PhD. BCS70 also provided a large sample, rich in observed variables that could exploit the use of robust quantitative methods and minimise sampling biases common to small, clinical, cross-sectional studies (e.g. omitted variable or diagnostic biases).

A third strength was the rigour in the approach to measuring ADHD. The mapping of questionnaire items was reviewed by a panel of international ADHD experts, and the derived measures validated against other mapped measures and epidemiological data. The method for measuring ADHD severity was piloted first, re-designed based feedback and further study of methods literature, then published in a peer-reviewed journal article.

Fourth, some of the statistical methods used in this thesis were non-standard and were selected to minimise bias and maximise precision of estimates for the specific characteristics of the data, based on current methods literature and simulation studies. Most of the methods were implemented in Mplus. Non-standard methods included the zero-inflated mixture model (ZIMM) used to estimate ADHD severity, coarsened exact matching (CEM) to create the matched/balanced sample, and the maximum likelihood robust (MLR) estimator with full information maximum likelihood (FIML) to estimate regression models with violations of multivariate normality and complex missingness. The use of these methods, as well as comparisons using multiple methods (e.g. robustness checks using BAYES estimators), strengthened findings.

Fifth, a pilot approach (chapter 5) was used to develop the design for the study of outcomes. The pilot work was presented in a poster and reviewed by my supervisors, and feedback informed changes to methods which addressed bias from missing data, increased robustness of estimates, and facilitated comparisons with other ADHD outcomes research[73].

In contrast to the exploratory approach taken testing many outcomes in other studies, only two outcomes were tested. Wellbeing and educational attainment were not highly correlated with each other, covered significant life domains, were person rather than event-centred, and the targeted focus reduced the risk of false positive findings from multiple comparisons.

A final strength was the successful refinement of methods used in a previous working paper on ADHD (Brassett-Grundy & Butler, 2004). As planned in chapter 3, the categorical measure of ADHD was more closely aligned with the current definition of ADHD per the DSM-5, a scaled score was calculated, and subtypes derived. Second, more robust methods were used to account for missing data. Finally, controlled confounds were linked more directly to

[73] E.g. separate results reported for girls and boys.

literature, and, quasi-experimental methods were used to improve balance on key covariates between ADHD and controls.

6.2 Limitations

There are limitations related to using observational data, and the BCS70 cohort in particular. Observational data does not have the characteristics of a randomised controlled experiment that allow for stronger inferences about causal effects. Also, the approach requires iterative development of research questions based on literature and data available, i.e. the researcher does not have creative freedom to design the data collection based on the research questions.

New measures had to be created using the BCS70 data because of changes over time to psychosocial constructs of interest. These included measures for ADHD, wellbeing, stressors, and engagement in leisure. It is a strength they were developed in a large sample, but since they were not validated in other samples, inferences about generalisability are limited.

Since the BCS70 was designed decades ago, the approach did not over-sample subsets of the population now known to attrit disproportionately over time, like more recent studies have done (e.g. the Millennium Cohort Study/MCS). Also, there have been considerable advances in knowledge about data collection and quality management practices since this study began, so unsurprisingly some of the data was difficult to use (i.e. required considerable cleaning and recoding) because of quality problems.

As discussed in section 5.1, the hypothesised conceptual model for my research Figure 3 depicted indirect relationships between stressors, protective factors, ADHD, and outcomes. The fact that I did not ultimately model these indirect (mediator/moderator) relationships is a limitation of this thesis. Addressing this limitation will be prioritised in my next phase of work, to shed further light on mechanisms at work in a complex system of constructs that consistently demonstrates equifinality.

The approach of testing a series of regression models for each outcome could have been affected by analyst bias, and explicit adjustments were not made for multiple comparisons.

Lastly, some methods used were non-mainstream, and it can be a disadvantage when methods are less widely known and understood, because they are less accessible to wide audiences. Since such methods require specialist knowledge, they may be less likely to be replicated.

References

Abadie, & Imbens. (2006). Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica*, *74*(1), 235–267.

Abadie, & Imbens. (2016). Matching on the Estimated Propensity Score. *Econometrica*, *84*(2), 781–807. https://doi.org/10.3982/ECTA11293

Achenbach, & Edelbroch. (1983). *Child Behavior Checklist*. University Associates in Psychiatry.

Achenbach, & Ruffle. (2000). The child behavior checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in Review*, *21*(8), 265–271. https://doi.org/10.1542/pir.21-8-265

ADDISS. (2006). *A report based on research amongst children and parents illustrating the impact of ADHD on the school day*.

Agnew-Blais, Polanczyk, Danese, Wertz, Moffitt, & Arseneault. (2016). Evaluation of the Persistence, Remission, and Emergence of Attention-Deficit/Hyperactivity Disorder in Young Adulthood. *JAMA Psychiatry*, *73*(7), 1–8. https://doi.org/10.1001/jamapsychiatry.2016.0465

Albert, & Anderson. (1984). On the existence of maximum likelihood estimates for weighted logistic regression. *Biometrika*, *71*(1), 1–10.

Allison. (2008). *Missing Data*. www.ssc.upenn.edu/~allison

Allison. (2018). *Structural Equation Modeling*. Statistical Horizons.

Alloway, & Passolunghi. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences*, *21*(1), 133–137. https://doi.org/10.1016/j.lindif.2010.09.013

Allsopp, Read, Corcoran, & Kinderman. (2019). Heterogeneity in psychiatric diagnostic classification. *Psychiatry Research*, *279*(July), 15–22. https://doi.org/10.1016/j.psychres.2019.07.005

Alwin, & Hauser. (1975). The decomposition of effects in path analysis. *American Sociological Review*, *40*(1), 37–47.

American Academy of Pediatrics, McNeil, & Wolraich. (2002). *NICHQ Vanderbilt Assessment Scales*. NICHQ. http://www.nichq.org/sites/default/files/resource-file/NICHQ_Vanderbilt_Assessment_Scales_0.pdf

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders (DSM-III-R)* (5th ed.). American Psychiatric Association.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.).

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

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

American Psychiatric Association. (2016). *History of the DSM*. Psychiatrists Practice. https://www.psychiatry.org/psychiatrists/practice/dsm/history-of-the-dsm

Andres. (2014). *Quantitatively measuring correlations*. Fathom Information Design Notebook. https://fathom.info/notebook/7028/

Angrist, & Pischke. (2010). Economics : How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, *24*(2), 3–30.

Antshel, Faraone, Stallone, Nave, Kaufmann, Doyle, Fried, Seidman, & Biederman. (2007). Is attention deficit hyperactivity disorder a valid diagnosis in the presence of high IQ? Results from the MGH Longitudinal Family Studies of ADHD. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *48*(7), 687–694. https://doi.org/10.1111/j.1469-7610.2007.01735.x

APA. (2020). *APA Dictionary of Psychology*. https://dictionary.apa.org

appg: wellbeing economics. (2019). *All party parliamentary group on wellbeing economics*. Website. https://wellbeingeconomics.co.uk

Araya, Hu, Heron, Enoch, Evans, Lewis, Nutt, & Goldman. (2008). Effects of stressful life events, maternal depression and 5-HTTLPR genotype on emotional symptoms in pre-

adolescent children. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *150*(5), 670–682. https://doi.org/10.1002/ajmg.b.30888
Arias, Esnaola, & Rodríguez-Medina. (2018). Identifying potentially marker symptoms of attention-deficit/hyperactivity disorder. *PeerJ*, *6*, e4820. https://doi.org/10.7717/peerj.4820
Arnett. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist*, *55*(5), 469–480. https://doi.org/10.1037/0003-066X.55.5.469
Astle, Bathelt, & Holmes. (2019). Remapping the cognitive and neural profiles of children who struggle at school. *Developmental Science*, *22*(1), 1–17. https://doi.org/10.1111/desc.12747
Austin. (2010). Statistical criteria for selecting the optimal number of untreated subjects matched to each treated subject when using many-to-one matching on the propensity score. *American Journal of Epidemiology*, *172*(9), 1092–1097. https://doi.org/10.1093/aje/kwq224
Austin. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, *10*(2), 150–161. https://doi.org/10.1002/pst.433
Baker. (2001). *The Basics of Item Response Theory* (2nd ed.). ERIC Clearing House on Assessment and Evaluation.
Baraldi, & Enders. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, *48*(1), 5–37. https://doi.org/10.1016/j.jsp.2009.10.001
Barkley. (1997). Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychological Bulletin*, *121*(1), 65–94. http://www.ncbi.nlm.nih.gov/pubmed/9000892
Barkley. (2015). *Attention-Deficit Hyperactivity Disorder, Fourth Edition: A Handbook for Diagnosis and Treatment* (Barkley (ed.); Fourth). The Guildford Press.
Baron, & Kenny. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. http://www.ncbi.nlm.nih.gov/pubmed/3806354
Bass, Dawkin, Muncer, Vigurs, & Bostock. (2016). Validation of Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in a population of people using Secondary Care Mental Health Services. *Journal of Mental Health*, *25*(4), 323–329. https://doi.org/10.3109/09638237.2015.1124401
Benjamini, & Hochberg. (1995). Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, *57*(1), 289–300.
BERA. (2011). *Ethical Guidelines for Educational Research*. BERA.
Berden, Althaus, & Verhulst. (1990). Major Life Events and Changes in the Behavioural Functioning of Children. *Journal of Child Psychology and Psychiatry*, *31*(6), 949–959. https://doi.org/10.1111/j.1469-7610.1990.tb00836.x
Bergey, Filipe, Conrad, & Singh (Eds.). (2018). *Global Perspectives on ADHD: Social Dimensions of Diagnosis and Treatment in Sixteen Countries*. Johns Hopkins University Press.
Birnbaum. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord & Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–472). Addison-Wesley.
Blackwell, Iacus, King, & Porro. (2009). cem : Coarsened exact matching in Stata. *The Stata Journal*, *4*, 524–546.
Blanchflower, & Oswald. (2008). Is well-being U-shaped over the life cycle? *Social Science and Medicine*, *66*(8), 1733–1749. https://doi.org/10.1016/j.socscimed.2008.01.030
Bo, Jun, & Palta. (2006). PSEUDO-R 2 in logistic regression model. *Statistica Sinica*, *16*(3), 847–860.
Bollen. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*, 605–634.
Boutwell, & Beaver. (2010). Journal of Criminal Justice The role of broken homes in the development of self-control: A propensity score matching approach. *Journal of Criminal Justice*, *38*(4), 489–495. https://doi.org/10.1016/j.jcrimjus.2010.04.018

Brady. (2014). Children and ADHD: Seeking control within the constraints of diagnosis. *Children and Society*, *28*(3), 218–230. https://doi.org/10.1111/chso.12069

Brassett-Grundy, & Butler. (2004). Prevalence and adult outcomes of Attention-Deficit / Hyperactivity Disorder. In *Bedford Group for Lifecourse & Statistical Studies* (No. 2; BG Occasional Paper). http://cls.ioe.ac.uk/library-media%5Cdocuments%5CBG_OccPaper2_ADHD_outcomes_2008.pdf

Brassett-Harknett, & Butler. (2007). Attention-deficit/hyperactivity disorder: An overview of the etiology and a review of the literature relating to the correlates and lifecourse outcomes for men and women. *Clinical Psychology Review*, *27*(2), 188–210. https://doi.org/10.1016/j.cpr.2005.06.001

Broadbent. (1958). *Perception and Communiation*. Pergamon Press.

Brookhart, Schneeweiss, Rothman, Glynn, Avorn, & Stürmer. (2006). Variable selection for propensity score models. *American Journal of Epidemiology*, *163*(12), 1149–1156. https://doi.org/10.1093/aje/kwj149

Brown, G. W., & Harris. (1978). *Social Origins of Depression*. Tavistock.

Brown, G. W., Harris, & Peto. (1973). Life events and psychiatric disorders. 2. Nature of causal link. *Psychological Medicine*, *3*(2), 159–176.

Brown, G. W., Sklair, Harris, & Birley. (1973). Life-events and psychiatric disorders Part 1: some methodological issues. *Psychological Medicine*, *3*, 74–87. https://doi.org/10.1017/S0033291700046365

Brown, M., & Hancock. (2014). *1970 British Cohort Study 2012 follow-up: a guide to the datasets* (Issue February). http://doc.ukdataservice.ac.uk/doc/7473/mrdoc/pdf/bcs70_2012_follow_up_guide_to_the_datasets.pdf

Brown, T. A. (2006). Confirmatory Factor Analysis for Applied Research. In David A Kenny (Ed.), *Methodology in the Social Sciences*. The Guildford Press. https://doi.org/10.1007/s11336-016-9548-3

Brown, T. E. (2013). *A New Understanding of ADHD in Children and Adults: Executive Function Impairment*. Routledge.

Budding, & Chidekel. (2012). ADHD and giftedness: A neurocognitive consideration of twice exceptionality. *Applied Neuropsychology: Child*, *1*(2), 145–151. https://doi.org/10.1080/21622965.2012.699423

Bujang, Sa'At, Tg Abu Bakar Sidik, & Lim. (2018). Sample size guidelines for logistic regression from observational studies with large population: Emphasis on the accuracy between statistics and parameters based on real life clinical data. *Malaysian Journal of Medical Sciences*, *25*(4), 122–130. https://doi.org/10.21315/mjms2018.25.4.12

Butler, Despotidou, & Shepherd. (1997). *1970 British Cohort Study Ten-year Follow-up: Guide to data available at the ESRC archive*.

Button, Ioannidis, Mokrysz, Nosek, Flint, Robinson, & Munafò. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Bynner, & Joshi. (2007). Building the Evidence Base From Longitudinal Data. *Innovation: The European Journal of Social Science Research*, *20*(2), 159–179. https://doi.org/10.1080/13511610701502255

Cadman, Findon, Eklund, Hayward, Howley, Cheung, Kuntsi, Glaser, Murphy, & Asherson. (2016). Six-year follow-up study of combined type ADHD from childhood to young adulthood: Predictors of functional impairment and comorbid symptoms. *European Psychiatry*, *35*, 47–54. https://doi.org/10.1016/j.eurpsy.2015.08.007

Cairncross, & Miller. (2016). The Effectiveness of Mindfulness-Based Therapies for ADHD. *Journal of Attention Disorders*, 108705471562530. https://doi.org/10.1177/1087054715625301

Caldwell, & Smith. (1988). Leisure: An overlooked component of health promotion. *Canadian Journal of Public Health*, *79*(2).

Caliendo, & Kopeinig. (2005). *Some Practical Guidance for the Implementation of Propensity Score Matching* (No. 1588; Discussion Paper Series).

Cambridge Libraries. (2019). *Education Library: Literature Searching*. LIbGuides. http://libguides.cam.ac.uk/litsearching

Carlton. (2012). *Measuring Socio-economic Status*. 1–22. %5C%5Ccgcna01%5CI_drive%5CSpecial Projects%5CWeb documents%5CPublications%5C2015 Review Report%5CIndigenous%5C2012-03%5CnMeasuring Socio-Economic Status.docx

Castellanos, & Tannock. (2002). Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. *Nature Reviews. Neuroscience*, *3*(8), 617–628. https://doi.org/10.1038/nrn896

Caye, Rocha, Anselmi, Murray, Menezes, Barros, Gonçalves, Wehrmeister, Jensen, Steinhausen, Swanson, Kieling, & Rohde. (2016). Attention-Deficit/Hyperactivity Disorder Trajectories From Childhood to Young Adulthood: Evidence From a Birth Cohort Supporting a Late-Onset Syndrome. *JAMA Psychiatry*, *73*(7). https://doi.org/10.1001/jamapsychiatry.2016.0383

Caye, Spadini, Karam, Grevet, Rovaris, Bau, Rohde, & Kieling. (2016). Predictors of persistence of ADHD into adulthood: a systematic review of the literature and meta-analysis. *European Child and Adolescent Psychiatry*, *25*(11), 1–9. https://doi.org/10.1007/s00787-016-0831-8

Caye, Swanson, Thapar, Sibley, Arseneault, & Hechtman. (2016). Life Span Studies of ADHD: Conceptual Challenges and Predictors of Persistence and Outcome. *Current Psychiatry Reports*, *18*(12). https://doi.org/10.1007/s11920-016-0750-x.Life

Celeux, & Soromenho. (1996). An entropy criterion for assessing the numbers of clusters in a mixture model. *Journal of Classification*, *13*(2), 195–212. https://doi.org/10.1016/S0167-8655(98)00144-5

Centers for Disease Control and Prevention. (2014). *Data and Statistics*. Attention-Deficit / Hyperactivity Disorder (ADHD). http://www.cdc.gov/ncbddd/adhd/data.html

Centers for Disease Control and Prevention. (2018a). *ADHD, Symptoms and Diagnosis*. http://www.cdc.gov/ncbddd/adhd/diagnosis.html

Centers for Disease Control and Prevention. (2018b). *ADHD Throughout the Years*. Attention Deficit / Hyperactivity Disorder (ADHD). https://www.cdc.gov/ncbddd/adhd/timeline.html

Centers for Disease Control and Prevention. (2018c). *DSM-5 Criteria for ADHD*. Attention Deficit / Hyperactivity Disorder (ADHD): Diagnosis. http://www.cdc.gov/ncbddd/adhd/diagnosis.html

Centre for Economic Performance. (2019). Wellbeing research programme. In *Website*. London School of Economics. http://cep.lse.ac.uk/_new/research/wellbeing/default.asp

Centre for Longitudinal Studies: UCL/IoE. (n.d.). *Rutter Behaviour Questions*.

Centre for Longitudinal Studies: UCL/IoE. (2012). *1970 BRITISH COHORT STUDY Age 42 Survey Self-completion Questionnaire*. UK Data Service.

Centre for Longitudinal Studies: UCL/IoE. (2014). *CLS Cohort Studies Code of Practice*.

Centre for Longitudinal Studies: UCL/IoE. (2019). *1970 British Cohort Study*. Home of the 1958, 1970, Millennium and Next Steps Cohort Studies. https://cls.ucl.ac.uk/cls-studies/1970-british-cohort-study/

Chamberlain, Chamberlain, & Institute of Education. (2013). 1970 British Cohort Study: Birth and 22-months Subsample 1970-1972. In *1970 British Cohort Study* (3rd ed., Issue May). UK Data Archive. https://doi.org/http://dx.doi.org/10.5255/UKDA-SN-2666-2

Chen, H., Cohen, & Chen. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics: Simulation and Computation*, *39*(4), 860–864. https://doi.org/10.1080/03610911003650383

Chen, M. H., Su, Chen, Hsu, Huang, Chang, Chen, Pan, & Bai. (2014). Is atopy in early childhood a risk factor for ADHD and ASD? A longitudinal study. *Journal of Psychosomatic Research*, *77*(4), 316–321. https://doi.org/10.1016/j.jpsychores.2014.06.006

Cherkasova, Sulla, Dalena, Ponde, & Hechtman. (2013). Developmental Course of Attention Deficit Hyperactivity Disorder and its Predictors. *Journal of Canadian Academy of Child and Adolescent Psychiatry*, *22*(1), 47–54. https://doi.org/10.2217/ebo.12.273

Cheung, Rijsdijk, McLoughlin, Faraone, Asherson, & Kuntsi. (2015). Childhood predictors of adolescent and young adult outcome in ADHD. *Journal of Psychiatric Research*, *62*, 92–100. https://doi.org/10.1016/j.jpsychires.2015.01.011

Chimiklis, Dahl, Spears, Goss, Fogarty, & Chacko. (2018). Yoga, Mindfulness, and Meditation

Interventions for Youth with ADHD: Systematic Review and Meta-Analysis. *Journal of Child and Family Studies*, *27*(10), 3155–3168. https://doi.org/10.1007/s10826-018-1148-7
Chronis, Jones, & Raggi. (2006). Evidence-based psychosocial treatments for children and adolescents with attention-deficit/hyperactivity disorder. *Clinical Psychology Review*, *26*(4), 486–502. https://doi.org/10.1016/j.cpr.2006.01.002
Chronis, Pelham, Baumann, Kipp, Lahey, Williams, Jones, & Rathouz. (2007). Maternal depression and early positive parenting predict future conduct problems in young children with attention-deficit/hyperactivity disorder. *Developmental Psychology*, *43*(1), 70–82. https://doi.org/10.1037/0012-1649.43.1.70
Cicchetti, & Rogosch. (1996). Equifinality and multifinality in developmental psychopathology. *Development and Psychopathology*, 597–600.
Clairivate Analytics. (2019). *Web of Science Core Collection*. Lib Guides. https://clarivate.libguides.com/webofscienceplatform/woscc
Clark, Fleche, & Lekfuangfu. (2017). *The Long-Lasting Effects of Family and Childhood on Adult Wellbeing : Evidence from British Cohort Data* (No. 1493; CEP Discussion Papers, Issue 1493).
Closer. (2019). *Introduction to Longitudinal Studies*. Learning Hub. https://learning.closer.ac.uk/introduction/what-can-longitudinal-studies-show-us/strengths-of-longitudinal-studies/
CLOSER. (2019). *Harmonisation of mental health measures in British birth cohorts*. About the Research We Fund.
Cohen, J. (1988). The effect size index: d. In *Statistical power analysis for the behavioral sciences* (pp. 284–288).
Cohen, L., Manion, & Morrisson. (2011). *Research Methods in Education* (7th ed.). Routledge.
Coleman, D., & Iso-Ahola. (1993). Leisure and Health: The Role of Social Support and Self-Determination. *Journal of Leisure Research*, *25*(2), 111–128. https://doi.org/10.1080/00222216.1993.11969913
Coleman, N. (2015). *Summary of longitudinal surveys: research report*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/444794/DFE-RR458_Summary_of_longitudinal_surveys.pdf
Compas, B. E. (1987). Stress and life events during childhood and adolescence. *Clinical Psychology Review*, *7*(3), 275–302. https://doi.org/Doi 10.1016/0272-7358(87)90037-7
Compas, Bruce E, Connor-Smith, Saltzman, Thomsen, & Wadsworth. (2001). Coping with stress during childhood and adolescence: Problems, progress, and potential in theory and research. *Psychological Bulletin*, *127*(1), 87–127. https://doi.org/10.1037/0033-2909.127.1.87
Conners. (1969). A teacher rating scale for use in drug studies with children. *The American Journal of Psychiatry*, *126*(6), 884–888.
Conners. (2008). *Conners Comprehansive Behavior Rating Scale (CBRS)*. Multi-Health Systems.
Conrad, & Singh. (2018). Reflections on ADHD in a Global Context. In *Global Perspectives on ADHD*. Johns Hopkins University Press.
Cordier, Bundy, Hocking, & Einfeld. (2009). A model for play-based intervention for children with ADHD. *Australian Occupational Therapy Journal*, *56*, 332–340. https://doi.org/10.1111/j.1440-1630.2009.00796.x
Cortese, Adamo, Del Giovane, Mohr-Jensen, Hayes, Carucci, Atkinson, Tessari, Banaschewski, Coghill, Hollis, Simonoff, Zuddas, Barbui, Purgato, Steinhausen, Shokraneh, Xia, & Cipriani. (2018). Comparative efficacy and tolerability of medications for attention-deficit hyperactivity disorder in children, adolescents, and adults: a systematic review and network meta-analysis. *The Lancet Psychiatry*, *5*(9), 727–738. https://doi.org/10.1016/S2215-0366(18)30269-4
Costello, & Maughan. (2015). Annual Research Review: Optimal outcomes of child and adolescent mental illness. *Journal of Child Psychology and Psychiatry*, *56*(3), 324–341. https://doi.org/10.1111/jcpp.12371
Cottle. (1965). Self-concept, ego ideal and the response to action. *Sociology and Social Research*, *50*(1), 78–88.
Cotton, & Baker. (2018a). A data mining and item response mixture modeling method to retrospectively measure Diagnostic and Statistical Manual of Mental Disorders-5

attention deficit hyperactivity disorder in the 1970 British Cohort Study. *International Journal of Methods in Psychiatric Research*. https://doi.org/10.1002/mpr.1753
Cotton, & Baker. (2018b). A data mining and item response mixture modeling method to retrospectively measure Diagnostic and Statistical Manual of Mental Disorders-5 attention deficit hyperactivity disorder in the 1970 British Cohort Study. *International Journal of Methods in Psychiatric Research*. https://doi.org/DOI: 10.1002/mpr.1753
Cotton, Baker, & Wilson. (2015). An exploratory case study of three children with ADHD and social difficulties: Child and parent responses to an educational intervention designed to facilitate self-regulation and deep learning. *The Psychology of Education Review*, *39*(2).
Coughlin. (1990). Recall bias in epidemiologic studies. *Journal of Clinical Epidemiology*, *43*(I), 87–91.
Daly, Delaney, Egan, & Baumeister. (2015). Childhood Self-Control and Unemployment Throughout the Life Span. *Psychological Science*, *26*(6), 709–723. https://doi.org/10.1177/0956797615569001
Daly, & Egan. (2017). Childhood cognitive ability and smoking initiation, relapse and cessation throughout adulthood: evidence from two British cohort studies. *Addiction*, *112*(4), 651–659. https://doi.org/10.1111/add.13554
Davies, Crombie, & Tavakoli. (1998). When can odds ratios mislead? *Bmj*, *316*(7136), 989–991. https://doi.org/10.1136/bmj.317.7166.1155a
de Boo, & Prins. (2007). Social incompetence in children with ADHD: possible moderators and mediators in social-skills training. *Clinical Psychology Review*, *27*(1), 78–97. https://doi.org/10.1016/j.cpr.2006.03.006
Denovan, & Macaskill. (2017). Building resilience to stress through leisure activities: a qualitative analysis. *Annals of Leisure Research*, *20*(4), 446–466. https://doi.org/10.1080/11745398.2016.1211943
Department for Education. (2019). *Schools, pupils, and their characteristics*.
Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750
Diamond, A., & Lee. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science (New York, N.Y.)*, *333*(6045), 959–964. https://doi.org/10.1126/science.1204529
Diamond, D. M., Campbell, Park, Halonen, & Zoladz. (2007). The temporal dynamics model of emotional memory processing: A synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. *Neural Plasticity*, *2007*. https://doi.org/10.1155/2007/60803
Dodgeon, & Parsons. (2011). *Deriving Highest Qualification in NCDS and BCS70* (No. 1; Data Note, Issue 1).
Dolan, Layard, & Metcalfe. (2011). *Measuring subjective well-being for public policy* (Issue February). http://www.statistics.gov.uk/
Dormann, Elith, Bacher, Buchmann, Carl, Carré, Marquéz, Gruber, Lafourcade, Leitão, Münkemüller, Mcclean, Osborne, Reineking, Schröder, Skidmore, Zurell, & Lautenbach. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 027–046. https://doi.org/10.1111/j.1600-0587.2012.07348.x
Dornbusch, Ritter, Leiderman, Roberts, & Fraleigh. (1987). The relation of parenting style to adolescent school performance. *Child Development*, *58*(5), 1244–1257. https://doi.org/10.1111/j.1467-8624.1987.tb01455.x
Durand, & Lambert. (1988). Don't know responses in surveys: Analyses and interpretational consequences. *Journal of Business Research*, *16*(2), 169–188. https://doi.org/10.1016/0148-2963(88)90040-9
Durlak. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, *34*(9), 917–928. https://doi.org/10.1093/jpepsy/jsp004
Durlak, Weissberg, Dymnicki, Taylor, & Schellinger. (2011). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. *Child Development*, *82*(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x
EBSCO Industries Inc. (2019). *EBSCOHost*. Database and Search Engine.

Edelen, & Reeve. (2007). Applying item response theory (IRT) modeling to questionnaire development , evaluation , and refinement. *Quality of Life Research*, *16*, 5–18. https://doi.org/10.1007/s11136-007-9198-0
Elliott, C. D., Murray, & Pearson. (1979). *British Ability Scales*. NFER.
Elliott, G. R., Blasey, Rekshan, Rush, Palmer, Clarke, Kohn, Kaplan, & Gordon. (2017). Cognitive Testing to Identify Children With ADHD Who Do and Do Not Respond to Methylphenidate. *Journal of Attention Disorders*, *21*(14), 1151–1160. https://doi.org/10.1177/1087054714543924
Elliott, J., & Shepherd. (2006). Cohort profile: 1970 British Birth Cohort (BCS70). *International Journal of Epidemiology*, *35*(4), 836–843. https://doi.org/10.1093/ije/dyl174
Elsevier. (2019). *Scopus*. Database and Search Engine. https://www2-scopus-com.ezp.lib.cam.ac.uk/search/form.uri?display=basic
Elton, Alcauter, & Gao. (2014). Network Connectivity Abnormality Profile Supports a Categorical-Dimensional Hybrid Model of ADHD. *Human Brain Mapping*, *35*(9), 4531–4543. https://doi.org/10.1016/j.neuron.2009.10.017.A
Embretson, & Reise. (2000). Item Response Theory for Psychologists. In *Multivariate Applications Series*. Lawrence Erlbaum Associates Inc.
Endler, & Parker. (1990). Stress and anxiety: Conceptual and assessment issues. *Stress Medicine*, *6*(3), 243–248. https://doi.org/10.1002/smi.2460060310
Enoch, Steer, Newman, Gibson, & Goldman. (2010). Early life stress, MAOA, and gene-environment interactions predict behavioral disinhibition in children. *Genes, Brain and Behavior*, *9*(1), 65–74. https://doi.org/10.1111/j.1601-183X.2009.00535.x
Erskine. (2016). *Supplementary Info*. https://ars.els-cdn.com/content/image/1-s2.0-S0890856716311571-mmc1.pdf
Erskine, Ferrari, Nelson, Polanczyk, Flaxman, Vos, Whiteford, & Scott. (2013). Research Review: Epidemiological modelling of attention-deficit/hyperactivity disorder and conduct disorder for the Global Burden of Disease Study 2010. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*(12), 1263–1274. https://doi.org/10.1111/jcpp.12144
Erskine, Norman, Ferrari, Chan, Copeland, Whiteford, & Scott. (2016). Long-Term Outcomes of Attention-Deficit/Hyperactivity Disorder and Conduct Disorder: A Systematic Review and Meta-Analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*, *55*(10), 841–850. https://doi.org/10.1016/j.jaac.2016.06.016
ESRC. (2016). *ESRC Framework for Research Ethics: Core Principles*. Website. http://www.esrc.ac.uk/funding/guidance-for-applicants/research-ethics/our-core-principles/
ESRC. (2019). *Health and Wellbeing*. Research Topics. https://esrc.ukri.org/research/research-topics/health-and-wellbeing/
Faraone, Biederman, & Mick. (2005). The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. *Psychological Medicine*, *36*(02), 159. https://doi.org/10.1017/S003329170500471X
Farrar, & Glauber. (1967). Multicollinearity in Regression Analysis : The Problem Revisited. *The Review of Economics and Statistics*, *49*(1), 92–107. https://doi.org/10.2307/30040678
Faul, Erdfelder, Lang, & Buchner. (2009). *G*Power 3* (3.1.7). http://www.gpower.hhu.de/en.html
Feinstein. (2000). The Relative Economic Importance of Academic Psychological and Behavioural Attributes Developed in Childhood. *Centre for Economic Performance*, 1–34.
Feinstein, & Symons. (1997). *Attainment in Secondary School* (No. 341; Discussion Paper, Issue March).
Ferguson. (2009). An Effect Size Primer: A Guide for Clinicians and Researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538. https://doi.org/10.1037/a0015808
Ferro, & Boyle. (2015). The Impact of Chronic Physical Illness, Maternal Depressive Symptoms, Family Functioning, and Self-esteem on Symptoms of Anxiety and Depression in Children. *Journal of Abnormal Child Psychology*, *43*(1), 177–187. https://doi.org/10.1007/s10802-014-9893-6
Field. (2009). *Discovering Statistics Using SPSS* (Third). Sage.
Finkelman, Green, Gruber, & Zaslavsky. (2011). A zero- and K-inflated mixture model for

health questionnaire data. *Statistics in Medicine*, *30*(9), 1028–1043. https://doi.org/10.1002/sim.4217

Flouri. (2006). Parental interest in children's education, children's self-esteem and locus of control, and later educational attainment: Twenty-six year follow-up of the 1970 British Birth Cohort. *British Journal of Educational Psychology*, *76*(1), 41–55. https://doi.org/10.1348/000709905X52508

Flouri. (2007). Early family environments may moderate prediction of low educational attainment in adulthood: The cases of childhood hyperactivity and authoritarian parenting. *Educational Psychology*, *27*(6), 737–751. https://doi.org/10.1080/01443410701309282

Flouri, Francesconi, Papachristou, Midouhas, & Lewis. (2019). Stressful life events, inflammation and emotional and behavioural problems in children: A population-based study. *Brain, Behavior, and Immunity*, *80*(November 2018), 66–72. https://doi.org/10.1016/j.bbi.2019.02.023

Floyd, Evans, & McGrew. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) abilities and mathematics achievement across the school-age years. *Psychology in the Schools*, *40*(2), 155–171. https://doi.org/10.1002/pits.10083

Frazier, & Wilson. (2018). Does physical activity in children improve ADHD symptoms? *Evidence Based Practice*, *21*(7). https://doi.org/10.1097/01.EBP.0000545192.28779.db

Furnham, & Cheng. (2017). Socio-Demographic Indicators, Intelligence, and Locus of Control as Predictors of Adult Financial Well-Being. *Journal of Intelligence*, *5*(4), 11. https://doi.org/10.3390/jintelligence5020011

Furnival, & Wilson. (1974). Regressions by Leaps and Bounds. *Technometrics*, *16*(4), 499–511.

Gale, Deary, Batty, & Schoon. (2007). IQ in childhood and vegetarianism in adulthood: 1970 British cohort study. *British Medical Journal*, *334*(7587), 245–248. https://doi.org/10.1136/bmj.39030.675069.55

Gale, Hatch, Batty, & Deary. (2009). Intelligence in childhood and risk of psychological distress in adulthood: The 1958 National Child Development Survey and the 1970 British Cohort Study. *Intelligence*, *37*(6), 592–599. https://doi.org/10.1016/j.intell.2008.09.002

Galobardes, Shaw, Lawlor, Lynch, & Smith. (2006). Indicators of socioeconomic position (part 1). *Journal of Epidemiology and Community Health*, *60*(1), 7–12. https://doi.org/10.1136/jech.2004.023531

Gammage. (1974). *Socialization schooling and locus of control*. University of Bristol.

Gammage. (1982). *Children and Schooling: Issues in Childhood*. George Allen & Unwin Ltd.

Garcia-Barrera, Kamphaus, & Bandalos. (2011). Theoretical and Statistical Derivation of a Screener for the Behavioral Assessment of Executive Functions in Children. *Psychological Assessment*, *23*(1), 64–79. https://doi.org/10.1037/a0021097

Garrido, Kelley, Paris, Roza, Meier, Morrison, & Aldridge. (2014). Methods for constructing and assessing propensity scores. *Health Services Research*, *49*(5), 1701–1720. https://doi.org/10.1111/1475-6773.12182

Gathercole, Brown, & Pickering. (2003). Memory and Education. *Educational and Child Psychology*, *20*(3), 109–122.

Gawrilow, & Gollwitzer. (2007). Implementation Intentions Facilitate Response Inhibition in Children with ADHD. *Cognitive Therapy and Research*, *32*(2), 261–280. https://doi.org/10.1007/s10608-007-9150-1

Gawrilow, Gollwitzer, & Oettingen. (2011). If-Then Plans Benefit Executive Functions in Children with ADHD. *Journal of Social and Clinical Psychology*, *30*(6), 616–646. https://doi.org/10.1521/jscp.2011.30.6.616

Gawrilow, Morgenroth, Schultz, Oettingen, & Gollwitzer. (2012). Mental contrasting with implementation intentions enhances self-regulation of goal pursuit in schoolchildren at risk for ADHD. *Motivation and Emotion*, *37*(1), 134–145. https://doi.org/10.1007/s11031-012-9288-3

Gelman, & Hill. (2006). Causal inference using regression on the treatment variable. In *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gol, & Jarus. (2005). Effect of a social skills training group on everyday activities of children with attention-deficit-hyperactivity disorder. *Developmental Medicine and Child Neurology*, *47*(8), 539–545. http://www.ncbi.nlm.nih.gov/pubmed/16108454

Gomez. (2007). Australian Parent and Teacher Ratings of the DSM-IV ADHD Symptoms. *Journal of Attention Disorders*, *11*(1), 17–27. https://doi.org/10.1177/1087054706295665
Gomez. (2008). Parent ratings of the ADHD items of the disruptive behavior rating scale: Analyses of their IRT properties based on the generalized partial credit model. *Personality and Individual Differences*, *45*(2), 181–186. https://doi.org/10.1016/j.paid.2008.04.001
Gomez. (2011). Item response theory analyses of adult self-ratings of the ADHD symptoms in the current symptoms scale. *Assessment*, *18*(4), 476–486. https://doi.org/10.1177/1073191110386341
Gomez. (2012). Item response theory analyses of adolescent self-ratings of the ADHD symptoms in the Disruptive Behavior Rating Scale. *Personality and Individual Differences*, *53*(8), 963–968. https://doi.org/10.1016/j.paid.2012.07.012
Gomez, Vance, & Gomez. (2011). Item response theory analyses of parent and teacher ratings of the ADHD symptoms for recoded dichotomous scores. *Journal of Attention Disorders*, *15*(4), 269–285. https://doi.org/10.1177/1087054709356404
Goodman, A., Joshi, Nasim, & Tyler. (2015). *Social and emotional skills in childhood and their long-term effects on adult life* (Issue March).
Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *38*(5), 581–586. http://www.ncbi.nlm.nih.gov/pubmed/9255702
Goodyer, Wright, & Altham. (1990). The friendships and recent life events of anxious and depressed school-age children. *British Journal of Psychiatry*, *156*(MAY.), 689–698. https://doi.org/10.1192/bjp.156.5.689
Gorter, Fox, & Twisk. (2015). Why item response theory should be used for longitudinal questionnaire data analysis in medical research. *BMC Medical Research Methodology*, *15*, 55-015-0050–x. https://doi.org/10.1186/s12874-015-0050-x [doi]
Gozal, & Molfese. (2005). *Attention Deficit Hyperacticity Disorder: From Genes to Patients* (Gozal & Molfese (Eds.)). Humana Press. https://doi.org/10.1017/CBO9781107415324.004
Green, Bunge, Briones Chiongbian, Barrow, & Ferrer. (2017). Fluid reasoning predicts future mathematical performance among children and adolescents. *Journal of Experimental Child Psychology*, *157*, 125–143. https://doi.org/10.1016/j.jecp.2016.12.005
Greenberg, Kindschi, Dupuy, & Hughes. (2007). *T.O.V.A. Clinical Manual* (Vol. 410). The TOVA Company.
Greenhill, Halperin, & Abikoff. (1999). Stimulant medications. *Journal of the American Academy of Child and Adolescent Psychiatry*, *38*(5), 503–512. https://doi.org/10.1097/00004583-199905000-00011
Groen-Blokhuis, Middeldorp, Kan, Abdellaoui, Van Beijsterveldt, Ehli, Davies, Scheet, Xiao, Hudziak, Hottenga, Neale, & Boomsma. (2014). Attention-deficit/hyperactivity disorder polygenic risk scores predict attention problems in a population-based sample of children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *53*(10), 1123–1129. https://doi.org/10.1016/j.jaac.2014.06.014
Gunter, & Daly. (2012). Computers in Human Behavior Causal or spurious: Using propensity score matching to detangle the relationship between violent video games and violent behavior. *Computers in Human Behavior*, *28*(4), 1348–1355. https://doi.org/10.1016/j.chb.2012.02.020
Habel, Cooper, Sox, Chan, Fireman, Arbogast, Cheetham, Quinn, Dublin, Boudreau, Andrade, Pawloski, Raebel, Smith, Achacoso, Uratsu, Go, Sidney, Nguyen-Huynh, … Selby. (2011). ADHD medications and risk of serious cardiovascular events in young and middle-aged adults. *JAMA: Journal of the American Medical Association*, *306*(24), 2673–2683. laurel.habel@kp.org,
Hackman, & Farah. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, *13*(2), 65–73. https://doi.org/10.1016/j.tics.2008.11.003
Hahs-Vaughn, & Onwuegbuzie. (2006). Estimating and using propensity score analysis with complex samples. *Journal of Experimental Education*, *75*(1), 31–65. https://doi.org/10.3200/JEXE.75.1.31-65
Halperin, Marks, Bedard, Chacko, Curchack, Yoon, & Healey. (2013). Training executive, attention, and motor skills: a proof-of-concept study in preschool children With ADHD. *Journal of Attention Disorders*, *17*(8), 711–721. https://doi.org/10.1177/1087054711435681

Hambleton, Swaminithan, & Rogers. (1991). *Fundamentals of Item Response Theory*. SAGE Publications. https://books.google.co.uk/books?id=gW05DQAAQBAJ&printsec=frontcover#v=onepage&q&f=false
Hanć, Szwed, Słopień, Wolańczyk, Dmitrzak-Węglarz, & Ratajczak. (2018). Perinatal Risk Factors and ADHD in Children and Adolescents: A Hierarchical Structure of Disorder Predictors. *Journal of Attention Disorders*, *22*(9), 855–863. https://doi.org/10.1177/1087054716643389
Happé, Booth, Charlton, & Hughes. (2006). Executive function deficits in autism spectrum disorders and attention-deficit/hyperactivity disorder: examining profiles across domains and ages. *Brain and Cognition*, *61*(1), 25–39. https://doi.org/10.1016/j.bandc.2006.03.004
Hardouin, Bonnaud-Antignac, & Sebille. (2011). Nonparametric item repsonse theory using Stata. *The Stata Journal*, *1*, 30–51. http://www.stata-journal.com/sjpdf.html?articlenum=st0216
Harris, & Horst. (2016). A Brief Guide to Decisions at Each Step of the Propensity Score Matching Process. *Practical Assessment, Research & Evaluation*, *21*(4), 1–11.
Hart. (1985). LAWSEQ: ITS RELATION TO OTHER MEASURES OF SELF-ESTEEM AND ACADEMIC ABILITY. *British Journal of Educational Psychology*, *55*, 167–169.
Hattie. (1992). *Self Concept*. Laurence Erlbaum Associates. https://books.google.co.uk/books?hl=en&lr=&id=Vh0BAwAAQBAJ&oi=fnd&pg=PP1&dq=self+concept&ots=NI5c_qmJOs&sig=dDzBPnwEVbf7N1OlEE8RfCpCaGM&redir_esc=y#v=onepage&q&f=false
Health Improvement Analytical Team. (2014). Wellbeing: Why it matters to health policy. In *Narrative*. https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjRzpmmoO7OAhWKJsAKHa7iBCkQFghFMAA&url=https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/277566/Narrative__January_201
Hechtman, Swanson, Sibley, Stehli, Owens, Mitchell, Arnold, Molina, Hinshaw, Jensen, Abikoff, Perez, Howard, Hoza, Etcovitch, Houssais, Lakes, & Nichols. (2016). Functional Adult Outcomes 16 Years After Childhood Diagnosis of Attention-Deficit/Hyperactivity Disorder: MTA Results. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*(11), 945–952. https://doi.org/10.1016/j.jaac.2016.07.774
Heidbreder. (2015). ADHD symptomatology is best conceptualized as a spectrum: a dimensional versus unitary approach to diagnosis. *ADHD Attention Deficit and Hyperactivity Disorders*, *7*(4), 249–269. https://doi.org/10.1007/s12402-015-0171-4
Heinze, Wallisch, & Dunkler. (2018). Variable selection – A review and recommendations for the practicing statistician. *Biometrical Journal*, *60*(3), 431–449. https://doi.org/10.1002/bimj.201700067
Helliwell, Layard, & Sachs. (2012). *World Happiness Report*. https://doi.org/10.1007/978-3-319-74336-3_371-1
Hennekens, & Buring. (1987). *Epidemiology in Medicine* (Mayrent (Ed.); First). Lippincott, Willions & Wilkins.
Higgins, Thompson, Deeks, & Altman. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557
Hinshaw, Scheffler, Fulton, Aase, Banaschewski, Cheng, Mattos, Holte, Levy, Sadeh, Sergeant, Taylor, & Weiss. (2011). International variation in treatment procedures for ADHD: social context and recent trends. *Psychiatric Services (Washington, D.C.)*, *62*(5), 459–464. https://doi.org/10.1176/appi.ps.62.5.459
Ho, Imai, King, & Stuart. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis*, *15*, 199–236. https://doi.org/10.1093/pan/mpl013
Hobbs, & Vignoles. (2010). Is children's free school meal 'eligibility' a good proxy for family income? *British Educational Research Journal*, *36*(4), 673–690. https://doi.org/10.1080/01411920903083111
Holland, & Thayer. (1986). Differential Item Functioning and the Mantel-Haenszel Procedure.

In *Program Statistics Research* (Issues 86–89). https://doi.org/10.1002/j.2330-8516.1986.tb00186.x
Holmes, & Rahe. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research*, *11*, 213–218. https://doi.org/http://dx.doi.org/10.1016/j.mhpa.2010.02.001
Howitt, & Cramer. (2008). *Introduction to Research Methods in Psychology* (Second). Pearson Education Limited.
Hoza. (2007). Peer Functioning in Children With ADHD. *Journal of Pediatric Psychology*, *32*(6), 655–663.
Hoza, Mrug, Gerdes, Hinshaw, Bukowski, Gold, Kraemer, Pelham, Wigal, & Arnold. (2005). What Aspects of Peer Relationships Are Impaired in Children With Attention-Deficit/Hyperactivity Disorder? *Journal of Consulting and Clinical Psychology*, *73*(3), 411–423. https://doi.org/10.1037/0022-006X.73.3.411
Hu, & Bentler. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118
Hung, & Gawrilow. (2016). A Selective Review of the Effect of Self-Regulation Strategies on Response Inhibition in Children with ADHD. 華醫學報 第, *45*(12), 110–131.
Hupfeld, Abagis, & Shah. (2019). Living "in the zone": hyperfocus in adult ADHD. *ADHD Attention Deficit and Hyperactivity Disorders*, *11*(2), 191–208. https://doi.org/10.1007/s12402-018-0272-y
Huppert. (2009). Psychological Well-being: Evidence Regarding its Causes and Consequences. *Applied Psychology: Health and Well-Being*, *1*(2), 137–164. https://doi.org/10.1111/j.1758-0854.2009.01008.x
Iacus, King, & Porro. (2011). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis*, *20*, 1–24. https://doi.org/10.1093/pan/mpr013
Iacus, King, & Porro. (2014). *cem: Coarsened Exact Matching Software* (1.1.14). https://gking.harvard.edu/cem
IDM Computer Solutions. (2019). *UltraCompare* (No. 18). http://www.ultraedit.com
Imai, King, & Stuart. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society*, *171*(2), 481–502.
Institute of Child Health. (1975). *The 1970 Birth Cohort 5-Year Follow Up User Guide*.
Iwasaki, & Mannell. (2000). Hierarchical dimensions of leisure stress coping. *Leisure Sciences*, *22*(3), 163–181. https://doi.org/10.1080/01490409950121843
Iwasaki, & Schneider. (2003). Leisure, stress, and coping: An evolving area of inquiry. *Leisure Sciences*, *25*(2–3), 107–113. https://doi.org/10.1080/01490400306567
Jann. (2017). *Kernel matching with automatic bandwidth selection*.
Jenkins, & Sabates. (2007). *The Classification of Qualifications in Social Surveys* (2007/2; CLS Cohort Studies, Issue March).
Jensen, Hinshaw, Kraemer, Lenora, Newcorn, Abikoff, March, Arnold, Cantwell, Conners, Elliott, Greenhill, Hechtman, Hoza, Pelham, Severe, Swanson, Wells, Wigal, & Vitiello. (2001). ADHD comorbidity findings from the MTA study: comparing comorbid subgroups. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*(2), 147–158. https://doi.org/10.1097/00004583-200102000-00009
Johnson, Wiersema, & Kuntsi. (2009). What would Karl Popper say? Are current psychological theories of ADHD falsifiable? *Behavioral and Brain Functions*, *3 March 2009*, 5–15. https://doi.org/10.1186/1744-9081-5-15
Joinson, Sullivan, Von Gontard, & Heron. (2016). Stressful Events in Early Childhood and Developmental Trajectories of Bedwetting at School Age. *Journal of Pediatric Psychology*, *41*(9), 1002–1010. https://doi.org/10.1093/jpepsy/jsw025
Joshi, Nasim, & Goodman. (2016). The Measurement of Social and Emotional Skills and their Association with Academic Attainment in British Cohort Studies. In Khine & Areepattamannil (Eds.), *Non-cognitive Skills and Factors in Educational Attainment* (pp. 239–264). Sense Publishers. https://doi.org/10.1007/978-94-6300-591-3_11
Judge, & Hurst. (2007). Capitalizing on one's advantages: Role of core self-evaluations. *Journal of Applied Psychology*, *92*(5), 1212–1227. https://doi.org/10.1037/0021-9010.92.5.1212
Kaat, & Farmer. (2017). Commentary : Lingering questions about the Social Responsiveness

Scale short form. A commentary on Sturm et al. (2017). *The Journal of Child Psychology and Psychiatry*, *58*(9), 1062–1064. https://doi.org/10.1111/jcpp.12801

Karalunas, Fair, Musser, Iyer, & Nigg. (2014). Title : Toward biologically-based nosology : ADHD subtyping using temperament dimensions. *JAMA Psychiatry*, *71*(9), 1015–1024.

Karalunas, & Nigg. (2019). Heterogeneity and subtyping in attention-deficit/hyperactivity disorder— Considerations for emerging research using person-centered computational approaches. *Biological Psychiatry*. https://doi.org/10.1016/J.BIOPSYCH.2019.11.002

Kempton, Vance, Maruff, Luk, Costin, & Pantelis. (1999). Executive function and attention deficit hyperactivity disorder: stimulant medication and better executive function performance in children. *Psychological Medicine*, *29*, 527–538. https://doi.org/10.1017/S0033291799008338

Kenny, D A. (2015). *Measuring model fit*. Davidakenny.Net. http://www.davidakenny.net/cm/fit.htm

Kessler, Adler, Barkley, Biederman, Conners, Faraone, Greenhill, Jaeger, Secnik, Spencer, Üstün, & Zaslavsky. (2005). Patterns and predictors of attention-deficit/hyperactivity disorder persistence into adulthood: Results from the national comorbidity survey replication. *Biological Psychiatry*, *57*(11), 1442–1451. https://doi.org/10.1016/j.biopsych.2005.04.001

Ketende, McDonald, & Dex. (2010). *Non-response in the 1970 British Cohort Study (BCS70) from birth to 34 years* (No. 4; CLS Cohort Studies, Issue October).

King. (2015). *Simplifying Matching Methods for Causal Inference*. https://doi.org/10.1080/19406940.2015.1063528

King, Lucas, & Nielsen. (2017). The Balance-Sample Size Frontier in Matching Methods for Causal Inference. *American Journal of Political Science*, *61*(2), 473–489. https://doi.org/10.1111/ajps.12272

King, & Nielsen. (2018). Why Propensity Scores Should Not Be Used for Matching. In *Political Analysis*.

Klein, Ph, Mannuzza, Olazagasti, Belsky, Hutchison, Lashua-shriftman, & Castellanos. (2012). Clinical and Functional Outcome of Childhood ADHD 33 Years Later. *Archive of General Psychiatry*, *69*(12), 1295–1303. https://doi.org/10.1001/archgenpsychiatry.2012.271.Clinical

Kortekaas-Rijlaarsdam, Luman, Sonuga-barke, & Oosterlaan. (2018). Does methylphenidate improve academic performance? A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *1*(0123456789). https://doi.org/10.1007/s00787-018-1106-3

Kotov, Waszczuk, Krueger, Forbes, Watson, Clark, Achenbach, Althoff, Ivanova, Michael Bagby, Brown, Carpenter, Caspi, Moffitt, Eaton, Forbush, Goldberg, Hasin, Hyman, … Zimmerman. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Kuntsi, J, Oosterlaan, & Stevenson. (2001). Psychological mechanisms in hyperactivity: I. Response inhibition deficit, working memory impairment, delay aversion, or something else? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(2), 199–210. https://doi.org/10.1017/S0021963001006709

Kuntsi, Jonna, & Klein. (2011). Intraindividual Variability in ADHD and Its Implications for Research of Causal Links. In Stanford & Tannock (Eds.), *Current Topics in Behavioral Neurosciences (9): Vol. Behavioral* (Behavioral, Issue July 2011, pp. 67–91). Springer-Verlag. https://doi.org/10.1007/7854

Kuntsi, Jonna, Wood, Rijsdijk, Johnson, Andreou, Albrecht, Arias-Vasquez, Buitelaar, McLoughlin, Rommelse, Sergeant, Sonuga-Barke, Uebel, van der Meere, Banaschewski, Gill, Manor, Miranda, Mulas, … Asherson. (2010). Separation of cognitive impairments in attention-deficit/hyperactivity disorder into 2 familial factors. *Archives of General Psychiatry*, *67*(11), 1159–1167. https://doi.org/10.1001/archgenpsychiatry.2010.139

Kurgan, & Musilek. (2006). A survey of Knowledge Discovery and Data Mining process models. *The Knowledge Engineering Review*, *21*(01), 1. https://doi.org/10.1017/S0269888906000737

Kuriyan, Pelham, Molina, Waschbusch, Gnagy, Sibley, Babinski, Walther, Cheong, Yu, & Kent. (2013). Young adult educational and vocational outcomes of children diagnosed

with ADHD. *Journal of Abnormal Child Psychology*, *41*(1), 27–41. https://doi.org/10.1007/s10802-012-9658-z
Lahey, Applegate, McBurnett, Biederman, Greenhill, Hynd, Barkley, Newcorn, Jensen, & Richters. (1994). DSM-IV field trials for attention deficit hyperactivity disorder in children and adolescents. *American Journal of Psychiatry*, *151*(11), 1673–1685. https://doi.org/10.1176/ajp.151.11.1673
Lange, Reichl, Lange, Tucha, & Tucha. (2010). The history of attention deficit hyperactivity disorder. *ADHD Attention Deficit and Hyperactivity Disorders*, *2*(4), 241–255. https://doi.org/10.1007/s12402-010-0045-8
Langton, Collishaw, Goodman, Pickles, & Maughan. (2011). An emerging income differential for adolescent emotional problems. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *52*(10), 1081–1088. https://doi.org/10.1111/j.1469-7610.2011.02447.x
Lara, Fayyad, de Graaf, Kessler, Aguilar-Gaxiola, Angermeyer, Demytteneare, de Girolamo, Haro, Jin, Karam, Lépine, Mora, Ormel, Posada-Villa, & Sampson. (2009). Childhood Predictors of Adult Attention-Deficit/Hyperactivity Disorder: Results from the World Health Organization World Mental Health Survey Initiative. *Biological Psychiatry*, *65*(1), 46–54. https://doi.org/10.1016/j.biopsych.2008.10.005
Lavoie, Pereira, & Talwar. (2016). Children’s Physical Resilience Outcomes: Meta-Analysis of Vulnerability and Protective Factors. *Journal of Pediatric Nursing*, *31*(6), 701–711. https://doi.org/10.1016/j.pedn.2016.07.011
Lawrence. (1973). *Improved reading through counselling*. Ward Lock.
Lawrence. (1978). *Counselling Students with Reading Difficulties: A Handbook for Tutors and Organisers*. Good Reading.
Lawrence. (1981). The Development of a Self-Esteem Questionnaire. *British Journal of Educational Psychology*, *51*, 245–251.
Layard, Clark, Cornaglia, Powdthavee, & Vernoit. (2014). What predicts a successful life? A life-course model of well-being. *Economic Journal*, *124*(580), F720–F738. https://doi.org/10.1111/ecoj.12170
Lebel, Walker, Leemans, Phillips, & Beaulieu. (2008). Microstructural maturation of the human brain from childhood to adulthood. *NeuroImage*, *40*(3), 1044–1055. https://doi.org/10.1016/j.neuroimage.2007.12.053
Lee, Schachar, Chen, Ornstein, Charach, Barr, & Ickowicz. (2008). Predictive validity of DSM-IV and ICD-10 criteria for ADHD and hyperkinetic disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *49*(1), 70–78. https://doi.org/10.1111/j.1469-7610.2007.01784.x
Leite, Stapleton, & Bettini. (2018). Propensity Score Analysis of Complex Survey Data with Structural Equation Modeling : a Tutorial with Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *5511*. https://doi.org/10.1080/10705511.2018.1522591
Lenhard, & Lenhard. (2016). *Computation of Effect Sizes*. Psychometrica. https://doi.org/10.13140/RG.2.1.3478.4245
Leuven, & Sianesi. (2018). *PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing*. https://econpapers.repec.org/RePEc:boc:bocode:s432001
Li, Reise, Chronis-Tuscano, Mikami, & Lee. (2015). Item Response Theory Analysis of ADHD Symptoms in Children With and Without ADHD. *Assessment*, *23*(6). https://doi.org/10.1177/1073191115591595
Lindhiem. (2013). *Quantifying Diagnostic Uncertainty Using Item Response Theory*. *25*(2), 456–466. https://doi.org/10.1037/a0031392.Quantifying
Lindhiem, Yu, Grasso, Kolko, & Youngstrom. (2015). Adapting the Posterior Probability of Diagnosis (PPOD) Index to Enhance Evidence-Based Screening: An Application to ADHD in Primary Care. *Assessment*, *22*(2), 198–207. https://doi.org/10.1097/MPG.0b013e3181a15ae8.Screening
Lindsey, & Sheather. (2010). Variable Selection in Linear Regression. *The Stata Journal*, *10*(4), 650–669.
Linton, Dieppe, & Medina-Lara. (2016). Review of 99 self-report measures for assessing well-being in adults: Exploring dimensions of well-being and developments over time. *BMJ Open*, *6*(7). https://doi.org/10.1136/bmjopen-2015-010641

Lloyd, & Devine. (2012). Psychometric properties of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in Northern Ireland. *Journal of Mental Health*, *21*(3), 257–263. https://doi.org/10.3109/09638237.2012.670883

Loe, & Feldman. (2007). Academic and educational outcomes of children with ADHD. *Journal of Pediatric Psychology*, *32*(6), 643–654. https://doi.org/10.1093/jpepsy/jsl054

Lovecky. (1999). Gifted Children with AD/HD. *Annual CHADD International Conference*, *4*, 1–10.

Luo, Weibman, Halperin, & Li. (2019). A review of heterogeneity in attention deficit/hyperactivity disorder (ADHD). *Frontiers in Human Neuroscience*, *13*(February), 1–12. https://doi.org/10.3389/fnhum.2019.00042

Lupien, McEwen, Gunnar, & Heim. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience*, *10*(6), 434–445. https://doi.org/10.1038/nrn2639

Maccoby, & Martin. (1983). Socialization in the context of the family: Parent-child interaction. In P. H. Mussen (Series Ed.) & E. M. Hetherington (Vol. Ed.) (Eds.), *Handbook of child psychology: Vol. 4: Socialization, personality and social development ; E. Mavis Hetherington, volume editor* (4th ed, pp. 1–101). Wiley.

Machin, & Vignoles (Eds.). (2005). *What's the Good of Education? The Economics of Education in the UK*. Princeton University Press. http://assets.press.princeton.edu/chapters/s7947.pdf

MacKinnon, Kingsbury, & Evans. (2018). The Association Between Prenatal Stress and Externalizing Symptoms in Childhood: Evidence From the Avon Longitudinal Study of Parents and Children. *Biological Psychiatry*, *83*(2), 100–108. https://research-information.bristol.ac.uk/files/132629834/Prenatal_stress_hyperactivity_conduct_disorder.pdf

Magis. (2013). A Note on the Item Information Function of the Four-Parameter Logistic Model. *Applied Psychological Measurement*, *37*(4), 304–315. https://doi.org/10.1177/0146621613475471

Magnus, & Thissen. (2017). Item Response Modeling of Multivariate Count Data with Zero Inflation, Maximum Inflation, and Heaping. *Journal of Educational and Behavioral Statistics*, *42*(5), 531–558. https://doi.org/10.3102/1076998617694878

Maldonado, & Greenland. (1993). Simulation Study of Confounder-Selection Strategies. *American Journal of Epidemiology*, *138*(11), 923–936.

Masten. (2001). Ordinary magic: Resilience processes in development. *American Psychologist*, *56*(3), 227–238. https://doi.org/10.1037/0003-066X.56.3.227

Masten. (2014). Global Perspectives on Resilience in Children and Youth. *Child Development*, *85*(1), 6–20. https://doi.org/10.1111/cdev.12205

Matza, Paramore, & Prasad. (2005). A review of the economic burden of ADHD. *Cost Effectiveness and Resource Allocation : C/E*, *3*, 5. https://doi.org/10.1186/1478-7547-3-5

Mayer, Lopez-Duran, Kovacs, George, Baji, Kapornai, Kiss, & Vetró. (2009). Stressful life events in a clinical sample of depressed children in Hungary. *Journal of Affective Disorders*, *115*(1–2), 207–214. https://doi.org/10.1016/j.jad.2008.08.018

Mayes, & Calhoun. (2007). Learning, attention, writing, and processing speed in typical children and children with ADHD, autism, anxiety, depression, and oppositional-defiant disorder. *Child Neuropsychology : A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *13*(6), 469–493. https://doi.org/10.1080/09297040601112773

Mayes, Calhoun, Bixler, & Zimmerman. (2009). IQ and neuropsychological predictors of academic achievement. *Learning and Individual Differences*, *19*(2), 238–241. https://doi.org/10.1016/j.lindif.2008.09.001

McKelvey, & Zavoina. (1975). A statistical model for the analysis of ordinal level dependent variables. In Aldrich, Alt, & Lupia (Eds.), *Journal of mathematical sociology* (Vol. 4, Issue 1, pp. 103–120). University of Michigan Press. https://doi.org/https://doi.org/10.1080/0022250X.1975.9989847

McQuade, Vaughn, Hoza, Murray-Close, Molina, Arnold, & Hechtman. (2014). Perceived social acceptance and peer status differentially predict adjustment in youth with and without ADHD. *Journal of Attention Disorders*, *18*(1), 31–43. https://doi.org/10.1177/1087054712437582

Mehan, Hertweck, & Meihls. (1986). *Handicapping the Handicapped: Decision Making in Students' Educational Careers*. Stanford University Press.
Mensah, & Hobcraft. (2008). Childhood deprivation, health and development: Associations with adult health in the 1958 and 1970 British prospective birth cohort studies. *Journal of Epidemiology and Community Health*, *62*(7), 599–606. https://doi.org/10.1136/jech.2007.065706
Merrell, & Wolfe. (1998). The relationship of teacher-rated social skills deficits and ADHD characteristics among kindergarten-age children. *Psychology in the Schools*, *35*(2), 101–109.
Metin, Wiersema, Verguts, Gasthuys, van Der Meere, Roeyers, & Sonuga-Barke. (2014). Event rate and reaction time performance in ADHD: Testing predictions from the state regulation deficit hypothesis using an ex-Gaussian model. *Child Neuropsychology : A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *7049*(December 2015), 1–11. https://doi.org/10.1080/09297049.2014.986082
Moffitt, Caspi, Taylor, Kokaua, Milne, Polanczyk, & Poulton. (2010). How common are common mental disorders? Evidence that lifetime prevalence rates are doubled by prospective versus retrospective ascertainment. *Psychological Medicine*, *40*(6), 899–909. https://doi.org/10.1017/S0033291709991036
Moffitt, Houts, Asherson, Belsky, Corcoran, Hammerle, Harrington, Hogan, Meier, Polanczyk, Poulton, Ramrakha, Sugden, Williams, Rohde, & Caspi. (2015). Is adult ADHD a childhood-onset neurodevelopmental disorder? Evidence from a four-decade longitudinal cohort study. *American Journal of Psychiatry*, *172*(10), 967–977. https://doi.org/10.1176/appi.ajp.2015.14101266
Moher, Shamseer, Clarke, Ghersi, Liberati, Petticrew, Shekelle, Stewart, & Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1), 1–9.
Molina, Hinshaw, Swanson, Arnold, Vitiello, Jensen, Epstein, Hoza, Hechtman, Abikoff, Elliott, Greenhill, Newcorn, Wells, Wigal, Gibbons, Hur, & Houck. (2009). The MTA at 8 years: prospective follow-up of children treated for combined-type ADHD in a multisite study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*(5), 484–500. https://doi.org/10.1097/CHI.0b013e31819c23d0
Mostafa. (2014). *The BCS70 sample , attrition and non- response*.
Mostafa, & Wiggins. (2014). *Handling attrition and non-response in the 1970 British Cohort Study* (Issue June).
Mostafa, & Wiggins. (2015). The impact of attrition and non-response in birth cohort studies: A need to incorporate missingness strategies. *Longitudinal and Life Course Studies*, *6*(2), 131–146. https://doi.org/10.14301/llcs.v6i2.312
MTA. (1999). Moderators and Mediators of Treatment Response for Children With Attention-Deficit/ Hyperactivity Disorder. *Archives of General Psychiatry*, *56*, 1088–1096.
Murasko. (2007). A lifecourse study on education and health: The relationship between childhood psychosocial resources and outcomes in adolescence and young adulthood. *Social Science Research*, *36*(4), 1348–1370. https://doi.org/10.1016/j.ssresearch.2007.01.001
Muthen & Muthen. (2004). *Interpretation of R-square in categorical data*. Mplus Discussion Boards.
Muthen & Muthen. (2007). *Chi Square fit of 0.000*. Mplus Discussion Boards. http://www.statmodel.com/discussion/messages/11/2134.html?1568271342
Muthen & Muthen. (2015). *Assessment of model with multinomial variable*. Mplus Discussion Boards. http://www.statmodel.com/discussion/messages/11/21690.html?1491876051
Muthen & Muthen. (2017). *Mplus* (No. 8).
Muthén, B. O. (2016). *Long running time*. Mplus Discussion Boards. http://www.statmodel.com/discussion/messages/11/22289.html?1453435696
Muthén, B. O., & Asparouhov. (2006). Item response mixture modeling: Application to tobacco dependence criteria. *Addictive Behaviors*, *31*(6), 1050–1066. https://doi.org/10.1016/j.addbeh.2006.03.026
Muthén, B. O., & Schultzberg. (2017). *Regression And Mediation Analysis Using Mplus* (August 16, Topic 11).
Muthén, L. K., & Muthén. (2017). Mplus User's Guide. Eighth Edition. *Los Angeles, CA: Muthén*

& Muthén. https://doi.org/10.1111/j.1600-0447.2011.01711.x
National Center for Biotechnology Information and U.S. National Library of Medicine. (2019). *PubMed*. Database and Search Engine. https://www-ncbi-nlm-nih-gov.ezp.lib.cam.ac.uk/pubmed?otool=igbcambulib
National Institute of Mental Health. (2019). *Research Domain Criteria*. Research Funded by NIMH. https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/index.shtml
Newman, Tay, & Diener. (2014). Leisure and Subjective Well-Being: A Model of Psychological Mechanisms as Mediating Factors. *Journal of Happiness Studies*, *15*(3), 555–578. https://doi.org/10.1007/s10902-013-9435-x
NHS. (2018). *Low birth weight term babies*. NHS Digital. https://digital.nhs.uk/data-and-information/national-indicator-library/low-birth-weight-term-babies
NHS Digital. (2018). Mental health of Children and Young People in England, 2017. In *Summary of key findings* (Issue November). https://doi.org/10.7748/paed2009.06.21.5.28.c7079
Nichols. (2007). Causal inference with observational data. *The Stata Journal*, *7*(4), 507–541.
Nichols. (2009). *Causal inference with observational data A brief review of quasi-experimental methods Virtually every set of estimates invites some kind of causal inference* . (Issue Iv). Boston College website.
Nigg. (2006). *What Causes ADHD? Understanding What Goes Wrong and Why*. The Guilford Press.
Noordzij, Van Diepen, Caskey, & Jager. (2017). Relative risk versus absolute risk: One cannot be interpreted without the other. *Nephrology Dialysis Transplantation*, *32*(February), ii13–ii18. https://doi.org/10.1093/ndt/gfw465
Novartis. (2019). *Prescribing information for Ritalin LA*. https://www.novartis.us/sites/www.novartis.us/files/ritalin_la.pdf
Nylund, Asparouhov, & Muthén. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396
O'Brien. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6
Obsuth, Murray, Di Folco, Ribeaud, & Eisner. (2020). Patterns of Homotypic and Heterotypic Continuity Between ADHD Symptoms, Externalising and Internalising Problems from Age 7 to 15. *Journal of Abnormal Child Psychology*, *48*(2), 223–236. https://doi.org/10.1007/s10802-019-00592-9
Office for National Statistics. (2016). *Explore 50 years of international migration to and from the UK*. International Migration. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/articles/explore50yearsofinternationalmigrationtoandfromtheuk/2016-12-01
Office for National Statistics. (2019). *Measures of National Well-being Dashboard*. People, Population and Community. https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/measuresofnationalwellbeingdashboard/2018-04-25
Office for National Statistics, National Records of Scotland, & Northern Ireland Statistics and Research Agency. (2017). *2011 Census aggregate data*. UK Data Service. https://doi.org/http://dx.doi.org/10.5257/census/aggregate-2011-2
Oldehinkel, Ormel, Verhulst, & Nederhof. (2014). Childhood adversities and adolescent depression: A matter of both risk and resilience. *Development and Psychopathology*, *26*(4), 1067–1075. https://doi.org/10.1017/S0954579414000534
Olivier, May, & Bell. (2017). Relative effect sizes for measures of risk. *Communications in Statistics - Theory and Methods*, *46*(14), 6774–6781. https://doi.org/10.1080/03610926.2015.1134575
Ouyang, Fang, Mervy, Perou, & Grosse. (2008). Attention-Deficit/Hyperactivity Disorder Symptoms and Child Maltreatment: A Population-Based Study. *Journal of Pediatrics*, *December*, 851–856. https://doi.org/10.1016/j.jpeds.2008.06.002

Pan, & Bai. (2018). Propensity score methods for causal inference: an overview. *Behaviormetrika*, *0123456789*. https://doi.org/10.1007/s41237-018-0058-8

Panksepp. (2007). Can PLAY Diminish ADHD and Facilitate the Construction of the Social Brain ? *Journal of the Canandian Academy of Child and Adolescent Psychiatry*, *16*(2), 57–66.

Parsons. (2014). *Childhood cognition in the 1970 British Cohort Study* (Issue November).

Pelham, & Bender. (1982). Peer relationships in hyperactive children: Description and treatment. *Advances in Learning & Behavioral Disabilities*, *1*, 365–436.

Peng, Lee, & Ingersoll. (2002). An introduction to logistic regression analysis and reporting. *Journal of Educational Research*, *96*(1), 3–14. https://doi.org/10.1080/00220670209598786

Percy, & Iwaniec. (2008). *Antecedents of hazardous teenage drinking: analysis of the1970 British Cohort Study* (Issue August).

Pfiffner, & McBurnett. (1997). Social skills training with parent generalization: treatment effects for children with attention deficit disorder. *Journal of Consulting and Clinical Psychology*, *65*(5), 749–757. http://www.ncbi.nlm.nih.gov/pubmed/9337494

Pfiffner, Villodas, Kaiser, Rooney, & McBurnett. (2013). Educational outcomes of a collaborative school-home behavioral intervention for ADHD. *School Psychology Quarterly : The Official Journal of the Division of School Psychology, American Psychological Association*, *28*(1), 25–36. https://doi.org/10.1037/spq0000016

Pinquart. (2016). Associations of Parenting Styles and Dimensions with Academic Achievement in Children and Adolescents: A Meta-analysis. *Educational Psychology Review*, *28*(3), 475–493. https://doi.org/10.1007/s10648-015-9338-y

Polanczyk, de Lima, Horta, Biederman, & Rohde. (2007). The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *The American Journal of Psychiatry*, *164*(6), 942–948. https://doi.org/10.1176/appi.ajp.164.6.942

Polanczyk, Laranjeira, Zaleski, Pinksy, Caetano, & Rohde. (2010). ADHD in a representative sample of the Brazilian population: estimated prevalence and comparative adequacy of criteria between adolescents and adults according to the item response theory. *International Journal of Methods in Psychiatric Research*, *19*(3), 177–184. https://doi.org/10.1002/mpr.319

Posner. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. http://www.ncbi.nlm.nih.gov/pubmed/3378136

Prevoo, & ter Weel. (2015a). The Effect of Family Disruption on Children's Personality Development: Evidence from British Longitudinal Data. *De Economist*, *163*(1), 61–93. https://doi.org/10.1007/s10645-014-9248-2

Prevoo, & ter Weel. (2015b). The importance of early conscientiousness for socio-economic outcomes: Evidence from the british cohort study. *Oxford Economic Papers*, *67*(4), 918–948. https://doi.org/10.1093/oep/gpv022

Primi, Ferrão, & Almeida. (2010). Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math. *Learning and Individual Differences*, *20*(5), 446–451. https://doi.org/10.1016/j.lindif.2010.05.001

Qualtrics. (2017). *Qualtrics*. http://www.qualtrics.com

Rajeev, & Wahba. (2002). Propensity Score Matching Methods for Non-experimental Causal Studies. *The Review of Economics and Statistics*, *84*(1), 151–161. https://doi.org/10.2139/ssrn.138259

Reeve, & Fayers. (2005). Applying item response theory modelling for evaluating questionniare item and scale properties. In Fayers & Hayes (Eds.), *Assessing Quality of Life in Clnical Trials: Methods and Practice* (Second). Oxford University Press.

Reise, Steven P, & Revicki (Eds.). (2015). *Handbook of Item Response Theory Modelling*. Routledge.

Reise, Steven Paul, & Waller. (2009). Item Response Theory and Clinical Measurement. *Annual Review of Clinical Psychology*, *5*(1), 27–48. https://doi.org/10.1146/annurev.clinpsy.032408.153553

Reuben, Moffitt, Caspi, Belsky, Harrington, Schroeder, Hogan, Ramrakha, Poulton, & Danese. (2016). Lest we forget: comparing retrospective and prospective assessments of adverse childhood experiences in the prediction of adult health. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *57*(10), 1103–1112. https://doi.org/10.1111/jcpp.12621

Rodgers, Pickles, Power, Collishaw, & Maughan. (1999). Validity of the Malaise Inventory in

general population samples. *Social Psychiatry and Psychiatric Epidemiology*, *34*(6), 333–341. https://doi.org/10.1007/s001270050153
Rogers, & Tannock. (2018). Are Classrooms Meeting the Basic Psychological Needs of Children With ADHD Symptoms? A Self-Determination Theory Perspective. *Journal of Attention Disorders*, *22*(14), 1354–1360. https://doi.org/10.1177/1087054713508926
Rosenbaum, & Rubin. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, *70*(1), 41–55.
Rosenbaum, & Rubin. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician*, *39*(1), 33–38.
Rotter. (1966). Generalized expectiencies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, *80*(1).
Roy, Hechtman, Arnold, Swanson, Molina, Sibley, & Howard. (2017). Childhood predictors of adult functional outcomes in the Multimodal Treatment study of ADHD (MTA). *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(8), 1–17. https://doi.org/10.1016/j.jaac.2017.05.020.Childhood
Rubin. (1976). Inference and Missing Data. *Biometrika*, *63*(3), 581–592.
Rubin. (1979). Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies. *Journal of the American Statistical Association*, *74*(366), 318–328.
Rubin. (1986). Comment : Which Ifs Have Causal Answers. *Journal of the American Statistical Association*, *81*(396), 961–962.
Rubin, & Thomas. (2000). Combining Propensity Score Matching With Additional Adjustments for Prognostic Covariates. *Journal of the American Statistical Association*, *95*(450), 573–585. https://doi.org/10.1080/01621459.2000.10474233
Russell, A. E., Ford, & Russell. (2015). Socioeconomic Associations with ADHD: Findings from a Mediation Analysis. *PloS One*, *10*(6), e0128248. https://doi.org/10.1371/journal.pone.0128248
Russell, G., Ford, Rosenberg, & Kelly. (2014). The association of attention deficit hyperactivity disorder with socioeconomic disadvantage: Alternative explanations and evidence. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *55*(5), 436–445. https://doi.org/10.1111/jcpp.12170
Rutter. (1967). A children's behaviour questionnaire for completion by teachers: Preliminary findings. *Journal of Child Psychology and Psychiatry*, *8*, 1–11.
Rutter. (1987). Psychosocial Resilience and Protective Mechanisms. *American Journal of Orthopsychiatry*, *57*(3).
Rutter, Tizard, & Whitmore. (1970). *Education, health and behaviour*. Longmans.
Ryan, & Deci. (2001). On Happiness and Human Potentials: A Review of Research on Hedonic and Eudaimonic Well-Being. *Annual Review of Psychology*, *52*(1), 141–166. https://doi.org/10.1146/annurev.psych.52.1.141
Sabeti Lab at Harvard University, Broad Institute of MIT and Harvard, & Fathom Information Design. (2018). *Mirador* (1.5.1). Fathom Information Design. www.fathom.info
Sacker, & Cable. (2006). Do adolescent leisure-time physical activities foster health and well-being in adulthood? Evidence from two British birth cohorts. *European Journal of Public Health*, *16*(3), 331–335. https://doi.org/10.1093/eurpub/cki189
SAGE journals and APSARD. (2019). *Most Read Articles December 2019*. Journal of Attention Disorders.
Sagiv, Epstein, Bellinger, & Korrick. (2013). Pre- and Postnatal Risk Factors for ADHD in a Nonclinical Pediatric Population. *Journal of Attention Disorders*, *17*(1), 47–57. https://doi.org/10.1177/1087054711427563
Sagvolden, Johansen, Aase, & Russell. (2005). A dynamic developmental theory of attention-deficit/hyperactivity disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes. *The Behavioral and Brain Sciences*, *28*(3), 397–419; discussion 419-68. https://doi.org/10.1017/S0140525X05000075
Sainani. (2015). Dealing With Missing Data. *PM&R*, *7*(9), 990–994. https://doi.org/10.1016/j.pmrj.2015.07.011
Salas, Driskell, & Hughes. (1996). Introduction: The Study of Stress and Human Performance.

In *Stress and Human Performance*. Lawrence Erlbaum Associates Inc.
Salum, Sonuga-Barke, Sergeant, Vandekerckhove, Gadelha, Moriyama, Graeff-Martins, Manfro, Polanczyk, & Rohde. (2014). Mechanisms underpinning inattention and hyperactivity: Neurocognitive support for ADHD dimensionality. *Psychological Medicine*, *44*(15), 3189–3201. https://doi.org/10.1017/S0033291714000919
Sanders. (1983). Towards a model of stress and human performance. *Acta Psychologica*, *53*, 61–97. https://doi.org/10.1016/0001-6918(83)90016-1
Sanders. (1998). *Elements of Human Performance: Reaction Processes and Attention in Human Skill*. Laurence Erlbaum Associates. https://doi.org/10.4324/9780203774250
Sass, Schmitt, & Walker. (2008). Estimating non-normal latent trait distributions within Item Response Theory using true and estimated item parameters. *Applied Measurement in Education*, *21*(1), 67–89. https://doi.org/10.1080/08957340701796415
Sayal, Prasad, Daley, Ford, & Coghill. (2018). ADHD in children and young people: prevalence, care pathways, and service provision. *The Lancet Psychiatry*, *5*(2), 175–186. https://doi.org/10.1016/S2215-0366(17)30167-0
Schachter, Pham, King, Langford, & Moher. (2001). How efficacious and safe is short-acting methylphenidate for the treatment of attention-decisit disorder in children and adolescents? A meta-analysis. *Canadian Medical Association Journal*, *165*(11), 1475–1488.
Schatz, & Rostain. (2006). ADHD with comorbid anxiety. A review of the current literature. *Journal of Attention Disorders*, *10*(2), 141–149. https://doi.org/10.1177/1087054706286698
Scheff. (1999). *Being Mentally Ill: A Sociological Theory* (Third). AldineTransaction.
Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson. (2007). Estimating Causal Effects: Using Experimental and Observational Designs. In *American Educational Research Association*. https://doi.org/10.1017/CBO9781107415324.004
Schoon, & Kneale. (2013). Becoming adults in Britain: lifestyles and wellbeing in times of social change. *Longitudinal and Life Course Studies*, *3*(2), 173–189. https://doi.org/10.14301/llcs.v3i2.181
Schutt, & O'Neil. (2019). *Doing Data Science*. Safari Books Online; O'Reily. https://www.oreilly.com/library/view/doing-data-science/9781449363871/ch01.html
Schwarz. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464.
Sciberras, Mulraney, Silva, & Coghill. (2017). Prenatal Risk Factors and the Etiology of ADHD—Review of Existing Evidence. *Current Psychiatry Reports*, *19*(1). https://doi.org/10.1007/s11920-017-0753-2
Semrud-Clikeman, Walkowiak, Wilkinson, & Butcher. (2010). Executive functioning in children with Asperger syndrome, ADHD-combined type, ADHD-predominately inattentive type, and controls. *Journal of Autism and Developmental Disorders*, *40*(8), 1017–1027. https://doi.org/10.1007/s10803-010-0951-9
Sergeant. (2000). The cognitive-energetic model: an empirical approach to attention-deficit hyperactivity disorder. *Neuroscience and Biobehavioral Reviews*, *24*(1), 7–12. https://doi.org/S0149-7634(99)00060-3
Sergeant. (2004). Modeling Attention-Deficit / Hyperactivity Disorder : A Critical Appraisal of the Cognitive – Energetic Model. *Biological Basis of Individual Behaviour*, *57*, 1248–1255. https://doi.org/10.1016/j.bps.2004.09.010
Shamseer, Moher, Clarke, Ghersi, Liberati, Petticrew, Shekelle, & Stewart. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *Bmj*, *7647*(January), 1–25. https://doi.org/10.1136/bmj.g7647
Shaw, M., Hodgkins, Caci, Young, Kahle, Woods, & Arnold. (2012). A systematic review and analysis of long-term outcomes in attention deficit hyperactivity disorder : effects of treatment and non-treatment. *BMC Medicine*, *10*(99), 1–15. http://www.biomedcentral.com/1741-7015/10/99%0ARESEARCH
Shaw, P., Malek, Watson, Greenstein, De Rossi, & Sharp. (2013). Trajectories of cerebral cortical development in childhood and adolescence and adult attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *74*(8), 599–606. https://doi.org/10.1016/j.biopsych.2013.04.007
Shepherd, & Gilbert. (2019). *1970 British Cohort Study Ethical Review and consent* (Issue January). http://www.cls.ioe.ac.uk/

Sherer, Maddux, Mercandante, Prentice-Dunn, Jacobs, & Rogers. (1982). The Self-Efficacy Scale: Construction and Validation. *Psychological Reports*, *51*, 663–671.
Shifrer. (2013). Stigma of a Label: Educational Expectations for High School Students Labeled with Learning Disabilities. *Journal of Health and Social Behavior*, *54*(4), 462–480. https://doi.org/10.1177/0022146513503346
Sigle-Rushton, Hobcraft, & Kiernan. (2005). Parental Divorce and Subsequent Disadvantage: A Cross-Cohort Comparison. *Demography*, *42*(3), 427–446. https://doi.org/10.1353/dem.2005.0026
Silva, Colvin, Hagemann, & Bower. (2014). Environmental Risk Factors by Gender Associated With Attention-Deficit/Hyperactivity Disorder. *Pediatrics*, *133*(1), e14–e22. https://doi.org/10.1542/peds.2013-1434
Singh. (2012). *VOICES Study: Final Report*.
Sjöwall, Roth, Lindqvist, & Thorell. (2013). Multiple deficits in ADHD: Executive dysfunction, delay aversion, reaction time variability, and emotional deficits. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*, 619–627. https://doi.org/10.1111/jcpp.12006
Slopen, Kubzansky, McLaughlin, & Koenen. (2013). Childhood adversity and inflammatory processes in youth: A prospective study. *Psychoneuroendocrinology*, *38*(2), 188–200. https://doi.org/10.1016/j.psyneuen.2012.05.013
Smith, A. L., Hoza, Linnea, McQuade, Tomb, Vaughn, Shoulberg, & Hook. (2013). Pilot physical activity intervention reduces severity of ADHD symptoms in young children. *Journal of Attention Disorders*, *17*(1), 70–82. https://doi.org/10.1177/1087054711417395
Smith, O. R. F., Alves, Knapstad, Haug, & Aarø. (2017). Measuring mental well-being in Norway: Validation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS). *BMC Psychiatry*, *17*(1), 1–9. https://doi.org/10.1186/s12888-017-1343-x
Sonuga-Barke, Brandeis, Cortese, Daley, Ferrin, Holtmann, Stevenson, Danckaerts, van der Oord, Döpfner, Dittmann, Simonoff, Zuddas, Banaschewski, Buitelaar, Coghill, Hollis, Konofal, Lecendreux, … Sergeant. (2013). Nonpharmacological interventions for ADHD: systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. *The American Journal of Psychiatry*, *170*(3), 275–289. https://doi.org/10.1176/appi.ajp.2012.12070991
Sonuga-Barke, Kennedy, Kumsta, Knights, Golm, Rutter, Maughan, Schlotz, & Kreppner. (2017). Child-to-adult neurodevelopmental and mental health trajectories after early life deprivation: the young adult follow-up of the longitudinal English and Romanian Adoptees study. *The Lancet*, *389*(10078), 1539–1548. https://doi.org/10.1016/S0140-6736(17)30045-4
Sowell, Peterson, Thompson, Welcome, Henkenius, & Toga. (2003). Mapping cortical change across the human life span. *Nature Neuroscience*, *6*(3), 309–315. https://doi.org/10.1038/nn1008
Spencer, Faraone, Tarko, McDermott, & Biederman. (2014). Attention-deficit/hyperactivity disorder and adverse health outcomes in adults. *Journal of Nervous and Mental Disease*, *202*(10), 725–731. tspencer@partners.org,
Springer, & Hauser. (2006). An assessment of the construct validity of Ryff's Scales of Psychological Well-Being: Method, mode, and measurement effects. *Social Science Research*, *35*(4), 1080–1102. https://doi.org/10.1016/j.ssresearch.2005.07.004
Staffa, & Zurakowski. (2018). Five Steps to Successfully Implement and Evaluate Propensity Score Matching in Clinical Research Studies. *International Anasthesia Research Society*, 1–8. https://doi.org/10.1213/ANE.0000000000002787
Stansfeld, & Rasul. (2006). Psychosocial factors, depression and illness. In Steptoe (Ed.), *Depression and Physical Illness* (pp. 19–52). Cambridge University Press. https://doi.org/10.1017/CBO9780511544293.003
Stata Corp LP. (2013). *Stata Structural Equation Modeling Reference Manual* (13th ed.). Stata Press.
Stata Corp LP. (2015). *Stata Item Response Theory Reference Manual* (Release 14). Stata Press.
StataCorp LLC. (2015). *Stata Statistics/Data Analysis SE* (14.2).
Steptoe, & Butler. (1996). Sports participation and emotional wellbeing in adolescents. *The Lancet*, *347*, 1789–1792.

Stevens, Lauinger, & Neville. (2009). Differences in the neural mechanisms of selective attention in children from different socioeconomic backgrounds: An event-related brain potential study. *Developmental Science*, *12*(4), 634–646. https://doi.org/10.1111/j.1467-7687.2009.00807.x
Stone, Otten, Engels, Vermulst, & Janssens. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: a review. *Clinical Child and Family Psychology Review*, *13*(3), 254–274. https://doi.org/10.1007/s10567-010-0071-2
Storebø, O J, Ramstad, Krogh, Nilausen, Skoog, Holmskov, Rosendal, Groth, Magnusson, R, Gillies, K, Gauci, Zwi, Kirubakaran, Forsbøl, Simonsen, & Gluud. (2015). Methylphenidate for children and adolescents with attention deficit hyperactivity disorder (ADHD). In *Cochrane Database of Systematic Reviews* (Issue 11). https://doi.org/10.1002/14651858.CD009885.pub2.Copyright
Storebø, Ole Jakob, Gluud, Winkel, & Simonsen. (2012). Social-skills and parental training plus standard treatment versus standard treatment for children with ADHD--the randomised SOSTRA trial. *PloS One*, *7*(6), e37280. https://doi.org/10.1371/journal.pone.0037280
Storebo, Rosendal, Skoog, Groth, Bille, Rasmussen, Simonsen, & Gluud. (2012). Methylphenidate for attention deficit hyperactivity disorder (ADHD) in children and adolescents. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.cd009996
Stuart-Smith, Thapar, Maughan, Thapar, & Collishaw. (2017). Childhood hyperactivity and mood problems at mid-life: evidence from a prospective birth cohort. *Social Psychiatry and Psychiatric Epidemiology*, *52*(1), 87–94. https://doi.org/10.1007/s00127-016-1285-5
Stuart. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1–21. https://doi.org/10.1214/09-STS313.Matching
Sturm, Kuhfeld, Kasari, & Mccracken. (2017). Development and validation of an item response theory-based Social Responsiveness Scale short form. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *9*, 1053–1061. https://doi.org/10.1111/jcpp.12731
Substance Abuse and Mental Health Services Administration. (2016). *Table 7, DSM-IV to DSM-5 Attention-Deficit/Hyperactivity Disorder Comparison*. DSM-5 Changes: Implications for Child Serious Emotional Disturbance [Internet]. https://www.ncbi.nlm.nih.gov/books/NBK519712/table/ch3.t3/
Substance Abuse and Mental Health Services Administration (US). (2016). Table 3.30, DSM-IV to DSM-5 Acute Stress Disorder Comparison. In *Substance Abuse and Mental Health Services Administration. Impact of the DSM-IV to DSM-5 Changes on the National Survey on Drug Use and Health*. https://www.ncbi.nlm.nih.gov/books/NBK519704/table/ch3.t30/
Sutherland. (2016). *A Primer on Propensity Score Matching* (SSRMC Advanced Topics).
Svensson. (2001). Guidelines to Statistical Evaluation of Data From Scales and Questionnaires. *Journal of Rehabilitation Medicine*, *33*, 47–48. https://doi.org/10.1080/165019701300006542
Swanson, J. M., Arnold, Molina, Sibley, Hechtman, Hinshaw, Abikoff, Stehli, Owens, Mitchell, Nichols, Howard, Greenhill, Hoza, Newcorn, Jensen, Vitiello, Wigal, Epstein, … Stern. (2017). Young adult outcomes in the follow-up of the multimodal treatment study of attention-deficit/hyperactivity disorder: symptom persistence, source discrepancy, and height suppression. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *58*(6), 663–678. https://doi.org/10.1111/jcpp.12684
Swanson, J., Schuck, Mann Porter, Carlson, Hartman, Sergeant, Clevenger, Wasdell, McClearly, Lakes, & Wigal. (2012). Categorical and Dimensional Definitions and Evaluations of Symptoms of ADHD: History of the SNAP and the SWAN Rating Scales James. *International Journal of Educational and Psychological Assessment*, *10*(1), 51–70. https://doi.org/10.1038/nchembio.1527.A
Tang, Ma, Wang, Fan, Feng, Lu, Yu, Sui, Rothbart, Fan, & Posner. (2007). Short-term meditation training improves attention and self-regulation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(43), 17152–17156. https://doi.org/10.1073/pnas.0707678104

Taub, Floyd, Keith, & McGrew. (2008). Effects of General and Broad Cognitive Abilities on Mathematics Achievement. *School Psychology Quarterly*, *23*(2), 187–198. https://doi.org/10.1037/1045-3830.23.2.187
Tavakol, & Dennick. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd
Tennant, Hiller, Fishwick, Platt, Joseph, Weich, Parkinson, Secker, & Stewart-Brown. (2007). The Warwick-Edinburgh mental well-being scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes*, *5*, 1–13. https://doi.org/10.1186/1477-7525-5-63
Ter-Stepanian, Grizenko, Cornish, Talwar, Mbekou, Schmitz, & Joober. (2017). Attention and Executive Function in Children Diagnosed. *J Can Acad Child Adolesc Psychiatry*, *26*, 21–30.
Thapar, Cooper, Eyre, & Langley. (2013). Practitioner review: What have we learnt about the causes of ADHD? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(1), 3–16. https://doi.org/10.1111/j.1469-7610.2012.02611.x
Thapar, Rice, Hay, Boivin, Langley, Van Den Bree, Rutter, & Harold. (2009). Prenatal smoking might not cause attention-deficit/hyperactivity disorder: Evidence from a novel design. *Biological Psychiatry*, *66*(8), 722–727. https://doi.org/10.1016/j.biopsych.2009.05.032
Thomas, Sanders, Doust, Beller, & Glasziou. (2015). Prevalence of Attention-Deficit/Hyperactivity Disorder: A Systematic Review and Meta-analysis. *Pediatrics*, *135*(4), e994–e1001. https://doi.org/10.1542/peds.2014-3482
Thompson, & Thompson. (1998). Neurofeedback combined with training in metacognitive strategies: effectiveness in students with ADD. *Applied Psychophysiology and Biofeedback*, *23*(4), 243–263. http://www.ncbi.nlm.nih.gov/pubmed/10457815
Toland. (2013). Practical Guide to Conducting an Item Response Theory Analysis. In *The Journal of Early Adolescence* (Vol. 34, Issue 1). https://doi.org/10.1177/0272431613511332
Touchette, Petit, Séguin, Boivin, Tremblay, & Montplaisir. (2007). Associations between sleep duration patterns and behavioral/cognitive functioning at school entry. *Sleep*, *30*(9), 1213–1219. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1978413&tool=pmcentrez&rendertype=abstract
UCLA: Statistical Consulting Group. (n.d.). *stata2mplus*. https://stats.idre.ucla.edu/stata/ado/analysis/
Uebel, Albrecht, Asherson, Börger, Butler, Chen, Christiansen, Heise, Kuntsi, Schäfer, Andreou, Manor, Marco, Miranda, Mulligan, Oades, van der Meere, Faraone, Rothenberger, & Banaschewski. (2010). Performance variability, impulsivity errors and the impact of incentives as gender-independent endophenotypes for ADHD. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *51*(2), 210–218. https://doi.org/10.1111/j.1469-7610.2009.02139.x
Ullebø, Posserud, Heiervang, Gillberg, & Obel. (2011). Screening for the attention deficit hyperactivity disorder phenotype using the strength and difficulties questionnaire. *European Child and Adolescent Psychiatry*, *20*(123), 451–458. https://doi.org/10.1007/s00787-011-0198-9
University College London. (2019). *Centre for Longitudinal Studies*. Home of the 1958, 1970, Millennium and Next Steps Cohort Studies. http://www.cls.ioe.ac.uk
University of Bristol. (2015). *Cohort details and data available*. Avon Longitudinal Study of Parents and Children. http://www.bristol.ac.uk/alspac/researchers/resources-available/
University of Bristol, & National. (1980). *Child health and education study: Maternal self-completion form*.
University of Bristol, & National Birthday Trust. (1980). *Child health and education study: Educational questionnaire*.
University of Essex, University of Manchester, & Jisc. (2012). *UK Data Service*. Website. https://ukdataservice.ac.uk
Van Belle, Van Hulst, & Durston. (2015). Developmental differences in intra-individual variability in children with ADHD and ASD. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *56*(12), 1316–1326. https://doi.org/10.1111/jcpp.12417
van der Eijk. (2016a). *Economic and Social Research Council Doctoral Training Centre Research*

Design , Practice and Ethics Module code M14151.
van der Eijk. (2016b). *Structural Equation Models 2: beyond the helicopter view*.
Van Der Eijk, & Rose. (2015). Risky business: Factor analysis of survey data - Assessing the probability of incorrect dimensionalisation. *PLoS ONE*, *10*(3), 1–31. https://doi.org/10.1371/journal.pone.0118900
van Lieshout, Luman, Twisk, van Ewijk, Groenman, Thissen, Faraone, Heslenfeld, Hartman, Hoekstra, Franke, Buitelaar, Rommelse, & Oosterlaan. (2016). A 6-year follow-up of a large European cohort of children with attention-deficit/hyperactivity disorder-combined subtype: outcomes in late adolescence and young adulthood. *European Child and Adolescent Psychiatry*, *25*(9). https://doi.org/10.1007/s00787-016-0820-y
Vartanian. (2011). *Secondary Data Anlaysis*. Oxford University Press.
Vaughn, Beaver, Wexler, Delisi, & Roberts. (2011). The Effect of School Dropout on Verbal Ability in Adulthood: A Propensity Score Matching Approach. *Journal of Youth and Adolescence*, *40*, 197–206. https://doi.org/10.1007/s10964-009-9501-1
Veall, & Zimmerman. (1996). Pseudo-R2 Measures for Some Common Limited Dependent Variable Models. *Journal of Economic Surveys*, *10*(3), 241–259. https://doi.org/https://doi.org/10.1111/j.1467-6419.1996.tb00013.x
Vignoles. (2014). *Secondary Data Analysis*.
Vignoles, & Alcott. (2018). *Evaluation Methods and Economic Appraisal*.
von Bertalanffy. (1968). *General System Theory*. George Braziller. https://doi.org/10.1016/B978-0-444-52076-0.50006-7
von Stumm, Gale, Batty, & Deary. (2009). Childhood intelligence, locus of control and behaviour disturbance as determinants of intergenerational social mobility: British Cohort Study 1970. *Intelligence*, *37*(4), 329–340. https://doi.org/10.1016/j.intell.2009.04.002
Wagner, Martel, Cogo, Renato, Maia, Pan, Rohde, & Salum. (2016). Attention - deficit / hyperactivity disorder dimensionality : the reliable ' g ' and the elusive ' s ' dimensions. *European Child & Adolescent Psychiatry*, *25*, 83–90. https://doi.org/10.1007/s00787-015-0709-1
Wall, Park, & Moustaki. (2015). IRT modeling in the presence of zero-inflation with application to psychiatric disorder severity. *Applied Psychological Measurement*, *39*(8), 583–597. https://doi.org/10.1177/0146621615588184
Webb, Wilson, & Chong. (2004). An Analysis of Quasi-complete Binary Data with Logistic Models: Applications to Alcohol Abuse Data. *Journal of Data Science*, *2*(3), 273–285. https://doi.org/10.6339/JDS.2004.02(3).155
Wechsler. (2011). *Wechsler Abbreviated Scale of Intelligence - Second Edition (WASI-II)*. Pearson Assessment.
Welford. (1973). Stress and Performance. *Ergonomics*, *16*(5), 567–580. https://doi.org/10.1080/00140137308924547
Werry, Sprague, & Cohen. (1975). Conners' Teacher Rating Scale for use in drug studies with children - an empirical study. *Journal of Abnormal Child Psychology*, *3*(3), 217–229. c:%5Cpdfs%5CRM xc666.pdf
Wewers, & Lowe. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing & Health*, *13*(4), 227–236. https://doi.org/10.1002/nur.4770130405
What Works Centre for Wellbeing. (2019). *What is wellbeing?* What Works Wellbeing. https://whatworkswellbeing.org/about/what-is-wellbeing/#
What Works Wellbeing. (2019). *Credible evidence for better decisions, to improve lives*. Website. https://whatworkswellbeing.org
White, Nicholls, Christie, Cole, & Viner. (2012). Childhood psychological function and obesity risk across the lifecourse: Findings from the 1970 British Cohort Study. *International Journal of Obesity*, *36*(4), 511–516. https://doi.org/10.1038/ijo.2011.253
Whitebread. (2017). Free play and children's mental health. *The Lancet Child and Adolescent Health*, *1*(3), 167–169. https://doi.org/10.1016/S2352-4642(17)30092-5
Wilkes-Gillan, Bundy, Cordier, Lincoln, & Chen. (2016). A randomised controlled trial of a play-based intervention to improve the social play skills of children with attention deficit hyperactivity disorder (ADHD). *PLoS ONE*, *11*(8), 1–22.

https://doi.org/10.1371/journal.pone.0160558
Willcutt. (2012). The Prevalence of DSM-IV Attention-Deficit/Hyperactivity Disorder: A Meta-Analytic Review. *Neurotherapeutics*, *9*(3), 490–499. https://doi.org/10.1007/s13311-012-0135-8
Willcutt, Doyle, Nigg, Faraone, & Pennington. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry*, *57*(11), 1336–1346. https://doi.org/10.1016/j.biopsych.2005.02.006
Willcutt, Nigg, Pennington, Solanto, Rohde, Tannock, Carlson, & Lahey. (2012). Validity of DSM-IV attention-deficit/hyperactivity disorder symptom dimensions and subtypes. *Journal of Abnormal Psychology*, *121*(4), 991–1010. https://doi.org/10.1037/a0027347.Validity
Williams. (2015a). *Heteroskedasticity*.
Williams. (2015b). Review of Alan Acock's Discovering Structural Equation Modeling Using Stata, Revised Edition. *Stata Journal*, *15*(1), 309–315. https://doi.org/The Stata Journal
Williamson, Birmaher, Anderson, Al-Shabbout, & Ryan. (1995). Stressful Life Events in Depressed Adolescents: The Role of Dependent Events during the Depressive Episode. *Journal of the American Academy of Child & Adolescent Psychiatry*, *34*(5), 591–598. https://doi.org/10.1097/00004583-199505000-00011
Williamson, Birmaher, Ryan, Shiffrin, Lusky, Protopapa, Dahl, & Brent. (2003). The Stressful Life Events Schedule for children and adolescents: Development and validation. *Psychiatry Research*, *119*(3), 225–241. https://doi.org/10.1016/S0165-1781(03)00134-3
Windmeijer. (1995). Goodness-of-fit measures in binary choice models. *Econometric Reviews*, *14*(1), 101–116. https://doi.org/10.1080/07474939508800306
Winship, & Morgan. (1999). The Estimation of Causal Effects From Observational Data. *Annual Review of Sociology*, *25*(1), 659–706. https://doi.org/doi:10.1146/annurev.soc.25.1.659
Witten, Frank, Hall, & Pal. (2017). *Data Mining: Practical Machine Learning Tools and Techniques* (Fourth). Elsevier. https://books.google.co.uk/books?hl=en&lr=&id=1SylCgAAQBAJ&oi=fnd&pg=PP1&dq=data+mining+approach&ots=8IALqiiAua&sig=ZBdxa-_QuQ-3_bz7oMdeno9EJwo#v=onepage&q=data mining approach&f=false
Wolf, A. (2011). The Wolf Report. In *Review of Vocational Education* (Issue March). https://doi.org/10.1037/h0066788
Wolf, E. J., Harrington, Clark, & Miller. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement*, *73*(6), 913–934. https://doi.org/10.1177/0013164413495237
Wood, Bann, Hardy, Gale, Goodman, Crawford, & Stafford. (2017). Childhood socioeconomic position and adult mental wellbeing: Evidence from four British birth cohort studies. *PLoS ONE*, *12*(10), 1–13. https://doi.org/10.1371/journal.pone.0185798
Woods-Giscombé, Lobel, & Crandell. (2010). The impact of miscarriage and parity on patterns of maternal distress in pregnancy. *Research in Nursing & Health*, *33*(4), 316–328. https://doi.org/10.1002/nur.20389
Woods. (2015). Estimating the Latent Density in Unidimensional IRT to Permit Non-Normality. In Steven P Reise & Revicki, Dennis (Eds.), *Handbook of Item Response Theory Modelling* (pp. 60–84). Routledge.
World Federation of ADHD. (2015). 5th World Congress on ADHD: From child to adult disorder. *Final Programme*.
World Federation of ADHD. (2017). 6th World Congress on ADHD. *From Child to Adult Disorder*.
World Federation of ADHD. (2019). 6th World Congress on ADHD. *From Child to Adult Disorder*.
World Health Organization. (1994). *Behavioural and emotional disorders with onset usually occurring in childhood and adolescence*. International Classification of Diseases (ICD-10). https://apps.who.int/classifications/apps/icd/icd10online2003/fr-icd.htm?gf90.htm+
Worthington, & Whittaker. (2006). Scale Development Research: A Content Analysis and Recommendations for Best Practices. *The Counseling Psychologist*, *34*(6), 806–838. https://doi.org/10.1177/0011000006288127

Wunder, Wiencierz, Schwarze, & Küchenhoff. (2009). Well-being over the life span: Semiparametric evidence from british and german longitudinal data. In *IZA Discussion Papers* (Vol. 4155). https://doi.org/10.1162/REST_a_00222

Wüstner, Otto, Schlack, Hölling, Klasen, & Ravens-Sieberer. (2019). Risk and protective factors for the development of ADHD symptoms in children and adolescents: Results of the longitudinal BELLA study. *PLoS ONE*, *14*(3), 1–19. https://doi.org/10.1371/journal.pone.0214412

Yang, & Zheng. (2018). Item Response Data Analysis Using Stata Item Response Theory Package. *Journal of Educational and Behavioral Statistics*, *43*(1), 116–129. https://doi.org/10.3102/1076998617749186

Young. (2012). *Don't know responses in survey research* (Issue May). The Pennsylvania State University.

youthinmind. (2012). *SDQ: Information for Researchers and Professionals about the Strengths & Difficulties Questionnaire*. http://www.sdqinfo.com

youthinmind. (2014a). *SDQ Norms (Britain)*. Sdqinfo.Org. http://www.sdqinfo.org/UKNorms.html

youthinmind. (2014b). *SDQ Scoring Site*. http://www.sdqscore.org/Amber

Zimmer-Gembeck, & Skinner. (2016). The Development of Coping: Implications for Psychopathology and Resilience. *Developmental Psychopathology*, 1–61. https://doi.org/10.1002/9781119125556.devpsy410

Zito. (2018). Children as Saviors? A Propensity Score Analysis of the Impact of Teenage Motherhood on Personal Transformation. *Youth and Society*, *50*(8), 1100–1122. https://doi.org/10.1177/0044118X16653872

Zolkoski, & Bullock. (2012). Children and Youth Services Review Resilience in children and youth : A review. *Children and Youth Services Review*, *34*(12), 2295–2303. http://dx.doi.org/10.1016/j.childyouth.2012.08.009

Chapter 1 appendix

A. DSM-5 ADHD criteria

A summary of the DSM-5 ADHD diagnostic criteria has been extracted from the Centers for Disease Control website, as follows:

> "People with ADHD show a persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development:
>
> Inattention: Six or more symptoms of inattention for children up to age 16, or five or more for adolescents 17 and older and adults; symptoms of inattention have been present for at least 6 months, and they are inappropriate for developmental level:
>
> 1. Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or with other activities.
> 2. Often has trouble holding attention on tasks or play activities.
> 3. Often does not seem to listen when spoken to directly.
> 4. Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., loses focus, side-tracked).
> 5. Often has trouble organizing tasks and activities.
> 6. Often avoids, dislikes, or is reluctant to do tasks that require mental effort over a long period of time (such as schoolwork or homework).
> 7. Often loses things necessary for tasks and activities (e.g. school materials, pencils, books, tools, wallets, keys, paperwork, eyeglasses, mobile telephones).
> 8. Is often easily distracted
> 9. Is often forgetful in daily activities.
>
> Hyperactivity and Impulsivity: Six or more symptoms of hyperactivity-impulsivity for children up to age 16, or five or more for adolescents 17 and older and adults; symptoms of hyperactivity-impulsivity have been present for at least 6 months to an extent that is disruptive and inappropriate for the person's developmental level:
>
> 1. Often fidgets with or taps hands or feet, or squirms in seat.
> 2. Often leaves seat in situations when remaining seated is expected.
> 3. Often runs about or climbs in situations where it is not appropriate (adolescents or adults may be limited to feeling restless).
> 4. Often unable to play or take part in leisure activities quietly.
> 5. Is often "on the go" acting as if "driven by a motor".
> 6. Often talks excessively.
> 7. Often blurts out an answer before a question has been completed.
> 8. Often has trouble waiting his/her turn.
> 9. Often interrupts or intrudes on others (e.g., butts into conversations or games)
>
> In addition, the following conditions must be met:
>
> 1. Several inattentive or hyperactive-impulsive symptoms were present before age 12 years.

2. Several symptoms are present in two or more settings, (such as at home, school or work; with friends or relatives; in other activities).
3. There is clear evidence that the symptoms interfere with, or reduce the quality of, social, school, or work functioning.
4. The symptoms are not better explained by another mental disorder (such as a mood disorder, anxiety disorder, dissociative disorder, or a personality disorder).
5. The symptoms do not happen only during the course of schizophrenia or another psychotic disorder."
6. Symptoms must persist for at least six months.

(Centers for Disease Control and Prevention, 2018c)

N.B. Numbering and formatting added for readability.

Chapter 3 appendix

A. Control variables used in previous analysis of ADHD outcomes in BCS70 (Brassett-Grundy & Butler, 2004)

Age 0	Age 5
1. Birthweight	1. Child guidance
2. Breast fed	2. Headaches
3. Mother's marital status	3. Stomach aches
4. Father present	4. Bilious
5. Siblings	5. Temper tantrums
6. Ethnic group	6. Sleeping difficulty
7. Mother's age group at birth (<18/not)	7. Wets – day
8. Father's age group at birth (<18/not)	8. Wets – night
9. Mother's region of origin (geographic)	9. Soils [pants]
10. Mother's occupation class	10. Eating difficulty
11. Father's occupation class	11. Rutter score group
12. Mother's qualifications	12. Father left
13. Father's qualifications	13. Father occupation class drop
14. Pregnancy complications	14. Mother occupation class drop
	15. Younger siblings
	16. Housing tenure
	17. Housing density
	18. House moves
	19. Lost parent
	20. Mother depressed/malaise
	21. Authoritarian parent
	22. Special needs
	23. In care
	24. Parent rating of hyperactivity

Chapter 4 appendix

A. Expert panel survey details (Appendix A)

Survey instructions and first question:

Screen 1

UNIVERSITY OF CAMBRIDGE

Background
I am developing a new method to identify Attention Deficit Hyperactivity Disorder (ADHD) symptomatology in existing longitudinal birth cohort data, to enable further study of long-term outcomes. The method relies on a mapping of DSM-5 (American Psychiatric Association, 2013) ADHD criteria to questionnaire items. In order to evaluate validity of this mapping, I am asking contacts with expertise in ADHD or related subjects to provide an independent assessment.

Please treat the content as confidential. If you would like to forward the survey link to someone else who would be well-suited to answering the questions, that would be fantastic, but please inform me you are doing so via email.

The survey should take between 5 and 10 minutes to complete.

Screen 2

UNIVERSITY OF CAMBRIDGE

Instructions
The questions below link DSM-5 ADHD criteria (wording slightly modified, see notes at end of survey) to questionnaire items designed for parent or teacher response regarding a 10-year old child (University College London, 2017). The items are mostly derived from Rutter (1967) and Conners (1969). There are five inattentive questions, nine hyperactive-impulsive, two conditional, three comments, and five questions about you and your role; 24 items in total. Please complete your response **by 30 September, 2017.** Responses are required for all statements; comments are optional.

Use the rating scale to indicate ***how well*** you think each statement ***captures the same meaning*** as the (slightly modified) DSM-5 criterion. Where there is more than one, evaluate each statement separately.

Screen 3

Inattentive items:

Has trouble holding attention to tasks or play activities (DSM-5 I2)
*(**How well** does each statement below, evaluated separately, **capture the same meaning**...)*

	Extremely well	Very well	Moderately well	Slightly well	Not well at all
Does not pay attention to what is being explained in class	○	○	○	○	○
Becomes bored during class	○	○	○	○	○
Cannot concentrate on any particular task, even though the child may return to it frequently	○	○	○	○	○
Child concentrates poorly on educational tasks, in comparison with the average 10 year old	○	○	○	○	○

Example of survey results:

Fidgets with or taps hands or feet, or squirms in seat (DSM-5 H1)

#	Question	Extremely well		Very well		Moderately well		Slightly well		Not well at all		Total
1	Is squirmy or fidgety	42.86%	6	50.00%	7	7.14%	1	0.00%	0	0.00%	0	14
2	Given to rhythmic tapping or kicking	0.00%	0	14.29%	2	57.14%	8	14.29%	2	14.29%	2	14
3	Fidgeting and indulging in minor distracting activities	14.29%	2	42.86%	6	42.86%	6	0.00%	0	0.00%	0	14

Four of the initially mapped BCS70 items were removed as a result of the review: di2/j138-bored during class, di4/j087-persevere with difficult tasks, di6/m241-sits still and concentrates more than 5 minutes, and di6/ j143-confused/hesitant with complex task (University of Bristol & National, 1980; University of Bristol & National Birthday Trust, 1980).

Those four items were clearly indicated by a majority of the experts as mapping 'not well at all'. Coincidentally, all four were somewhat redundant, as there were other BCS70 items that did map well to the relevant DSM-5 criteria. Two further items: dh7/m73-impulsive excitable, and dc3/moderate or severe behavior problems on the Rutter scale, did not have a clear majority of opinion from the experts, but mixed views. These two were the only candidate items from BCS70 that could map to the two relevant DSM-5 criteria, so we decided to keep them and adhere as closely to the DSM-5 as possible.

B. Mplus code used for 2-class zero-inflated mixture model (Appendix B)

Derived from Wall et al., (2015), adapted with advice from Jung Yeon Park

```
TITLE:
ZI Mixture IRT (2 latent classes) based on derived scale of 16 DSM-5 criteria mapped to BCS70 Age 10 behaviour data;

DATA:
FILE IS <<filename.dat>>

VARIABLE:
NAMES = rowid dh1 dh2 dh3 dh4 dh5 dh6 dh7 dh8 dh9 di2 di4 di6 di8 di9 dc1 dc2;
IDVARIABLE IS rowid;
USEVARIABLES = dh1 dh2 dh3 dh4 dh5 dh6 dh7 dh8 dh9 di2 di4 di6 di8 di9 dc1 dc2;
CATEGORICAL = dh1 dh2 dh3 dh4 dh5 dh6 dh7 dh8 dh9 di2 di4 di6 di8 di9 dc1 dc2;
MISSING = ALL(999);
CLASSES = c (2);

ANALYSIS:
ESTIMATOR = MLR;
TYPE = MIXTURE;
ALGORITHM=INTEGRATION ODLL;
! the algorithm = odll is needed because of the model constraint command
STARTS 400 50;
STSEED 170056;
PROCESS = 6 (STARTS);

MODEL:
%OVERALL%
f BY dh1* dh2 dh3 dh4 dh5 dh6 dh7 dh8 dh9 di2 di4 di6 (lam1-lam12)
 di8 di9 dc1 dc2 (lam13-lam16);

[dh1$1* dh2$1 dh3$1 dh4$1 dh5$1 dh6$1 dh7$1 dh8$1 dh9$1] (tau1-tau9)
[di2$1 di4$1 di6$1 di8$1 di9$1 dc1$1 dc2$1] (tau10-tau16);
```

```
[c#1] (logitp1);

%c#1%
f*  (phi1);
[f*] (m1);

%c#2%
f*  (phi2);
[f*] (m2);

MODEL CONSTRAINT:
new(b1,b2,b3,b4,b5,b6,b7,b8,b9,b10,b11,b12,b13,b14,b15,b16);

m1= -100;
phi1= 0.0001;

m2= 0;
phi2= 1;

b1 = tau1/lam1;
b2 = tau2/lam2;
b3 = tau3/lam3;
b4 = tau4/lam4;
b5 = tau5/lam5;
b6 = tau6/lam6;
b7 = tau7/lam7;
b8 = tau8/lam8;
b9 = tau9/lam9;
b10 = tau10/lam10;
b11 = tau11/lam11;
b12 = tau12/lam12;
b13 = tau13/lam13;
b14 = tau14/lam14;
b15 = tau15/lam15;
```

```
b16 = tau16/lam16;

OUTPUT: TECH1 TECH8;
! plots for ICC curves
PLOT: TYPE = PLOT3;

SAVEDATA: FILE IS <<filename.dat>>;
	SAVE = FSCORES;
	SAVE = CPROB;
```

C. Note on sharing of derived categorical and dimensional measures data (Appendix C)

Other researchers may wish to use the categorical and dimensional ADHD indicators we derived without replicating the entire analysis. The indicators would not be useful without the related identifier (BCSID) to allow linking with other variables in the BCS70 datasets. The BCSID is owned by the Centre for Longitudinal Studies (CLS). They encourage sharing of derived data, and we will share our variables with them, which they may share more widely at their discretion. In the meantime, requests for access to our data will be coordinated between us and the CLS on a case-by-case basis.

Chapter 5 appendix

A. Pilot study propensity score matching procedure

The dataset used was SEMData_Short4_SI2.dta, which includes single mean (by ADHD subgroup) imputation values for all predictor variables, and dummy variables to indicate observations that were imputed. The procedure used was based on guidance from a taught course on PSM (Sutherland, 2016). Propensity score matching was conducted using both native Stata commands and the user-written PSMATCH2 (Leuven & Sianesi, 2018) software package.

First, a logistic regression model was built using membership in the ADHD subgroup as the dependent variable and all the predictors identified in the previous section, as well as their missingness indicators. Imputed mean values were included within the predictor variables. Stata code was as follows:

```
logistic adhd_sg sex10b F5DadEd F5HomeLS F5MumMal F0SmokeLevel
F0Unmarried F5HVOCIQ F5DE_m F5HLS_m F5MM_m F5SL_m F0UM_m
F5HI_m
```

The model was significant ($N = 7{,}242$, χ^2 (13) = 282.42, $p < 0.001$, McFadden's $R^2 = 0.11$). Each actual individual predictor was significant at the 99% confidence level, except F5HomeLS ($p = 0.009$) and F0Unmarried ($p = 0.063$).

N.B. Dataset included single-mean imputed values; missingness indicators were not significant.

Next, a postestimation command was used to predict a propensity score (probability of membership in the ADHD subgroup) for each observation. The result was stored in the variable 'pscore'.

```
predict pscore
```

A caliper was calculated based on findings that a control observation should not be selected unless it is within 1/5th of a standard deviation of the propensity score for a corresponding treatment group observation (Austin, 2011).

```
sum pscore, d
dis "caliper " = r(sd)/5
[displayed result: .00960962]
```

The two (2) nearest-neighbours were matched, based on guidance from simulation study findings (Austin, 2010).

```
psmatch2 adhd_sg, out(HWB_FS) pscore(pscore) n(2) caliper
.00960962)
```

The matched dataset was tested for differences in distributions on each predictor variable using 'pstest', with the option 'both' to show statistics for matched and unmatched samples.

```
pstest sex10b F5DadEd F5HomeLS F5MumMal F0SmokeLevel
F0Unmarried F5HVOCIQ, both
```

The pstest procedure indicated that the matched sample using these covariates and matching estimators did not effectively reduce bias. Matching should reduce bias by 90% (Rosenbaum & Rubin, 1985), and variance ratios should be as close as possible to 1.0 (Garrido et al., 2014). The bias reduction was acceptable for sex10b (98.9), F5DadEd (95.8), and F5MumMal (91.7). The other bias reductions were:
F5HomeLS: 18.7%, F0SmokeLevel: 82.4%, F0Unmarried: -77.8, and F5HVOCIQ: 75.4%

The bias reduction was negative for F0Unamarried, so actually bias was increased. These results indicate the matching approach should be tweaked to improve balance before reporting an estimated treatment effect.

Further variations were tested for different numbers of neighbours (1-4), a wider caliper (1/4 SD), using the common option (to drop propensity score outliers), and using ties. None of these produced the desired 90% reduction in bias for all the covariates. The F5HomeLS variable was particularly problematic, with low reductions in bias for all options. Another two options for matching estimators include kernel and mahalanobis distance. These were tested next:
Kernel matching (with default options)

```
psmatch2 adhd_sg, kernel out(HWB_FS) pscore(pscore)
```

The kernel match performed better, with bias reductions ranging from 63-84%, and variance ratios between 0.72 and 1.10.

Mahalanobis distance matching (default options)

```
psmatch2 adhd_sg, mahalanobis(sex10b F5DadEd F5HomeLS F5MumMal
F5HVOCIQ F0Unmarried F0SmokeLevel) outcome(HWB_FS) ate
```

Although it did still not reach the desired 90% bias reduction for all variables, Mahalanobis produced the best bias reduction, ranging from 80-100% and variance ratios between 0.94 and 1.0. The difference between the treated group and controls per the average treatment effect on the treated (ATT) using this matched sample was -0.222, nearly half the size of the -0.404 difference for the unmatched sample. ATT has been selected here as the more relevant measure, because in this context it is the average effect of membership in the ADHD subgroup, for an ADHD population (Abadie & Imbens, 2006). I am less interested in the overall population effect (average treatment effect, or ATE) for people who are not classifiable as ADHD.

The standard errors reported in PSMATCH2 do not take into account that the propensity score was estimated (pscore was estimated based on the logistic model, see above). Therefore it is recommended that an additional method is used to calculate robust standard errors (Abadie & Imbens, 2016). This can be done for some PSM matching estimators using the native Stata 14 teffects commands, but those commands do not support Mahalanobis distance estimators.

The teffects commands do support other estimators, such as inverse probability weights (IPW). Following an example provided by Stata, matching using our variables and teffects with an IPW estimator is shown here:

```
teffects ipw (HWB_FS) (adhd_sg sex10b F5DadEd F5HomeLS
F5MumMal F0SmokeLevel F0Unmarried F5HVOCIQ), atet
```

Output includes standard errors that are adjusted for the estimated propensity score (*ATT* = -0.282, *SE* = .074, z = -3.84, $p < 0.001$, 95% *CI*: [-0.427 to -0.137]). Although the output does not report percentage variance reduction, it does provide a variance ratio. The ratio ranged from 0.76 – 1.0, with F5HVOIQ, F5HomeLS, and F5MumMal all < 0.90.

Given I was unable to achieve an acceptable balance, and the strong arguments in literature against using PSM particularly with categorical data (all of my variables are categorical) (Blackwell et al., 2009; Iacus et al., 2011; King & Nielsen, 2018), I decided at this point not to continue with PSM and proceed only with CEM.

Chapter 6 appendix

A. Mplus file naming convention

Since numerous (200+) Mplus files were created and generated by these analyses, a file-naming convention was defined to help keep them organised. The convention is as follows:

```
WB/ED + B/M + B/G/0/1 + Sn + U/W/Em/A + n
```

Where:

WB = wellbeing / ED = education level +
B = Bayes / M = MLR +
B = both girls and boys / G = grouped (by sex) / 0= girls only / 1 = boys only +
Sn = step number +
U = unmatched full sample / W = weighted matched sample / Em = exact matched sample / A = ADHD sample +
N (1:n) = numbered simple regressions for each X within Sn model (optional)

For example, for the first analysis (step 1) on the wellbeing outcome, where MLR was the estimator, output was grouped by sex, and the unmatched full sample was used, the input filename was:

```
WBMGS1U.inp
```

For the same model with a Bayes estimator, the filename was:

```
WBBGS1U.inp
```

Mplus automatically creates output files with the same root name and a .out extension. The same input file (MOAF2.dat, for 'Mplus output analysis file, second version') was used for all regressions (see variable list above).

B. Correlations of core variables

	adhd	adhdsev	maths	lowses	zplay	locatheta	setheta	zext	cstress	estress
adhd	1.000									
adhdsev	**0.456**	1.000								
maths	-0.178	**-0.395**	1.000							
lowses	0.069	0.141	-0.161	1.000						
zplay	-0.050	-0.084	0.161	-0.102	1.000					
locatheta	-0.106	-0.255	**0.426**	-0.102	0.127	1.000				
setheta	-0.099	-0.230	0.246	-0.103	0.058	**0.372**	1.000			
zext	**0.360**	**0.474**	-0.193	0.152	-0.107	-0.143	-0.153	1.000		
cstress	0.219	**0.357**	**-0.344**	0.124	-0.148	-0.201	-0.178	0.229	1.000	
estress	0.085	0.162	-0.100	**0.364**	-0.056	-0.067	-0.092	0.170	0.144	1.000

r>.30 in bold

All correlations significant at p < .001

C. Correlations of all variables with significance in models with good fit

	adhd	adhdsev	maths	lowses	epvt	mummal	homelow	daded	wheez	backward	ecz	sepfmum	poornbhd	smokelvl	smokelvl	unmar	zext	zint	cstress	estress	loctheta	setheta	zplay
adhd	1.000																						
adhdsev	0.456	1.000																					
maths	-0.178	-0.395	1.000																				
lowses	0.069	0.141	-0.161	1.000																			
epvt	-0.083	-0.181	0.365	-0.177	1.000																		
mummal	0.107	0.174	-0.137	0.128	-0.124	1.000																	
homelow	0.072	0.120	-0.145	0.185	-0.152	0.086	1.000																
daded	-0.085	-0.190	0.313	-0.147	0.260	-0.139	-0.081	1.000															
wheez	0.049	0.064	-0.015	0.031	-0.009	0.052	0.021	-0.016	1.000														
backward	0.113	0.136	-0.199	0.121	-0.212	0.057	0.202	-0.080	0.021	1.000													
ecz	0.011	0.019	0.029	-0.021	0.043	0.009	-0.024	0.083	0.087	0.015	1.000												
sepfmum	0.035	0.053	-0.041	0.072	-0.087	0.077	0.054	-0.002	0.026	0.026	0.012	1.000											
poornbhd	0.022	0.065	-0.103	0.105	-0.127	0.067	0.205	-0.103	-0.013	0.068	-0.009	0.042	1.000										
smokelvl	0.060	0.135	-0.141	0.112	-0.109	0.112	0.098	-0.170	0.048	0.073	-0.039	0.052	0.064	1.000									
smokelvl	0.048	0.117	-0.130	0.100	-0.099	0.099	0.085	-0.155	0.045	0.060	-0.043	0.046	0.051	0.925	1.000								
unmar	0.036	0.075	-0.060	0.108	-0.078	0.036	0.072	-0.061	0.027	0.026	-0.014	0.087	0.080	0.092	0.086	1.000							
zext	0.360	0.474	-0.193	0.152	-0.131	0.173	0.158	-0.132	0.052	0.096	0.011	0.062	0.082	0.114	0.099	0.079	1.000						
zint	0.142	0.224	-0.093	0.036	-0.029	0.162	0.015	-0.034	0.049	0.043	0.020	0.030	0.012	0.015	0.014	0.018	0.254	1.000					
cstress	0.219	0.357	-0.344	0.124	-0.184	0.120	0.130	-0.084	0.084	0.215	0.037	0.045	0.054	0.087	0.076	0.040	0.229	0.198	1.000				
estress	0.085	0.162	-0.100	0.364	-0.144	0.125	0.122	-0.027	0.043	0.093	0.000	0.083	0.088	0.096	0.092	0.115	0.170	0.088	0.144	1.000			
loctheta	-0.106	-0.255	0.426	-0.102	0.265	-0.098	-0.093	0.222	0.005	-0.101	0.041	-0.027	-0.082	-0.080	-0.075	-0.036	-0.143	-0.084	-0.201	-0.067	1.000		
setheta	-0.099	-0.230	0.246	-0.103	0.153	-0.095	-0.073	0.136	-0.021	-0.080	0.027	-0.037	-0.045	-0.090	-0.082	-0.043	-0.153	-0.105	-0.178	-0.092	0.372	1.000	
zplay	-0.050	-0.084	0.161	-0.102	0.173	-0.079	-0.100	0.120	-0.006	-0.094	0.017	-0.035	-0.085	-0.049	-0.037	-0.026	-0.107	-0.083	-0.148	-0.056	0.127	0.058	1.000

r > 0.10
r > 0.20
r > 0.30

Chapter 7 appendix

A. SWB model robustness checks using Bayes estimator on unmatched sample

Model 1

Girls	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	-0.317	0.091	< 0.001	-0.496	-0.144	*
R2	0.003			N	4,387	
PPP[74]	0.539			BIC	12159.738	

Boys	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	-0.281	0.064	< 0.001	-0.408	-0.156	*
R2	0.005			N	4,132	
PPP	0.549			BIC	11135.493	

Model 2

Girls	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	0.109	0.096	0.134	-0.064	0.299	
ADHDSEV	-0.200	0.020	< 0.001	-0.237	-0.161	*
R2	0.027			N	4,387	
PPP	0.441			BIC	12063.080	

Boys	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	-0.074	0.071	0.160	-0.198	0.068	
ADHDSEV	-0.113	0.018	< 0.001	-0.147	-0.077	*
R2	0.015			N	4,132	
PPP	0.480			BIC	11103.946	

Model 3

[74] PPP for a BAYES model in Mplus is an index based on the chi-square value. It is more powerful with larger samples, the threshold for acceptability is PPP > 0.05. 0.50 is an example of excellent fit, and overall, higher is better (B. O. Muthén & Schultzberg, 2017; L. K. Muthén & Muthén, 2017).

Girls	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	0.136	0.097	0.084	-0.062	0.330	
ADHDSEV	-0.161	0.020	< 0.001	-0.198	-0.122	*
BACKWARD	0.032	0.101	0.367	-0.168	0.222	
DADED	0.035	0.009	< 0.001	0.018	0.054	*
HOMELOW	-0.403	0.096	< 0.001	-0.592	-0.209	*
MUMMAL	-0.124	0.038	< 0.001	-0.201	-0.050	*
SMOKE	0.160	0.079	0.024	0.002	0.315	*
SMOKELVL	-0.098	0.034	0.001	-0.165	-0.031	*
WHEEZ	-0.057	0.040	0.072	-0.134	0.021	
R2	0.038			N	4,387	
PPP	0.000			BIC	37672.866	

Boys	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	-0.063	0.070	0.180	-0.202	0.075	
ADHDSEV	-0.095	0.017	< 0.001	-0.128	-0.061	*
BACKWARD	-0.101	0.087	0.116	-0.280	0.052	
DADED	0.023	0.010	0.008	0.004	0.041	*
HOMELOW	0.004	0.092	0.481	-0.177	0.179	
MUMMAL	-0.074	0.040	0.039	-0.156	0.006	
SMOKE	0.009	0.079	0.455	-0.146	0.170	
SMOKELVL	-0.033	0.035	0.162	-0.106	0.034	
WHEEZ	0.000	0.038	0.493	-0.073	0.076	
R2	0.019			N	4,132	
PPP	0.000			BIC	37203.927	

Model 4

Girls	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	0.182	0.097	0.032	-0.005	0.368	
ADHDSEV	-0.069	0.023	0.003	-0.114	-0.023	*
BACKWARD	0.144	0.105	0.080	-0.058	0.358	
CSTRESS	-0.027	0.013	0.017	-0.055	-0.002	*
DADED	0.016	0.010	0.042	-0.002	0.037	
ECZ	0.025	0.047	0.295	-0.070	0.115	
EPVT	0.038	0.020	0.024	0.000	0.075	*

ESTRESS	0.004	0.007	0.305	-0.010	0.017	
HOMELOW	-0.248	0.100	0.009	-0.436	-0.051	*
LOCTHETA	0.034	0.019	0.041	-0.004	0.073	
LOWSES	-0.084	0.046	0.034	-0.176	0.006	
MATHS	0.032	0.020	0.062	-0.009	0.070	
MUMMAL	-0.078	0.039	0.025	-0.153	-0.001	*
POORNBHD	-0.089	0.066	0.094	-0.215	0.043	
SEPFMUM	-0.204	0.078	0.005	-0.357	-0.052	*
SETHETA	0.052	0.019	0.003	0.014	0.088	*
SMOKE	0.156	0.078	0.027	-0.002	0.307	
SMOKELVL	-0.087	0.034	0.007	-0.150	-0.021	*
UNMAR	0.000	0.063	0.499	-0.121	0.122	
WHEEZ	-0.049	0.041	0.109	-0.130	0.031	
ZEXT	-0.062	0.020	< 0.001	-0.101	-0.023	*
ZINT	-0.020	0.015	0.092	-0.048	0.008	
ZLEIS	0.037	0.014	0.004	0.011	0.068	*
R2	0.069			N	4,387	
PPP	0.451			BIC	154235.93	

Boys	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	-0.001	0.068	0.496	-0.138	0.130	
ADHDSEV	-0.004	0.020	0.414	-0.044	0.032	
BACKWARD	0.030	0.086	0.358	-0.144	0.203	
CSTRESS	-0.046	0.013	< 0.001	-0.070	-0.020	*
DADED	0.008	0.010	0.212	-0.012	0.027	
ECZ	0.034	0.047	0.223	-0.061	0.126	
EPVT	0.022	0.019	0.133	-0.017	0.058	
ESTRESS	0.004	0.007	0.302	-0.009	0.018	
HOMELOW	0.128	0.099	0.096	-0.050	0.329	
LOCTHETA	0.043	0.020	0.016	0.002	0.080	*
LOWSES	-0.061	0.048	0.085	-0.159	0.029	
MATHS	0.038	0.019	0.021	0.002	0.075	*
MUMMAL	-0.030	0.039	0.217	-0.105	0.048	
POORNBHD	-0.111	0.070	0.066	-0.241	0.030	
SEPFMUM	0.011	0.077	0.448	-0.139	0.176	
SETHETA	0.064	0.020	< 0.001	0.024	0.102	*
SMOKE	0.023	0.076	0.386	-0.133	0.162	
SMOKELVL	-0.032	0.034	0.178	-0.097	0.031	
UNMAR	-0.046	0.068	0.242	-0.189	0.080	
WHEEZ	0.011	0.037	0.392	-0.062	0.082	
ZEXT	-0.037	0.017	0.012	-0.071	-0.005	*
ZINT	-0.057	0.016	0.001	-0.089	-0.025	*

ZLEIS	0.032	0.015	0.019	0.001	0.060	*
R2	0.051			N	4,132	
PPP	0.392			BIC	149947.013	

Model 5

Girls	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	0.189	0.096	0.025	0.000	0.385	
ADHDSEV	-0.070	0.022	< 0.001	-0.116	-0.025	*
CSTRESS	-0.024	0.013	0.031	-0.050	0.001	
EPVT	0.039	0.019	0.018	0.003	0.077	*
ESTRESS	0.004	0.007	0.287	-0.010	0.017	
HOMELOW	-0.238	0.094	0.009	-0.414	-0.042	*
LOCTHETA	0.038	0.019	0.026	-0.001	0.076	
LOWSES	-0.092	0.046	0.019	-0.184	-0.005	*
MATHS	0.036	0.019	0.040	-0.003	0.072	
MUMMAL	-0.076	0.038	0.025	-0.150	-0.001	*
SEPFMUM	-0.207	0.080	0.004	-0.369	-0.053	*
SETHETA	0.054	0.019	0.003	0.017	0.089	*
SMOKELVL	-0.029	0.013	0.019	-0.055	-0.001	*
ZEXT	-0.066	0.020	< 0.001	-0.104	-0.026	*
ZINT	-0.021	0.015	0.074	-0.049	0.007	
ZLEIS	0.039	0.015	0.006	0.009	0.068	*
R2	0.064			N	4,387	
PPP	0.471			BIC	140751.314	

Boys	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	-0.003	0.071	0.484	-0.142	0.135	
ADHDSEV	-0.008	0.021	0.367	-0.045	0.035	
CSTRESS	-0.044	0.012	< 0.001	-0.067	-0.020	*
ESTRESS	0.002	0.007	0.388	-0.011	0.015	
LOCTHETA	0.045	0.019	0.013	0.009	0.083	*
LOWSES	-0.074	0.045	0.053	-0.164	0.015	
MATHS	0.050	0.018	0.001	0.015	0.087	*
SETHETA	0.066	0.019	< 0.001	0.030	0.104	*
ZEXT	-0.042	0.016	0.006	-0.074	-0.011	*
ZINT	-0.057	0.016	< 0.001	-0.088	-0.025	*
ZLEIS	0.036	0.014	0.006	0.007	0.065	*
R2	0.045			N	4,132	

PPP	0.529	BIC	116185.226

Model 6

Girls	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ACED5	0.026	0.009	< 0.001	0.008	0.046	*
ADHD	0.143	0.086	0.054	-0.033	0.298	
ADHDSEV	-0.031	0.022	0.078	-0.074	0.011	
CSTRESS	-0.003	0.012	0.381	-0.026	0.020	
EPVT	0.018	0.017	0.140	-0.016	0.051	
ESTRESS	0.012	0.006	0.024	0.000	0.024	*
HOMELOW	-0.148	0.085	0.042	-0.316	0.016	
LOCTHETA	0.014	0.019	0.221	-0.024	0.048	
LOWSES	-0.055	0.043	0.110	-0.135	0.031	
MATHS	0.001	0.018	0.481	-0.035	0.038	
MUMMAL	-0.034	0.036	0.170	-0.102	0.038	
OALCGRP	-0.158	0.034	< 0.001	-0.225	-0.087	*
ODIS	-0.230	0.040	< 0.001	-0.307	-0.151	*
OHLTH	-0.264	0.014	< 0.001	-0.292	-0.236	*
OLWPART	0.279	0.031	< 0.001	0.219	0.340	*
OSOC	-0.034	0.019	0.033	-0.071	0.003	
SEPFMUM	-0.118	0.075	0.059	-0.271	0.030	
SETHETA	0.021	0.017	0.103	-0.011	0.054	
SMOKELVL	-0.013	0.012	0.146	-0.035	0.011	
ZEXT	-0.047	0.018	0.005	-0.083	-0.013	*
ZINT	-0.035	0.014	0.005	-0.060	-0.008	*
ZLEIS	0.030	0.013	0.012	0.004	0.057	*
R2	0.212			N	4,387	
PPP	0.48			BIC	189425.587	

Boys	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ACED5	0.003	0.009	0.375	-0.015	0.021	
ADHD	-0.053	0.064	0.204	-0.179	0.065	
ADHDSEV	0.026	0.019	0.083	-0.010	0.063	
CSTRESS	-0.016	0.011	0.072	-0.036	0.005	
ESTRESS	0.002	0.006	0.349	-0.010	0.014	
LOCTHETA	0.019	0.019	0.154	-0.017	0.055	
LOWSES	0.016	0.042	0.344	-0.067	0.099	

MATHS	0.010	0.017	0.284	-0.026	0.043	
OALCGRP	-0.059	0.020	0.002	-0.099	-0.020	*
ODIS	-0.241	0.044	< 0.001	-0.326	-0.151	*
OHLTH	-0.314	0.015	< 0.001	-0.341	-0.285	*
OLWPART	0.327	0.029	< 0.001	0.273	0.383	*
OSOC	-0.046	0.019	0.010	-0.084	-0.009	*
SETHETA	0.044	0.018	0.006	0.008	0.078	*
ZEXT	-0.024	0.015	0.040	-0.053	0.003	
ZINT	-0.052	0.014	< 0.001	-0.078	-0.026	*
ZLEIS	0.022	0.014	0.062	-0.004	0.049	
R2	0.235			N	4,132	
PPP	0.461			BIC	165054.083	

Bayes regressions on 1 to 1 matched sample

Model 1

Girls	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	-0.235	0.13	0.032	-0.499	0.016	
R2	0.014			N	226	
PPP	0.461			BIC	638.038	

Boys	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	-0.163	0.100	0.048	-0.364	0.027	
R2	0.006			N	436	
PPP	0.48			BIC	1276.713	

Model 2

Girls	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	0.309	0.243	0.100	-0.163	0.788	
ADHDSEV	-0.277	0.108	0.006	-0.493	-0.079	*

R2	0.048		N	226	
PPP	0.480		BIC	636.696	

Boys	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	0.281	0.166	0.032	-0.019	0.618	
ADHDSEV	-0.259	0.079	< 0.001	-0.417	-0.107	*

R2	0.034		N	436	
PPP	0.549		BIC	1271.516	

Model 3

Girls	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	0.200	0.239	0.217	-0.265	0.679	
ADHDSEV	-0.226	0.105	0.011	-0.429	-0.032	*
BACKWARD	-0.217	0.302	0.235	-0.803	0.411	
DADED	0.134	0.066	0.027	-0.004	0.259	
HOMELOW	-0.428	0.236	0.030	-0.905	0.032	
MUMMAL	0.001	0.149	0.496	-0.293	0.279	
SMOKE	-0.089	0.383	0.416	-0.822	0.700	
SMOKELVL	0.054	0.145	0.372	-0.236	0.317	
WHEEZ	0.327	0.181	0.040	-0.060	0.678	

R2	0.153		N	226	
PPP	0.569		BIC	2347.461	

Boys	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	0.272	0.163	0.042	-0.032	0.597	
ADHDSEV	-0.252	0.078	< 0.001	-0.413	-0.108	*
SMOKE	-0.237	0.294	0.219	-0.825	0.332	
SMOKELVL	0.096	0.119	0.224	-0.145	0.318	
DADED	0.015	0.044	0.359	-0.074	0.099	
HOMELOW	0.426	0.342	0.109	-0.292	1.074	
MUMMAL	-0.101	0.116	0.191	-0.317	0.132	
BACKWARD	-0.472	0.198	0.008	-0.848	-0.084	*
WHEEZ	-0.060	0.122	0.313	-0.302	0.171	

R2	0.072		N	436	

PPP	0.627			BIC	4355.255

Model 4

Girls	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	0.220	0.262	0.190	-0.322	0.741	
ADHDSEV	-0.199	0.123	0.053	-0.444	0.034	
BACKWARD	-0.052	0.331	0.436	-0.700	0.580	
CSTRESS	-0.014	0.050	0.391	-0.111	0.081	
DADED	0.102	0.073	0.073	-0.038	0.245	
ECZ	0.112	0.228	0.307	-0.327	0.588	
EPVT	0.136	0.097	0.086	-0.049	0.326	
ESTRESS	0.020	0.034	0.261	-0.048	0.088	
HOMELOW	-0.425	0.297	0.075	-1.041	0.122	
LOCTHETA	-0.024	0.100	0.393	-0.219	0.177	
LOWSES	0.130	0.175	0.237	-0.248	0.450	
MATHS	0.004	0.098	0.486	-0.187	0.210	
MUMMAL	0.061	0.153	0.347	-0.253	0.347	
POORNBHD	0.364	0.354	0.164	-0.342	1.031	
SEPFMUM	-0.417	0.327	0.097	-1.088	0.199	
SETHETA	0.154	0.085	0.034	-0.017	0.318	
SMOKE	-0.268	0.397	0.247	-1.044	0.520	
SMOKELVL	0.129	0.155	0.209	-0.177	0.437	
UNMAR	0.045	0.226	0.394	-0.387	0.505	
WHEEZ	0.317	0.181	0.045	-0.050	0.666	
ZEXT	0.010	0.074	0.444	-0.137	0.151	
ZINT	0.016	0.069	0.413	-0.123	0.153	
ZLEIS	0.029	0.071	0.356	-0.112	0.180	

R2	0.284			N	226
PPP	0.069			BIC	10499.376

Boys	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	0.345	0.174	0.019	0.013	0.658	*
ADHDSEV	-0.164	0.085	0.025	-0.341	-0.003	*
BACKWARD	-0.338	0.217	0.066	-0.780	0.100	
CSTRESS	-0.077	0.035	0.010	-0.147	-0.011	*
DADED	0.002	0.045	0.481	-0.087	0.086	
ECZ	0.089	0.143	0.266	-0.212	0.365	
EPVT	-0.058	0.064	0.179	-0.185	0.073	
ESTRESS	0.019	0.023	0.180	-0.025	0.060	

HOMELOW	0.539	0.356	0.072	-0.179	1.191	
LOCTHETA	0.028	0.070	0.343	-0.125	0.160	
LOWSES	-0.194	0.150	0.100	-0.481	0.099	
MATHS	0.081	0.060	0.092	-0.042	0.196	
MUMMAL	-0.067	0.119	0.284	-0.303	0.164	
POORNBHD	-0.483	0.205	0.007	-0.886	-0.061	*
SEPFMUM	0.147	0.219	0.267	-0.307	0.558	
SETHETA	0.019	0.068	0.372	-0.115	0.154	
SMOKE	-0.293	0.297	0.171	-0.841	0.293	
SMOKELVL	0.129	0.121	0.167	-0.119	0.346	
UNMAR	-0.097	0.248	0.347	-0.567	0.381	
WHEEZ	-0.056	0.123	0.316	-0.299	0.191	
ZEXT	-0.064	0.047	0.086	-0.159	0.024	
ZINT	-0.024	0.048	0.300	-0.119	0.071	
ZLEIS	-0.024	0.055	0.326	-0.134	0.085	
R2	0.166			N	436	
PPP	0.098			BIC	19188.699	

Model 5

Girls	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	0.327	0.251	0.094	-0.143	0.833	
ADHDSEV	-0.196	0.122	0.055	-0.438	0.034	
CSTRESS	-0.020	0.046	0.338	-0.114	0.070	
ESTRESS	0.019	0.032	0.265	-0.043	0.082	
LOCTHETA	0.043	0.095	0.325	-0.135	0.239	
LOWSES	0.176	0.174	0.139	-0.191	0.513	
MATHS	0.012	0.099	0.454	-0.181	0.209	
SETHETA	0.200	0.077	0.006	0.058	0.349	*
ZEXT	-0.061	0.068	0.177	-0.196	0.074	
ZINT	0.011	0.062	0.436	-0.118	0.126	
ZLEIS	0.065	0.066	0.152	-0.065	0.197	
R2	0.144			N	226	
PPP	0.294			BIC	7282.269	

Boys	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ADHD	0.309	0.170	0.033	-0.020	0.657	
ADHDSEV	-0.150	0.085	0.049	-0.325	0.027	
CSTRESS	-0.087	0.033	0.005	-0.152	-0.024	*

ESTRESS	0.025	0.022	0.124	-0.018	0.066	
LOCTHETA	0.017	0.068	0.381	-0.106	0.153	
LOWSES	-0.182	0.147	0.106	-0.483	0.088	
MATHS	0.075	0.062	0.097	-0.044	0.202	
POORNBHD	-0.427	0.200	0.015	-0.840	-0.029	*
SETHETA	0.022	0.069	0.370	-0.117	0.155	
ZEXT	-0.055	0.044	0.097	-0.141	0.026	
ZINT	-0.023	0.047	0.303	-0.121	0.063	
ZLEIS	-0.024	0.052	0.329	-0.129	0.074	
R2	0.117			N	436	
PPP	0.402			BIC	14206.892	

Model 6

Girls	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig
ACED5	0.064	0.049	0.104	-0.036	0.155	
ADHD	0.258	0.236	0.147	-0.177	0.731	
ADHDSEV	-0.141	0.116	0.113	-0.360	0.080	
CSTRESS	-0.050	0.045	0.135	-0.141	0.036	
ESTRESS	0.026	0.030	0.184	-0.031	0.084	
LOCTHETA	0.042	0.089	0.319	-0.135	0.224	
LOWSES	0.170	0.157	0.151	-0.155	0.459	
MATHS	-0.002	0.097	0.492	-0.181	0.200	
OALCGRP	-0.357	0.178	0.018	-0.712	-0.027	*
ODIS	-0.353	0.183	0.026	-0.727	0.008	
OHLTH	-0.234	0.069	0.001	-0.373	-0.105	*
OLWPART	0.077	0.138	0.299	-0.207	0.337	
OSOC	0.130	0.086	0.073	-0.044	0.288	
SETHETA	0.116	0.073	0.056	-0.030	0.247	
ZEXT	-0.038	0.063	0.276	-0.162	0.087	
ZINT	0.003	0.062	0.476	-0.119	0.121	
ZLEIS	0.036	0.060	0.284	-0.079	0.147	
R2	0.317			N	226	
PPP	0.137			BIC	10227.068	

Boys	Est	Posterior SD	p-value	Lower 2.5%	Upper 2.5%	Sig

ACED5	-0.001	0.032	0.490	-0.061	0.065	
ADHD	0.137	0.152	0.188	-0.153	0.438	
ADHDSEV	-0.022	0.074	0.373	-0.170	0.119	
CSTRESS	-0.036	0.029	0.100	-0.092	0.018	
ESTRESS	0.048	0.018	0.003	0.013	0.085	*
LOCTHETA	-0.038	0.056	0.251	-0.147	0.080	
LOWSES	-0.134	0.124	0.166	-0.355	0.123	
MATHS	0.080	0.055	0.079	-0.032	0.184	
OALCGRP	-0.120	0.065	0.034	-0.257	0.011	
ODIS	-0.585	0.145	< 0.001	-0.869	-0.312	*
OHLTH	-0.258	0.045	< 0.001	-0.343	-0.171	*
OLWPART	0.395	0.093	< 0.001	0.213	0.578	*
OSOC	-0.031	0.070	0.334	-0.171	0.101	
POORNBHD	-0.325	0.170	0.026	-0.645	0.000	
SETHETA	0.047	0.057	0.215	-0.071	0.152	
ZEXT	-0.070	0.038	0.033	-0.150	0.003	
ZINT	-0.024	0.040	0.257	-0.102	0.057	
ZLEIS	-0.019	0.046	0.329	-0.107	0.076	

R2	0.352	N	436
PPP	0.275	BIC	19860.49

B. Robustness check of SWB factor score as a measure of wellbeing

Two of the chapter 7 regression models were replicated using the Warwick Edinburg Mental Wellbeing scale (WEMWBS), as a robustness check of the SWB factor score measure.

As described in chapter 6, the SWB was a factor score constructed from three observed variables: the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) score, self-rated life satisfaction, and a Rutter Malaise Scale score. The WEMWBS scale has been previously validated against several samples (Bass et al., 2016; Lloyd & Devine, 2012; Tennant et al., 2007), and in this sense would have been preferable to the unvalidated SWB score. However, the factor score was created in part to address the large percentage (29.8%) of missing responses to the WEMWBS scale in the BCS70 data. So, two models were replicated using WEMWBS: 1) ADHD alone as a predictor, and 2) a core and significant subset of covariates as identified in model 5 of the matched sample. Both models used the weighted matched sample. Results were compared to the corresponding models using the SWB measure (steps 1 and 5 from the matched sample section).

Model 1 (from matched section) comparison

Girls	WEMWBS	Sig	SWB	Sig
ADHD	-0.688		-0.100	
N = 3,179	R^2 = 0.000		R^2 = 0.000	

Boys	WEMWBS	Sig	SWB	Sig
ADHD	-1.136		-0.173	*
N = 2,802	R^2 = 0.001			R^2 = 0.002

Regression coefficients of WEMWBS score (30% missing) vs. SWB on ADHD. MLR estimator used on weighted matched sample in Mplus (Total N = 5,981)

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

Model 5 (from matched section) comparison

Girls	WEMWBS	Sig	SWB	Sig
ADHD	3.923	*	0.303	*
ADHDSEV	-0.541		-0.031	
BACKWARD	5.261		0.523	*
CSTRESS	0.059		0.000	
ESTRESS	0.221		0.013	
LOCTHETA	0.477		0.044	
LOWSES	-1.285		-0.133	~

MATHS	0.998	*	0.101	*
SETHETA	0.689		0.085	~
ZEXT	-1.609	*	-0.142	*
ZINT	0.019		0.002	
ZLEIS	1.133	*	0.091	*
N = 3,179	R2 = 0.121		R2 = 0.083	

Boys	WEMWBS	Sig	SWB	Sig
ADHD	1.533		0.072	
ADHDSEV	-0.070		-0.037	
CSTRESS	-0.852	**	-0.041	~
ESTRESS	-0.085		-0.009	
LOCTHETA	0.199		0.055	
LOWSES	2.287	~	0.055	
MATHS	0.968	**	0.080	*
SETHETA	-0.236		0.047	
ZEXT	-0.841	*	-0.027	
ZINT	0.485	~	-0.014	
ZLEIS	-0.168		0.009	
N = 2,902	R^2 = 0.057			R^2 = 0.037

Regression of WEMWBS score (30% missing) on matched step 5 covariates, compared to SWB. MLR on weighted matched sample in Mplus (Total N = 5,981)

**** p<0.001, ** p<0.01, * p<0.05*
Coefficients are unstandardized, for consistency with the convention used throughout chapter 7. WEMWBS scale range = 14 to 70, M = 49.3, SD = 8.3; SWB scale range = -4.6 to 2.5, M = 0, SD = 0.95)

The results were similar, in that signs were the same direction and there was good overlap in the sets of significant predictors. ADHD alone explained almost no variance for either outcome variable. The differences are highlighted below:

Model 1

- ADHD alone was negative for Warwick and SWB, for both girls and boys. The only significant coefficient was for boys, predicting SWB

Model 5

- ADHD, MATHS, ZEXT (externalising score), and ZLEIS (engagement in leisure score) were significant for girls for both WEMWBS and SWB. BACKWARD was significant for SWB (and discounted as spurious) but was not for WEMWBS
- For boys, MATHS was the only significant predictor for SWB, but MATHS, CSTRESS, and ZEXT were significant for WEMWBS. The CSTRESS variable had a $p < 0.20$ for SWB, but ZEXT did not.
- The R^2 values were higher for WEMWBS

Boys with ADHD had slightly higher missingness than girls with ADHD (47% v 43%), which could partly explain the greater disparity between SWB and WEMWBS results for boys compared to girls. There was higher missingness in the ADHD group than in the overall sample (31 v 27%), and the mean ADHDSEV was lower. Also, ADHDSEV had a significant relationship with ZEXT ($r = 0.474$, $p < 0.001$). These factors could explain the slightly different behaviour for the ZEXT variable between the two samples. Finally, BACKWARD was found in the previous analysis of the matched sample to interact with ADHD when predicting SWB, and thus coefficients deemed to be unreliable. The sample differences and interactions are reasonable explanations for the minor differences between the results for the two outcome measures, and thus provide some evidence that WEMWBS and SWB measure the same construct.