Early infant candidate markers of autistic social trait developmental trajectories in the general population

Gemma Halliday

D.Clin.Psy. thesis (Volume 1), 2019
University College London
I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature: 

Name: Gemma Louise Halliday

Date: 5\textsuperscript{th} July 2019
Overview

This thesis investigated early infant candidate markers of developmental trajectories of autistic social traits (ASTs) in a large general population cohort: the Avon Longitudinal Study of Parents and Children (ALSPAC). Previous research on the relationship between internalising and externalising traits and autistic traits (ATs) in the general population was reviewed, and limitations and directions for research and clinical implications have been highlighted.

Part 1, the literature review, quantified the association between ATs and internalising (anxiety and depression) and externalising (attention deficit hyperactivity disorder [ADHD], conduct disorder and oppositional defiant disorder) symptoms in the general population for the first time. A positive correlation between ATs and internalising and externalising symptoms (ADHD only) was identified, with medium to large effect sizes. Individuals with more severe ATs experience higher levels of psychopathologic traits across the lifespan.

Part 2, the empirical paper, investigated early infant behavioural predictors of AST latent developmental trajectories from mid-childhood to mid-adolescence (seven to 16 years) and explored gender differences in predictors in a large general population cohort (n=7,773) using multi-group Latent Growth Curve Analysis. Social communication, language and temperament characteristics at 15 and 24 months predicted initial AST status. Trends towards gender differences were observed, with markers more strongly associated with male ASTs. Developmental differences from infancy predict childhood AST severity and this may differ between genders, however such predictors are limited in explaining AST chronogeneity.

Part 3, the critical appraisal, considers limitations, dilemmas, directions for future research, and offers personal reflections on the research process.
Impact Statement

This study has important implications across academic research and clinical practice domains. The present findings indicate researchers and clinicians need to move towards dimensional autistic trait (AT) measurement across settings. In line with increasing recognition of gender differences in the autism phenotype, this study acutely highlights the concerning neglect of gender differences in population-based AT research. Adopting a fundamentally dimensional approach to the study of autism is more likely to bear fruit in progressing our understanding of AT chronogeneity and etiology for males and females, therefore improving clinical assessment, formulation and intervention.

In the academic domain, this study progressed the field of autism comorbidity research. Meta-analyses demonstrated an aggregate moderate association between subthreshold ATs and a range of subthreshold psychopathologic traits in population-based samples for the first time. Importantly, this indicates that even at the subthreshold level ATs can provide valuable contributions to our evolving understanding of the relationship between autistic and psychopathologic traits. A future research priority is to use population-based longitudinal twin designs to disentangle the developmental sequence between subthreshold autistic and psychopathologic traits, and to establish the extent of shared and distinct genetic and environmental etiology.

The findings in Part 2 support the call for identification of early infant markers of AT developmental trajectories from the first two years of life to better understand AT chronogeneity. The study extends previous research which identified associations between early behavioural markers and autism by demonstrating similar markers also predict subthreshold AST severity in the general population. Multi-group Latent Growth Curve Modelling was demonstrated as a flexible and
powerful analytic method to answer complex research questions about
developmental trajectories of traits. This study highlighted the importance of
exploring gender differences in infant candidate markers of AST trajectories. Trends
towards gender differences in developmental behaviours related to the severity of
ASTs were demonstrated for the first time. Given the male-centric definition of
autism, it is crucial that future research characterises and understands the
chronogeneity of the understudied female autism phenotype.

The implications for clinical practice are that routine screening of ATs from
childhood should be implemented across mental health and developmental clinical
services to improve early detection of higher ATs. Screening for behavioural
markers of higher ASTs from as young as 15 to 24 months using measures of social
communication, language and temperament could be clinically useful and
meaningful. Clinicians should consider that the ‘red flags’ (developmental
behavioural markers) indicative of possible high ATs in females may be different to
males. This traverses back into the research domain, indicating a need for
development and validation of female-specific AT screening and diagnostic tools.

Dissemination of study findings through scholarly journals and relevant
media outlets such as Spectrumnews.org would encourage dialogue and feedback
from academics, professionals and, importantly, the identifying autistic community.
Table of Contents

Appendices.................................................................................................................9

Contents of tables and figures......................................................................................10

Acknowledgements..........................................................................................................12

Part 1. Literature Review: Autistic traits and internalising and externalising traits are associated in the general population: A meta-analysis.................................13

Abstract.........................................................................................................................14

1.0 Introduction................................................................................................................15
  1.1 Autism Spectrum Disorder (ASD)..........................................................................15
  1.2 Autistic-like traits (ATs)........................................................................................15
  1.3 A dimensional model of psychopathology.........................................................16
  1.4 Comorbidity in autism.........................................................................................16
  1.5 Comorbidity with internalising and externalising disorders.............................17
  1.6 Co-occurrence of autistic and internalising and externalising traits..................18
  1.7 Moderating factors: life stage and study design...............................................18
  1.8 Review aims........................................................................................................19

2.0 Method.......................................................................................................................20
  2.1 Search strategy......................................................................................................20
  2.2 Inclusion and exclusion criteria..........................................................................20
  2.3 Risk of bias..........................................................................................................22
  2.4 Analytic strategy..................................................................................................23
  2.5 Effect size calculation..........................................................................................23
  2.6 Statistical procedures..........................................................................................24

3.0 Results.........................................................................................................................25
  3.1 Study selection......................................................................................................25
  3.2 Corpus of studies..................................................................................................27
    3.2.1 Internalising outcomes..................................................................................28
    3.2.2 Externalising outcomes................................................................................28
  3.3 Measures................................................................................................................29
    3.3.1 AT measures..................................................................................................29
    3.3.2 Internalising measures..................................................................................29
    3.3.3 Externalising measures...............................................................................30
  3.4 Risk of bias............................................................................................................31
  3.5 Meta-analyses........................................................................................................32
    3.5.1 Internalising: anxiety symptoms.................................................................32
    3.5.2 Internalising: social anxiety symptoms.......................................................33
    3.5.3 Internalising: depression symptoms............................................................33
    3.5.4 Externalising: ADHD symptoms.................................................................33
  3.6 Moderation analyses.............................................................................................34
    3.6.1 Study design..................................................................................................34
    3.6.2 Sample age...................................................................................................34
  3.7 Publication bias........................................................................................................35
  3.8 Narrative synthesis.................................................................................................35
    3.8.1 Internalising: other anxiety symptoms.........................................................35
    3.8.2 Internalising: depression symptoms.............................................................36
    3.8.3 Externalising: conduct symptoms.................................................................36

4.0 Discussion...................................................................................................................55
Part 2. Empirical Paper: Early infant candidate markers of autistic social trait developmental trajectories in the general population

Abstract

1.0 Introduction

1.1 Autism Spectrum Disorder (ASD)

1.2 Autistic-like traits (ATs)

1.3 Individual differences in developmental trajectories of autistic social traits (ASTs)

1.4 Gender differences in AST trajectories

1.5 Early candidate markers of AST trajectories

1.5.1 Social communication and language

1.5.2 Motor development

1.5.3 Temperament

1.6 Early candidate markers of ATs and ASD in the general population

1.7 Study aims

1.7.1 Research questions

2.0 Method

2.1 Setting

2.2 Participants

2.2.1 Recruitment

2.2.2 Sample

2.2.3 Demographic characteristics

2.3 Power calculation

2.4 Measures

2.4.1 Autistic social traits (ASTs): Social and Communication Disorders Checklist

2.4.2 Early behavioural candidates

2.4.2.1 Denver Developmental Screening Test

2.4.2.2 MacArthur Communicative Development Inventories

2.4.2.3 Carey Infant Temperament Scales

2.5 Ethics

2.6 Data analysis

2.6.1 Reproduction of Mandy et al. (2018) multi-group LGCM

2.6.2 Hypothesised covariate multi-group LGCM

2.6.3 Gender differences

2.7 Missing data handling

3.0 Results

3.1 Descriptive statistics

3.1.1 Missing data

3.1.2 Correlations amongst predictors

3.2 Multi-group LGCM

3.2.1 Multi-group LGCM identification

3.3 Hypothesised predictors of multi-group LGCM

3.4 Exploratory analyses

3.4.1 Gender differences in predictors of LGCM
4.0 Discussion ........................................................................................................130
  4.1 Summary of key findings ..................................................................................130
  4.2 Findings in the context of previous research ..................................................131
  4.3 Limitations and Research Implications ..........................................................136
  4.4 Conclusions: Scientific Implications ..............................................................138
  4.5 Conclusions: Clinical implications .................................................................139
5.0 References ........................................................................................................140

Part 3. Critical Appraisal ......................................................................................160

1.0 Dealing with dilemmas and methodological decisions .....................................161
  1.1 Alternative statistical approaches ...................................................................161
  1.2 Selection of early candidates of AST chronogeneity .......................................163
2.0 Measurement of ATs in population-based research ..........................................165
3.0 We need to talk about women ..........................................................................167
4.0 Other limitations and directions for research ..................................................169
5.0 References ........................................................................................................169
Appendices

Appendix I: Search terms used in PsychINFO and Medline.............................................177

Appendix II: Newcastle-Ottawa Quality Assessment Scale (Wells et al., 2014).....179

Appendix III: Participant demographic information......................................................181

Appendix IV: Social and Communication Disorders Checklist (Skuse et al., 2005).................................................................................................................................184

Appendix V: Example syntax specifying multi-group quadratic LGCM with predictors of model intercept, slope and quadratic.................................................................186
Contents of tables and figures

Part 1: Literature review

Figure 1: PRISMA diagram...........................................................................................................26

Table 1: Characteristics of included studies measuring the association ATs and internalising outcome(s)..................................................................................................................38

Table 2: Characteristics of included studies measuring the association between ATs and externalising outcome(s)............................................................................................................42

Table 3: Summary of measures used to assess ATs in the included studies..................44

Table 4: Newcastle Ottawa Scale (NOS; 2014) quality assessment scores........48

Figure 2: Forest plot of ATs and anxiety symptoms meta-analysis..................49

Figure 3: Forest plot of ATs and social anxiety symptoms meta-analysis..............50

Figure 4: Forest plot of ATs and depression symptoms meta-analysis...............51

Figure 5: Forest plot of ATs and ADHD symptoms meta-analysis........................52

Table 5: Effect of study design on AT and mental health outcome association......53

Table 6: Effect of sample age on AT and mental health outcome association......53

Figure 6: Funnel plots.................................................................................................54

Part 2: Empirical paper

Figure 1: SCDC trajectories for boys and girls...............................................................94

Figure 2: Developmental gender difference in SCDC across childhood and adolescence.................................................................94

Figure 3: ALSPAC recruitment phases, from Boyd et al. (2013)..........................103

Figure 4: ALSPAC study attrition, from Boyd et al., (2013)...................................104

Table 1: Predictor variables, measure and timepoint of measurement by developmental domain..................................................................................................................107

Figure 5: Path diagram of quadratic LGCM showing male group as example......113

Figure 6: Path diagram of quadratic LCGM with predictors of growth factors for male group ..................................................................................................................114

Table 2: Number and percent of missing data points (n=7773)..............................116

Table 3: Missing data at each time point (n=7773).................................................116
Table 4: Number and percent with missing data for predictors form Mandy et al., (2018) sample (n=9744)

Table 5: Correlations among predictor variables, sample mean and standard deviation (n=7773)

Table 6: Growth factor means for multigroup quadratic LGCM

Figure 7: AST trajectories for males and females

Figure 8: Gender differences in AST across childhood and adolescence

Table 7: Estimated mean SCDC scores for males and females and mean differences in gender across childhood and adolescence

Table 8: Summary of the 11 predictors of model intercept and slope for males and females (n=7773)

Figure 9: AST trajectories for males and females scoring high and low on the CTS Activity subscale at 24 months old

Figure 10: AST trajectories for males and females scoring high and low on the DDST Language subscale at 24 months old

Table 9: Gender differences in predictors of model intercept and slope (n=7773)
Acknowledgements

I am indebted to the children and families who participated in the ALSPAC study and to the ALSPAC research team for supporting this project. I am extremely grateful to my supervisors Pasco Fearon and Will Mandy for the opportunity to work with them on this study, and for their generous support, encouragement, and insightful feedback. I am also very grateful to Jon Heron at the University of Bristol for his guidance to a total beginner in MPlus, and his generosity in sharing his thinking around latent growth modelling with the ALSPAC data. I would also like to extend thanks to Richard Pender, a fellow UCL DClinPsy alumni whose DClinPsy thesis inspired further examination of the ALSPAC data, for his time in rating studies for the meta-analysis.

I am forever grateful to Tony Roth and Steve Pilling for seeing potential in me during our work together prior to training. And to my clinical tutor, Kate Sherratt, for her support and provision of a reflective space for my journey through the doctorate. I would also like to thank the friends, colleagues and clinical supervisors, particularly during my final placement, whose sensitivity, empathy and humour made the road as smooth as it could be.

Finally, I would like to thank my partner Joe for his unwavering patience, calm reassurance, prompts to self-care, and for his slightly worrying yet incredibly helpful excitement about equations and graphs. And of course, my family, who always believe in me and my dreams.
Part 1: Literature Review

Autistic traits and internalising and externalising traits are associated in the general population: A meta-analysis.
Abstract

Aims: To quantify the association between autistic traits (ATs) and internalising and externalising traits in the general population. Internalising traits included anxiety and depressive symptoms. Externalising traits included attention deficit hyperactivity disorder (ADHD), oppositional defiant disorder, and conduct disorder symptoms.

Method: EMBASE, Medline and PsycINFO were systematically searched using terms related to autistic traits, internalising and externalising symptoms. Thirty five studies were included, reporting on over 80,000 participants. Standardised effect sizes of correlations between ATs and internalising and externalising symptoms were calculated, and meta-analyses used a random effects model. Categorical analysis was used to explore factors (study design and sample age) influencing the association between ATs and psychopathologic symptoms.

Results: A positive correlation between ATs and internalising and externalising symptoms (ADHD only) in the general population was identified, with medium to medium to large effect sizes. The largest effect was for ATs and social anxiety symptoms ($r = .44$). Meta-analysis was not possible for conduct symptoms and no studies met inclusion criteria for oppositional symptoms. Sample age had no effect on trait association. Study design significantly influenced the effect size for depression symptoms only, with cross-sectional studies evidencing larger effects.

Conclusion: Individuals in the general population with more severe ATs experience higher levels of psychopathologic traits. Subthreshold ATs are important in developing our understanding of the relationships between autistic and psychopathologic traits. Dimensional AT measurement is needed across settings. Conclusions are tentative due to high heterogeneity between studies and a focus on internalising symptoms with adults using cross-sectional designs.
1.0 Introduction

1.1 Autism Spectrum Disorder (ASD)

Autism Spectrum Disorder (ASD) is a lifelong neurodevelopmental condition affecting around 1% of the population, or 52 million people globally (Baxter et al., 2015; Elsabbagh et al., 2012; Robinson et al., 2016). Diagnosis of ASD requires atypicalities in two domains of behaviour present from early childhood; social communication and interaction, and restricted, repetitive behaviours, interests or activities (Diagnostic and Statistical Manual of Mental Disorders fifth edition [DSM-5]; American Psychiatric Association [APA], 2013). Males are more commonly diagnosed than females, at a ratio of three to one (Loomes, Hull & Mandy, 2017). Autism is associated with poorer physical, psychological, social and quality of life outcomes (Levy & Perry, 2011; Rydzewska et al., 2019; van Heijst & Geurts, 2015).

1.2 Autistic-like traits (ATs)

Autism is a heterogenous condition, with varied neurobiological pathways contributing to different presentations and severities of autistic behaviours and symptoms (Robinson et al., 2016). It is now understood as a dimensional condition instead of a categorical disorder, representing the extreme manifestation of a ‘constellation’ of autistic-like behavioural traits (ATs) that exist on a continuum of impairment from subthreshold to exceeding diagnostic threshold (Constantino, 2009). Population-based studies demonstrate that ATs are continuously distributed in children and adults in the general population and represent a genetic liability for autism across the continuum of trait severity, known as the Broader Autism Phenotype (BAP) (Baron-Cohen, Wheelwright, Skinner, Martin & Clubley, 2001; Constantino & Todd, 2003; Lundström et al., 2012; Robinson et al., 2011a; 2011b; Robinson et al., 2016; Skuse et al., 2009; St. Pourcain et al., 2014).
1.3 A *dimensional model of psychopathology*

The internalising-externalising model is an established two-dimensional structure of psychopathology which proposes any psychopathology can be characterised as one of two behavioural trait dimensions that reflect two underlying core psychopathological processes (Achenbach & Edelbrock, 1981). These processes are: internalising; anxiety and depression symptoms, and externalising; hyperactive-impulsive, aggressive and delinquent behaviour symptoms.

This model is highly replicable and empirically supported across general community and clinical population studies (e.g., Forbush & Watson, 2013; Kendler, Prescott, Myers & Neale, 2003; Slade & Watson, 2006), accounting for the comorbidity of different child and adult mental disorders across the lifespan (Caspi et al., 2013). Further support comes from converging evidence that, similar to ATs, traits of externalising disorders such as attention deficit hyperactivity disorder (ADHD) and oppositional defiant (ODD), and internalising disorders such as anxiety and depression exist on a continuum in the general population (e.g., Bosman et al., 2019; Coghill & Sonuga-Barke, 2012; Judd et al., 2002; Lubke et al., 2009; Shear et al., 2007; Sterba et al., 2010).

1.4 *Comorbidity in autism*

Co-morbidity is the co-occurrence of two or more categorically defined disorders, with a second diagnosis of different core symptoms to the first (Matson & Nebel-Schwalm, 2007). Comorbidity is the norm in autism, with at least 50% of individuals meeting criteria for another disorder (Mannion, Brahm & Leader, 2014), and both persists and changes over the lifespan. Autism comorbidity has been studied categorically using clinical samples with an autism diagnosis, or clinically-ascertained cohorts with a psychiatric disorder at increased risk of comorbidity.
Intellectual Disability, defined as Intelligence Quotient (IQ) less than 70, is arguably the most common, affecting 50% to 70% of autistic children (Chakrabarti & Fombonne, 2005; Charman et al., 2011; Isaksen, Diseth, Schjolberg & Shjeldal, 2013). A population-representative study reported 70% of autistic children aged 12 years met criteria for one additional mental disorder and 41% met criteria for two (Siminoff et al., 2008). Systematic and meta-analytic reviews of clinically-ascertained samples are lower, between 50% to 70% (Mannion et al., 2014). A meta-analytic review of over 26,000 autistic adults found a staggering 54.8% had at least one comorbid psychiatric disorder (Lugo-Marin et al., 2019).

1.5 Comorbidity with internalising and externalising disorders

Autism is associated with a range of internalising and externalising disorders. Most commonly co-occurring in childhood are the anxiety disorders (48%), followed by ADHD (28%) and ODD (28%) (Gadow et al., 2005; Simonoff et al., 2008; van Steensel et al., 2011). The neurodevelopmental disorder of ADHD (25.7%) is more common than mood and anxiety disorders in adulthood (18.8% and 17.8% respectively), but this may be an artefact of heterogeneous assessment and study design (Lugo-Marin et al., 2019). ADHD typically emerges in childhood and shares some overlapping symptoms and etiological factors with autism including a male majority (Antshel & Russo, 2019; Larsson et al., 2012; Posthuma & Polderman, 2013; Robinson et al., 2011a; Ros & Graziano, 2018). Comorbid ODD and conduct disorder (CD) are relatively less understood (Mandy, Roughan & Skuse, 2014), despite one in four autistic children having a comorbid diagnosis (Kaat & Lecavalier, 2013).

Individuals with autism also experience higher rates of anxiety and depression than the neurotypical population (Costello et al., 2005; Spiers et al., 2012; Vasa & Mazurek, 2015; Wigham, Barton, Parr & Rodgers, 2017). Most
common in childhood and adolescence is specific phobia (30%), followed by obsessive-compulsive disorder (OCD; 17%), and social anxiety disorder and agoraphobia (17%) (van Steensel et al., 2011). Social anxiety disorder overtakes in adulthood, likely due to increased social demands and the reduced fear of social evaluation due to reduced mentalising abilities or difficulty verbalising in childhood (Hallett et al., 2013), followed by OCD (Lugo-Marin et al., 2019).

1.6 Moderating factors: life stage and study design

There are inconsistencies in reported autism comorbidity prevalence rates which could be explained by numerous factors. One likely contributing factor is the life stage of the sample; childhood versus adolescence versus adulthood. Some disorders develop normatively earlier or later in life in individuals without autism; for example, it is well-established that anxiety and ADHD typically manifest earlier in life than depression. Within disorder categories, there is also changeability over time in line with normative neurological, cognitive and social developments and/or changing environmental demands; for example, social anxiety disorder prevalence increases with age (e.g., Merikangas et al., 2010).

Another moderating factor could be study design and/or measurement, with the reviewed literature using variable cross-sectional and longitudinal designs. The developmental nature of longitudinal designs makes them more advantageous in providing information on the continuity and possible direction of causality of traits (Hawks, Marrus, Glowinski & Constantino, 2019; Krueger, Tackett & MacDonald, 2016).

1.7 Co-occurrence of autistic and internalising and externalising traits

The highly debated question is what does this categorical comorbidity represent: true comorbidity where symptoms are genuine manifestations of distinct
mental disorders, phenocopies (a trait resembling another trait without its genotype) that are epiphenomena of autism, or something else?

Given the dimensionality of autism and psychopathology (Achenbach & Edelbrock, 1981; Caspi et al., 2013), ATs may be an important and understudied risk factor for understanding a range of common psychiatric syndromes (Constantino, 2018). Clinical studies show autistic symptom severity is associated with behavioural disability (Dworzynski et al., 2009), increased risk of various emotional and behavioural impairments (Constantino & Frazier, 2013), and population studies suggest this is the case across the continuum of trait severity (e.g., Lundstrom et al., 2011b; Hallett et al., 2009; Hallett, 2012; Hallett et al., 2013; Tick et al., 2016). Understanding how and why traits co-occur can elucidate how ATs influence the development of internalising and externalising behaviours and vice versa, informing accurate and appropriate assessment and clinical formulation and intervention for people with ATs and autism, and other psychopathology symptoms.

1.8 Review aims

Despite the dimensional re-conceptualisation of autism and psychopathology, the comorbidity literature is primarily focused on clinical samples or clinically-ascertained cohorts with a psychiatric disorder at increased risk of comorbidity (e.g., Towbin et al., 2005; Pine et al., 2008). This does not provide direct information on the co-occurrence of subclinical ATs and psychopathology symptoms. This review therefore aims to quantify the strength and direction of these associations for the first time. Factors that may account for disparities seen in the literature will be explored. Studies reporting on associations between ATs and internalising and externalising symptoms in general population samples will be systematically searched for in order to answer the following:
1. What is the nature of the association between ATs and internalising symptoms of depression and anxiety in the general population, across the lifespan?

2. What is the nature of the association between ATs and externalising symptoms of ADHD, oppositionality and conduct problems in the general population, across the lifespan?

3. Are associations moderated by study characteristics?
   a. study design.
   b. sample age.

2.0 Method

2.1 Search strategy

A systematic search was conducted in EMBASE, PsycINFO and Medline from 1st January 1974 (Embase), 1st January 1806 (PsycINFO) and 1st January 1946 (Medline) up to and including 30th August 2018. The search was limited to original research studies published in English in peer-reviewed journals. Cross-references of the obtained studies were checked for studies that might have been missed by electronic search. Search terms related to autistic traits (ATs) were combined with terms related to depression, anxiety, ADHD, oppositional and conduct behaviour symptoms, adapted to database index. Terms used to search EMBASE are reported in the PRISMA diagram as an example (see Figure 1), with the same terms for the concept of ATs and adapted search terms for the concept of mental health outcome in PsychInfo and Medline (see Appendix I).

2.2 Inclusion and exclusion criteria

The inclusion criteria were that a study:
(i) Unselectively sampled from the general population at any age. Where a study selectively sampled but included a typically developing healthy control group from the general population, correlation(s) reported for this control group were included if criteria (ii) to (vi) were met;

(ii) Reported a correlation coefficient between ATs and at least one mental health outcome;

(iii) Assessed ATs via total score on a validated self-report instrument with a continuous scale indicating degree of severity (details of these measures are summarised in Table 1);

(iv) Reported on at least one mental health outcome, which included: symptoms of depression, anxiety, eating or feeding disorders (internalising), and/or symptoms of ADHD, conduct problems, oppositional problems (externalising);

(v) Assessed the mental health outcome via total score or subscale of a validated self-report instrument with a continuous scale indicating degree of severity;

(vi) Cross-sectional or longitudinal cohort design. For longitudinal studies, the same AT and mental health outcome measure were applied at all time points.

The exclusion criteria were that a study:

(i) Selectively sampled from a clinical setting (e.g., inpatients or participants attending a mental health clinic or autism clinic);

(ii) Included an overrepresentation of participants with autism who met diagnostic criteria, greater than 5% of total n, according to either the DSM-III-R (APA, 1987), DSM-IV (APA, 1994), DSM 5 (APA, 2013), or ICD-10 (International Statistical Classification of Diseases and Related
Health Problems – 10th revision; World Health Organization [WHO], 1992);

(iii) Used a case control, case series or case report design;

(iv) Reported a correlation for a general measure of ‘internalising’ and/or ‘externalising’ mental health outcome only (e.g., Strengths and Difficulties Questionnaire [SDQ], Child Behaviour Checklist [CBCL]).

An additional exclusion criterion was applied post-search due to the initial large size of the review in agreement with supervising expert reviewers:

(i) Studies measuring symptoms of eating or feeding disorders.

Reviews showing an elevated prevalence of psychotic and eating disorders in people with ASD and of ATs in both clinical and general population samples exist and therefore these psychopathologies are outside the scope of this review (Kincaid, Doris, Shannon & Mulholland, 2017; Westwood & Tchanturia, 2017).

2.3 Risk of bias

Risk of bias for individual studies was assessed by the reviewer using the Newcastle Ottawa Quality Assessment Scale (NOS) for cohort studies designed to aid interpretation of meta-analytic results for nonrandomised studies (Wells et al., 2014) (see Appendix II). The 9-item measure gives an overall risk of bias rating denoted by total number of stars awarded per item summed across 3 domains: cohort selection, comparability of cohorts, and assessment of outcome. A study can be awarded up to seven stars in total across the selection and outcome categories, and a maximum of two stars for comparability, yielding a maximum of nine stars. A higher number of stars indicates lower risk of bias and therefore higher quality. Ratings were independently double-rated by a qualified clinical psychologist with experience in systematic reviews of ATs. A Two-Way Random consistency intra-
class correlation coefficient (ICC) using average measures assessed inter-rater reliability, with high agreement of ICC (2,2) = .90 (Shrout & Fleiss, 1979).

2.4 Analytic strategy

Data on study characteristics and outcome(s) was extracted from included studies by the reviewer. Meta-analyses were conducted where there were more than two studies that could be aggregated because standard random-effects meta-analytic methods perform poorly when applied to a very small number of studies (Guolo & Varin, 2017; Liberati et al., 2009; Seide, Röver & Friede, 2019). Sensitivity analyses were pre-specified; effects were to be examined according to study design and sample age. Where studies could not be included in meta-analysis, a brief narrative synthesis was conducted (Popay et al., 2006). This was only the case for a few studies (n=5). Measures that would be included in the analysis were finalised before effect sizes were calculated so as to reduce bias. Where a study occasionally reported correlations for more than one AT measure, the most commonly used AT measure in the included studies was extracted.

2.5 Effect size calculation

The AT and mental health symptom measures generated continuous outcomes from which the correlation coefficient (Pearson’s r statistic) was extracted as the raw value of the strength of the (linear) relationship (Borenstein, 2009). Standardised effect sizes were calculated using the Fisher’s r-to-z transformation of r into a z-score to give a standardised effect size for aggregation. This is a variance stabilising and normalising transformation, whereby the skewed distribution of the sample correlation (r) is converted into an approximation of a normal distribution (z), enabling comparison of the difference between coefficients where the variance of
the sampling distribution is independent of the correlation, summarised as (from Lipsey & Wilson, 2001):

\[
\text{Effect size } z' = 0.5[\ln(1+r) - \ln(1-r)]
\]

where \( z' \) is the effect size estimate, \( r \) is the raw correlation coefficient effect size, and ‘\( \ln \)’ is the natural logarithm. Calculation of the standard error of the \( z \)-transformed correlation was (from Lipsey & Wilson, 2001):

\[
\text{Standard error } z' = \frac{1}{\sqrt{n-3}}
\]

Effect sizes were interpreted in line with the guidance by Cohen (1988; 1992): small = .10, medium = .30, and large = .50. Homogeneity of effects among studies was tested with the chi-squared \( Q \) test (Hedges & Olkin, 1985), and the \( I^2 \) statistic to measure inconsistency (the percentage of total variation across studies due to heterogeneity) of effects. The latter is more meaningful for a random-effects model; it does not inherently depend on the number of studies and is accompanied by an uncertainty interval (Borenstein et al., 2009; Higgins & Thompson, 2002).

Heterogeneity and publication bias were assessed by visual evaluation of funnel plots of the back-transformed correlation effect sizes for asymmetry, with the latter resulting from the non-publication of small studies with negative results (Rosenthal, 1979). Asymmetry was assessed using weighted linear regression tests (Egger et al., 1997).

2.6 Statistical procedures

Data were analysed using the Metafor package for R (Viechtbauer, 2010). A random-effects meta-analysis model with restricted maximum-likelihood estimation provided an average effect size for the population of studies, as the included studies
were assumed to be a random selection of the entire study population and there was variability at the study and participant level (Hedges & Vevea, 1998; Lipsey & Wilson, 2001; Viechtbauer 2005).

A mixed-effects model then determined the moderating effects of two study characteristics on the size of this effect: design, and sample life stage based on mean sample age. Two proposed categorical dummy variables were coded and calculated for moderators: design (cross-sectional or longitudinal), and life stage (childhood [0 to 9.9 years], adolescence [10.9 to 19.9 years], adulthood [20 years and above]). The definition of adolescence was informed by the current definition of adolescence as people aged 10-19 years (World Health Organisation [WHO], 2019). Effect sizes were regressed on the study characteristics in a restricted maximum-likelihood meta-regression (Viechtbauer, 2005; Viechtbauer, 2010).

3.0 Results

3.1 Study selection

The flow of papers through screening for inclusion in the review is shown in Figure 1.
After combination and de-duplication, the search retrieved 1039 papers from EMBASE, 636 papers from PsychINFO and 1830 papers from Medline. After combination and de-duplication,
2356 papers remained. Titles and abstracts were screened against inclusion and exclusion criteria, resulting in exclusion of 2161 and inclusion of 195 studies, including two papers which reported on two separate studies respectively, counted as individual studies: Dickter et al., 2018a; 2018b, Lundström et al. 2011a; 2011b. Full-text articles were screened for eligibility, resulting in exclusion of 160 papers for the following reasons: correlation not reported (n=70); no appropriate mental health outcome (n=38); selectively sampled or clinical population (n=34); unvalidated, idiosyncratic or non-continuous AT measure (n=11); general measure of ‘internalising’ and/or ‘externalising’ mental health outcome (n=9); correlation with subscale(s) of AT measure (n=5); relative risk ratio or other statistic (n=4); and dissertation (n=1). The references of included papers and three excluded reviews were searched for additional papers not retrieved by the search, but none were included after screening.

The 35 included studies were categorised into those measuring the relationship between ATs and internalising mental health symptom(s) (n=25), and those measuring the relationship between ATs and externalising mental health symptom(s) (n=12). Two studies were included in both quantitative and qualitative synthesis due to reporting additional outcomes that could not be included in meta-analysis (Liew et al., 2015; Zhou et al., 2018).

### 3.2 Corpus of studies

The search yielded 35 included studies. Studies measuring an internalising outcome are described in Table 1, and an externalising outcome in Table 2. Across all outcomes, participants ranged in mean age from 7.9 years to 69.2 years. The majority of studies were cross-sectional (n=31), with four longitudinal studies included, and the majority involved adults (n=24), with seven involving adolescents and four involving children.
3.2.1 Internalising outcomes

Twenty-five studies measured an internalising outcome of depression and/or anxiety symptoms, involving 45,042 participants. Ten studies reported on an association between ATs and trait anxiety symptoms, eight studies on ATs and social anxiety symptoms, three studies on ATs and generalised anxiety symptoms, and two studies on ATs and obsessive-compulsive symptoms. Seventeen studies measured an association between ATs and depressive symptoms. Across studies, 23 were cross-sectional and three were longitudinal in design. Of the longitudinal studies, one followed up participants over three timepoints with the longest period being three months (Asano et al., 2014), and two followed up participants once, at six months (Geurts et al., 2016) and at 12 months (Hallett et al., 2009). Nineteen studies involved adults, five involved adolescents and two involved children, ranging in mean age from eight to 69.2 years.

3.2.2 Externalising outcomes.

Twelve studies measured an externalising outcome of ADHD and/or conduct symptoms, involving 44,115 participants. Eleven studies measured an association between ATs and ADHD symptoms. Two studies measured an association between ATs and conduct symptoms. No studies measured an association between ATs and oppositional symptoms or traits. Across studies, 11 were cross-sectional and one was longitudinal in design (Wallace et al., 2012), albeit the latter was variable in length and in number of adolescents followed up, see ‘3.5.2 Externalising: conduct symptoms’. Seven studies involved adults, two involved adolescents and three involved children, ranging from 7.9 to 41.5 years.
3.3 Measures

3.3.1 AT measures.

Although all studies included in the meta-analysis used a validated, self- or parent-report questionnaire measure to assess ATs, the measures varied. Across all measures, a higher score indicated higher severity of ATs. The measures in the included studies are described in Table 3. Twenty-eight studies used a self-report measure for adolescent and/or adult samples, four studies used a parent-report measure for child samples, and one administered an interview. Twenty-three studies used the Autism Quotient-50 item (AQ-50; Baron-Cohen et al., 2001) and two studies used the shorter self-report Autism Quotient-Short to reduce participant burden (AQ-Short; Hoekstra et al., 2011). Only one study used a psychiatric telephone interview, a 12-item measure yielding a single DSM-IV score for ATs from the parent-report Autism-Tics And Comorbidities interview (A-TAC; Hansson et al., 2005).

3.3.2 Internalising measures.

Nineteen studies used a self- or parent-report anxiety symptom measure. 16 studies reported on the total score, of which eight reported on a general measure of trait anxiety, eight reported on a measure of social anxiety, two reported on a measure of OCD, and two reported on a measure of Generalised Anxiety Disorder (GAD; APA, 2013). Two studies reported on more than one anxiety outcome using two or three different measures (Liew, Thevaraja, Hong & Magiati, 2015; Zhou, Wang & Chasson, 2018).

Studies used the following measures: Liebowitz Social Anxiety Scale (n=3; LSAS; Liebowitz, 1987), Beck Anxiety Inventory (n=2; BAI; Beck, Epstein, Brown &
Steer, 1988), State Trait Anxiety Inventory (n=2; STAI; Spielberger, Gorsuch & Lushene, 1970); Social Phobia and Anxiety Inventory (n=2; SPAI- 23; Turner et al., 1989, Social Interaction Anxiety Scale (n=2; SIAS; Mattick & Clarke, 1989), Zung Self-Rating Anxiety Scale (n=2; SRAS; Zung, 1971), Obsessive Compulsive Inventory-Revised (n=2; OCI-R; Foa et al., 2002), Anxiety Related Behaviours Questionnaire (n=1; ARBQ; Eley et al., 2003), Penn State Worry Questionnaire (n=1; PSWQ; Meyer et al., 1990), 6-item A-TAC DSM-IV score for GAD (n=1; Hansson et al., 2005); anxiety symptom subscale from the Depression Anxiety & Stress Scale-21 item (n=3; DASS-21; Lovibond & Lovibond, 1995).

16 studies reported on a self- or parent-report depressive symptom measure, one of which reported on post-natal depression symptoms included in the meta-analysis due to high symptom overlap (APA, 2013). 11 studies reported on the total score of a measure using the following measures: Beck Depression Inventory-II (n=4; BDI-II; Beck, Steer & Brown, 1996), Centre for Epidemiological Studies-Depression Scale (n=3; CES-D or CESD-Revised; Carpenter et al., 1993; Eaton et al., 2004), Inventory of Depressive Symptomatology-Self Report (n=2; IDS-SR; Rush, Carmody & Reimitz, 2000), Patient Health Questionnaire–Nine Item Version (n=1; PHQ-9; Kroenke, Spitzer & Williams, 2001), Edinburgh Postnatal Depression Scale (n=1; EPDS; Cox, Holden & Sagovsky, 1987). Five studies reported on a depressive symptom subscale score for a combined depression/anxiety symptom measure, including the DASS-21 (n=3; Lovibond & Lovibond, 1995), Inventory of Depression and Anxiety Symptoms (n=1; IDAS; Watson et al., 2007), Hospital Anxiety and Depression Scale (n=1; HADS; Zigmand & Snaith, 1983).

3.3.3 Externalising measures

Ten studies reported on a self- or parent-report measure of ADHD symptoms. Eight studies used the total score of a measure and two studies reported
two correlation coefficients on the inattentive and hyperactive/impulsive subscales respectively of the Adult ADHD Self-Report Scale (ASRS; Kessler et al., 2005), and Conners Parent Rating Scale (CPRS; Conners, 1998). For these two studies the average of the two correlations per study was taken so as not to double-enter the sample nor privilege one symptom domain over another.

The other studies used the following measures: self- or parent-report versions of the Conner’s ADHD Ratings Scales \((n=4;\) Conners, Erhardt & Sparrow, 1999), ASRS \((n=2;\) Kessler et al., 2003), ADHD Rating Scale \((n=1;\) ADHD-RS; Reid et al., 1998), 18-item A-TAC DSM-IV score for ADHD \((n=1;\) Hansson et al., 2005), a validated 12-item self-report measure of DSM–IV ADHD symptoms \((n=1;\) Reiersen et al., 2008), SSAGA-OZ interview comprising nine items derived from the DSM-IV criterion for ADHD \((n=1;\) Lynskey et al., 2012). Two studies reported on a parent-report measure of conduct-like symptoms. One study used the total score from the Antisocial Process Screening Device which screens for antisocial behaviour traits \((APSD;\) McMahon et al., 2010), and the other study derived a five-item measure yielding a total DSM-IV score for conduct disorder symptoms from the A-TAC \((Hansson et al., 2005)\).

3.4 Risk of bias

The mean score across the 35 studies was 1.5 stars \((SD=1.1)\) out of nine, ranging from zero to five stars. Table 4 summarises the risk of bias scores for each study on the full NOS (Wells et al., 2014). After scoring, the reviewers reflected the NOS was not ideally suited to capturing and comparing variability in study quality. Some items were inappropriately suited to the design of studies taking a dimensional approach, i.e., the vast majority were not comparing an exposed and unexposed cohort. Therefore, five of eight items were retained on the basis of face validity: Selection; items one and four, Outcome; items one, two and three. The following items were excluded: for ‘Selection’; items two and four as these related to
selection of a non-exposed cohort and demonstrated that autism/mental health disorder diagnosis was not present; and for ‘Comparability’; item one as this related to comparability of exposed and non-exposed cohorts. On the revised NOS, the mean score was one star (SD=.6) out of five, ranging from zero to three stars.

Problems with study quality included how representative the sample was of the general population (15 of 35 studies comprised university student samples), a predominance of cross-sectional studies, and four studies with comparatively small sample sizes (n<100). Two studies had selection bias, selectively sampling female participants (Amos et al., 2018), and participants scoring low and high on the AQ-50 (25% percentile rank and below, and 75% percentile rank and above; Kunihara et al., 2006). One study had a higher proportion (15.7%) scoring above the validated cut-off on the AT measure (Pelton et al., 2017), compared to under 8.5% in other studies (n=8). Most studies were rated lower for use of self-report.

3.5 Meta-analyses

3.5.1 Internalising: anxiety symptoms

Data was available for 10 studies reporting on 11,576 participants. Figure 2 illustrates a forest plot of the back-transformed correlation between ATs and trait anxiety symptoms, showing an overall positive correlation of .35 (95% CI=.25-.44, p<.0001) which is a medium effect size (Cohen, 1977), suggesting that a higher severity of ATs was associated with a higher severity of anxiety symptoms. Only one study found a non-significant association (Geurts et al., 2016). There was evidence of significant and very large heterogeneity between studies, $\hat{\tau}^2$=96.56%, 95% CI=92.10-99.01, with $Q=208.70$, $df=9$, $p<.0001$ meaning that the variability across the effect sizes was considerably greater than would be expected from sampling error alone (Lipsey & Wilson, 2001).
3.5.2 Internalising: social anxiety symptoms

Data was available for eight studies reporting on 4,002 participants. Figure 3 illustrates a forest plot of the back-transformed correlation between ATs and social anxiety symptoms, showing an overall positive correlation of .44 (95% CI=.31-.55, \( p<.0001 \)), a medium to large effect size (Cohen, 1977), suggesting that a higher severity of ATs was associated with a higher severity of social anxiety symptoms. Only one study found a non-significant association (Lamport & Zlomke, 2014). Again, there was evidence of significant and very large heterogeneity between studies, \( I^2=94.17\% \), 95% CI=85.23-98.64, with variability greater than expected from sampling error (\( Q =66.77 \), \( df =7 \), \( p<.0001 \)).

3.5.3 Internalising: depression symptoms

Data was available for 16 studies reporting on 10,932 participants. Figure 4 illustrates a forest plot of the back-transformed correlation of .33 (95% CI=.27-.39, \( p<.0001 \)) which is a medium effect size (Cohen, 1977), suggesting that a higher severity of ATs was associated with a higher severity of depressive symptoms. Two studies found non-significant associations (Geurts et al., 2013; Domes et al., 2016). Again, there was evidence of significant and very large heterogeneity between studies, \( I^2=90.33\% \), 95% CI=80.33-95.99, with variability greater than expected from sampling error (\( Q =160.85 \), \( df =15 \), \( p<.0001 \)).

3.5.4 Externalising: ADHD symptoms

Data was available for nine studies reporting on 10,427 participants. Figure 5 illustrates a forest plot of the back-transformed correlation of .34 (95% CI=.21-.45, \( p<.0001 \)), a medium effect size (Cohen, 1977), suggesting that a higher severity of ATs was associated with a higher severity of ADHD symptoms. Three studies found non-significant associations (Geurts et al., 2013; Naajen et al., 2017; Panagiotidi et
al., 2017). Again, there was evidence of significant and very large heterogeneity between studies, $I^2=98.50\%$, 95% CI=96.54-99.59, with variability greater than expected from sampling error ($Q=405.00$, $df=8$, $p<.0001$).

### 3.6 Moderation analyses

Heterogeneity among the studies was explored using informal categorical analysis as there were fewer than 20 effect sizes and therefore insufficient power for a meaningful meta-regression (as used by Wykes et al., 2011).

#### 3.6.1 Study design

Studies included in each meta-analysis were grouped into two subgroups according to whether the study design was cross-sectional or longitudinal. Analyses were only possible for internalising symptoms of anxiety and depression due to sole use of cross-sectional studies included in the other meta-analyses. Table 5 presents the findings. For anxiety symptoms, study design did not significantly influence the average transformed correlation ($Q_M=1.55$, $df=1$, $p=.21$). For depression symptoms, study design significantly influenced the average transformed correlation ($Q_M=5.73$, $df=1$, $p=.02$); there was a significant effect for cross-sectional studies ($n=14$), transformed effect size =.37, 95% CI =.310.43, $p<.001$.

#### 3.6.2 Sample age

Studies included in each meta-analysis were grouped into two subgroups according to the mean age of the sample; childhood and adolescence (0-19.9 years), or adulthood (20 years and above) (WHO, 2019). Analyses were not possible using a-priori specified age categories due to few studies of children and adolescents, therefore childhood and adolescence was collapsed into one category. Age of participants did not significantly influence any of the average transformed
correlations (anxiety: $Q_M = 1.55, df = 1, p = .22$; social anxiety: $Q_M = .02, df = 1, p = .88$; depression: $Q_M = 1.10, df = 1, p = .30$; ADHD: $Q_M = .06, df = 1, p = .82$).

3.7 Publication bias

Publication bias was visually explored via funnel plots of the back-transformed estimated effect size for each random-effects model (see Figure 6) (Sterne et al., 2011). All plots showed possible asymmetry, with ATs and depression symptoms appearing more symmetrical. All plots showed considerable horizontal scatter, likely representing the very high heterogeneity between studies.

A weighted regression test for funnel plot asymmetry (with standard error as predictor) found evidence of significant asymmetry only for ATs and ADHD symptoms, ($t = -2.44, df = 7, p = .05$) suggesting possible publication bias reflecting a lack of publication of null findings. There was no significant asymmetry evidence for the other funnel plots confirming publication bias was unlikely (anxiety: $t = 1.15, df = 8, p = .28$; social anxiety: $t = -0.78, df = 6, p = .67$; depression: $t = .91, df = 14, p = .38$).

However, tests for funnel plots asymmetry lack power to distinguish chance from real asymmetry at 10 or fewer studies, particularly in presence of substantial heterogeneity, therefore they should be interpreted with caution (Sedgwick, 2013).

3.8 Narrative synthesis

3.8.1 Internalising: other anxiety symptoms

Five studies reported on other anxiety symptoms additionally to reporting on anxiety, social anxiety symptoms or other externalising symptoms and so are briefly synthesised. For generalised anxiety symptoms, three studies involving 27,720 children and adults reported significant positive correlation between ATs and generalised anxiety of $r = .34$ ($p < .001$; Liew et al., 2015), $r = .33$ ($p$ value not
reported; Lundström et al.; 2011a) and \( r =.17 \) (\( p \) value not reported; Lundström et al.; 2011b). For generalised anxiety symptoms, the two Swedish twin study cohorts reported by Lundström et al. (2011a; 2011b) comprise the majority (\( n=10,773 \) and \( n=16,695 \), respectively).

For obsessive-compulsive symptoms, two studies involving 2860 adults reported a significant positive correlation between ATs and obsessive-compulsive symptoms of \( r =.33 \) (\( p <.01 \); Zhou et al., 2018) and \( r =.34 \) (\( p <.001 \); Liew et al., 2015), a medium effect size suggesting increased severity of ATs was associated with increased severity of obsessive-compulsive symptoms (Cohen, 1977). This was very similar to the estimated effect size from the meta-analyses of anxiety symptoms.

3.8.2 Internalising: depression symptoms

One study reported on an association between AT severity and severity of depressive symptoms but was not included in the meta-analysis due to the nature of the AT measure and is therefore briefly summarised here (Lundström et al., 2011b). Depressive traits were measured by the 11-item CES-D (Carpenter et al., 1998). A significant positive correlation of \( r =.31 \) (\( p \) value not reported) was reported for 16,695 twin adults from the general population, a very similar medium raw effect size compared to the estimated effect size from the meta-analysis (Cohen, 1977).

3.8.3 Externalising: conduct symptoms

Two studies reported associations between AT severity and severity of conduct disorder symptoms involving 11,071 child and adolescents. The majority (\( n=10,773 \)) came from the Child and Adolescent Twin Study in Sweden (CATSS) twin cohort assessed at nine or 12 years old (Lundström et al., 2011a), with 298 adolescents with a mean age of 14.7 years from Wallace et al., (2012).
studies reported a positive correlation between AT severity and severity of conduct symptoms of $r = .36$ ($p$ value not reported; Lundström et al., 2011a) and $r = .39$, $p < .001$ (Wallace et al., 2012). This is an encouragingly similar medium effect size despite much heterogeneity between sample size, design and trait measures (Cohen, 1977), suggesting a trend whereby increased severity of ATs was associated with increased severity of conduct symptoms.
Table 1

Characteristics of included studies measuring the association between ATs and internalising outcome(s)

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>N</th>
<th>Mean Age(SD)</th>
<th>% Male</th>
<th>Design</th>
<th>AT Measure</th>
<th>Outcome(s); Measure(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amos et al. (2018)</td>
<td>Adult</td>
<td>458</td>
<td>30.6(12.9)</td>
<td>50.0</td>
<td>Cross-sectional</td>
<td>BAPQ</td>
<td>Anxiety, depression; Depression Anxiety &amp; Stress Scale (DASS-21).</td>
</tr>
<tr>
<td>Asano et al. (2014)</td>
<td>Adult</td>
<td>841</td>
<td>30.9(NR)</td>
<td>100.0</td>
<td>Longitudinal</td>
<td>BPASS</td>
<td>Postnatal depression; Edinburgh Postnatal Depression Scale (EPDS).</td>
</tr>
<tr>
<td>Dickter et al. (2018a)</td>
<td>Adolescent</td>
<td>104</td>
<td>19.3(1.3)</td>
<td>44.5</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Social anxiety; Social Phobia and Anxiety Inventory (SPAI-23).</td>
</tr>
<tr>
<td>(Study 1, Dickter et al., 2018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dickter et al. (2018b)</td>
<td>Adolescent</td>
<td>195</td>
<td>18.9(1.3)</td>
<td>40.7</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Social anxiety; SPAI-23.</td>
</tr>
<tr>
<td>(Study 2, Dickter et al., 2018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doherty et al. (2017)</td>
<td>Adult</td>
<td>37</td>
<td>NR(NR)</td>
<td>43.2</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Social anxiety; Liebowitz Social Anxiety Scale (LSAS).</td>
</tr>
<tr>
<td>Domes et al. (2016)</td>
<td>Adult</td>
<td>55</td>
<td>45.8(11.2)</td>
<td>58.2</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Depression; Beck Depression Inventory-II (BDI-II).</td>
</tr>
<tr>
<td>Study</td>
<td>Age</td>
<td>N</td>
<td>Mean (SD)</td>
<td>Crossover</td>
<td>Scale</td>
<td>Test Measure</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------</td>
<td>-----</td>
<td>-----------</td>
<td>-----------</td>
<td>-------</td>
<td>-------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Ebrahimi et al (2017)</td>
<td>Adult</td>
<td>336</td>
<td>22.7(3.7)</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Anxiety, depression; Beck Anxiety Inventory (BAI), BDI-II.</td>
<td></td>
</tr>
<tr>
<td>Freeth et al. (2012)</td>
<td>Adult</td>
<td>1325</td>
<td>20.1(3.3)</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Social anxiety; LSAS.</td>
<td></td>
</tr>
<tr>
<td>Geurts et al. (2016)</td>
<td>Adult</td>
<td>114</td>
<td>69.2(6.5)</td>
<td>Longitudinal</td>
<td>AQ-28</td>
<td>Anxiety, depression; BAI-21 item, Inventory of Depressive Symptomatology Self Report (ISD-SR).</td>
<td></td>
</tr>
<tr>
<td>Hallett et al. (2009)</td>
<td>Child</td>
<td>3233</td>
<td>8-9(NR)</td>
<td>Longitudinal</td>
<td>CAST</td>
<td>Anxiety; Anxiety Related Behaviours Questionnaire.</td>
<td></td>
</tr>
<tr>
<td>Horibe et al. (2018)</td>
<td>Adult</td>
<td>294</td>
<td>20.1(2.8)</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Depression; BDI-II.</td>
<td></td>
</tr>
<tr>
<td>Jackson et al. (2016)</td>
<td>Adult</td>
<td>230</td>
<td>21.3(2.5)</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Depression; BDI-II.</td>
<td></td>
</tr>
<tr>
<td>Kitzakoe et al. (2014)</td>
<td>Adolescent</td>
<td>1479</td>
<td>18.2(NR)</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Social anxiety;</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Age</td>
<td>Sample Size</td>
<td>Mean (SD)</td>
<td>Study Type</td>
<td>Scale</td>
<td>Measure</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----</td>
<td>-------------</td>
<td>-----------</td>
<td>----------------</td>
<td>-------</td>
<td>-------------------------------</td>
<td></td>
</tr>
<tr>
<td>Kunihara et al. (2006)</td>
<td>Adolescent</td>
<td>96</td>
<td>19.3(2.5)</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Anxiety, depression; STAI, ISD-SR.</td>
<td></td>
</tr>
<tr>
<td>Liew et al. (2015)</td>
<td>Adult</td>
<td>252</td>
<td>20.6(1.7)</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Social anxiety, depression; SIAS, Inventory of Depression and Anxiety Symptoms (IDAS).</td>
<td></td>
</tr>
<tr>
<td>Lundström et al. (2011a)</td>
<td>Child</td>
<td>10733</td>
<td>NR(NR)</td>
<td>Cross-sectional</td>
<td>A-TAC</td>
<td>Anxiety; 6 items for Generalised Anxiety Disorder (GAD) from A-TAC.</td>
<td></td>
</tr>
<tr>
<td>(CATSS cohort)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lundström et al. (2011b)</td>
<td>Adult</td>
<td>16695</td>
<td>NR(NR)</td>
<td>Cross-sectional</td>
<td>DSM-IV score</td>
<td>Anxiety, depression; 6 DSM-IV items for GAD; Centre for Epidemiologic Studies Depression Scale (CES-D).</td>
<td></td>
</tr>
<tr>
<td>(STAGE cohort)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oliver at al. (2016)</td>
<td>Adult</td>
<td>90</td>
<td>21.7(3.2)</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Anxiety; STAI.</td>
<td></td>
</tr>
<tr>
<td>Pelton et al. (2017)</td>
<td>Adult</td>
<td>163</td>
<td>21.6(3.0)</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Depression;</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Age</td>
<td>Sample Size</td>
<td>Mean (SD) Anxiety</td>
<td>Mean (SD) Depression</td>
<td>Design</td>
<td>AQ Scale</td>
<td>Additional Measures</td>
</tr>
<tr>
<td>------------------</td>
<td>--------</td>
<td>-------------</td>
<td>-------------------</td>
<td>----------------------</td>
<td>-----------------</td>
<td>----------</td>
<td>---------------------------------------------------------</td>
</tr>
<tr>
<td>Mealey et al. (2014)</td>
<td>Adult</td>
<td>144</td>
<td>25.3 (7.7)</td>
<td>39.6</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Anxiety, depression; DASS-21.</td>
</tr>
<tr>
<td>Reed et al. (2016)</td>
<td>Adult</td>
<td>413</td>
<td>21.4 (3.4)</td>
<td>40.7</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Social anxiety, depression; LSAS, HADS.</td>
</tr>
<tr>
<td>Rosbrook et al. (2010)</td>
<td>Adolescent</td>
<td>231</td>
<td>18.9 (2.8)</td>
<td>49.4</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Anxiety, depression; Zung Self-Rating Anxiety Scale (SRAS), CES-D.</td>
</tr>
<tr>
<td>Xu et al. (2015)</td>
<td>Adult</td>
<td>3229</td>
<td>20.5 (1.6)</td>
<td>58.0</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Anxiety, depression; SAS, CES-D.</td>
</tr>
<tr>
<td>Zhou et al. (2018)</td>
<td>Adult</td>
<td>2608</td>
<td>20.6 (1.1)</td>
<td>51.3</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>Anxiety, obsessive compulsive symptoms, depression; SAS, OCI-R.</td>
</tr>
</tbody>
</table>

**Note.** A-TAC = Autism-Tics, ADHD And Co-morbidities Inventory, AQ-50 = Autism Spectrum Quotient-50 item, AQ-28 = Autism Spectrum Quotient-28 item, BAPQ = Broad Autism Phenotype Questionnaire, SCQ = Social Communication Questionnaire, SRS = Social Responsiveness Scale, CATSS cohort = Child and Adolescent Twin Study in Sweden cohort, STAGE cohort = Study of Twin Adults: Genes and Environment. NR = Not Reported.
Table 2

Characteristics of included studies measuring the association between ATs and externalising outcome(s)

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>N</th>
<th>Mean Age (SD)</th>
<th>% Male</th>
<th>Design</th>
<th>AT Measure</th>
<th>Outcome(s); Measure(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Christ et al. (2010)</td>
<td>Adolescent</td>
<td>1847</td>
<td>18.4(1.0)</td>
<td>37.3</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>ADHD; 12 DSM-IV ADHD items.</td>
</tr>
<tr>
<td>DeAlwis et al. (2013)</td>
<td>Adult</td>
<td>3080</td>
<td>27-40(NR)</td>
<td>NR</td>
<td>Cross-sectional</td>
<td>SRS</td>
<td>ADHD; SSAGA-OZ interview (9 DSM-IV ADHD items).</td>
</tr>
<tr>
<td>Geurts et al. (2013)</td>
<td>Adult</td>
<td>85</td>
<td>21.5(2.4)</td>
<td>37.6</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>ADHD; ADHD Rating Scale (ADHD-RS)</td>
</tr>
<tr>
<td>Horibe et al. (2018)</td>
<td>Adult</td>
<td>294</td>
<td>20.1(2.8)</td>
<td>33.0</td>
<td>Cross-sectional</td>
<td>AQ-50</td>
<td>ADHD (inattention only); Adult ADHD Self-Report Scale (ASRS).</td>
</tr>
<tr>
<td>Lundström et al. (2011a)</td>
<td>Child</td>
<td>10733</td>
<td>9-12(NR)</td>
<td>51.0</td>
<td>Cross-sectional</td>
<td>A-TAC</td>
<td>ADHD, conduct problems; A-TAC ADHD DSM-IV score, 5 DSM-IV items from A-TAC.</td>
</tr>
<tr>
<td>(CATSS cohort)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lundström et al. (2011b)</td>
<td>Adult</td>
<td>16695</td>
<td>NR(NR)</td>
<td>40.0</td>
<td>Cross-sectional</td>
<td>DSM-IV score</td>
<td>ADHD; 18 items similar to A-TAC algorithm.</td>
</tr>
<tr>
<td>(STAGE cohort)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Age Group</td>
<td>Sample Size</td>
<td>Mean (SD)</td>
<td>Cross-sectional</td>
<td>Measure</td>
<td>Subscale(s)</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------</td>
<td>-----------------</td>
<td>---------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Naajen et al. (2017)</td>
<td>Child</td>
<td>922</td>
<td>10.9(2.8)</td>
<td>87.5</td>
<td>SCQ</td>
<td>ADHD; Conners’ Parent Rating Scale (inattention, hyperactivity/impulsivity subscales).</td>
<td></td>
</tr>
<tr>
<td>Panagiotidi et al., (2017)</td>
<td>Adult</td>
<td>276</td>
<td>34.0(13.6)</td>
<td>25.7</td>
<td>AQ-50</td>
<td>ADHD; Adult ADHD Self-Report Scale (ASRS-v1.1).</td>
<td></td>
</tr>
<tr>
<td>Park et al. (2017)</td>
<td>Adult</td>
<td>3170</td>
<td>32.3(2.5)</td>
<td>NR</td>
<td>SRS</td>
<td>ADHD; ASRS (modified version).</td>
<td></td>
</tr>
<tr>
<td>Polderman et al. (2013)</td>
<td>Adult</td>
<td>559</td>
<td>41.5(11.3)</td>
<td>44.0</td>
<td>AQ-28</td>
<td>ADHD; CAARS-S ADHD index score.</td>
<td></td>
</tr>
<tr>
<td>Ronald et al. (2008)</td>
<td>Child</td>
<td>6107</td>
<td>7.9(0.5)</td>
<td>NR</td>
<td>CAST</td>
<td>ADHD; Revised Conners’ Parent Rating Scale (CPRS-R).</td>
<td></td>
</tr>
</tbody>
</table>

Table 3

Summary of measures used to assess autistic traits (ATs) in the included studies

<table>
<thead>
<tr>
<th>Study(s)</th>
<th>AT measure</th>
<th>Description of AT measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Christ et al. (2010), Dickter et al. (2018a,b), Doherty et al. (2017), Domes et al. (2016), Ebrahimi et al. (2017), Geurts et al. (2013), Freeth et al. (2012), Horibe et al. (2018), Jackson et al. (2016), Keenan et al. (2018), Kitzakoe et al. (2014), Kunihara et al. (2006), Lamport et al. (2014), Liew et al. (2015), Oliver et al. (2016), Panagiotidi et al. (2017), Pelton et al. (2017), Mealey et al. (2014), Reed et al. (2016), Rosbrook et al. (2010), Xu et al. (2015), Zhou et al. (2018)</td>
<td><strong>Autism Quotient 50 item (AQ-50)</strong>, Baron-Cohen et al., (2001).</td>
<td>Three AQ versions exist for: children, adolescents and adults. The 50-item AQ is a self-report measure structured around five subdomains (subscales): social interaction, communication, attention to detail, attention switching, and imagination. Items are rated on a 4-point Likert scale (1=definitely agree, 2=slightly agree, 3=slightly disagree, and 4=definitely disagree). A higher total score denotes higher AT severity. Cut-off score of 76 or greater indicates a positive screen for ASD and need for further evaluation.</td>
</tr>
<tr>
<td>Geurts et al. (2016), Polderman et al. (2013)</td>
<td><strong>Autism Quotient-Short (AQ-S;AQ28)</strong>, Hoekstra et al., (2011).</td>
<td>28-item abridged version of the AQ-50 designed to reduce participant burden. Composed of five subscales: Social Skills, Routine, Attentional Switching, Imagination and Numbers and Patterns. Items are rated on a 4-point Likert scale as per the AQ-50. A higher total score denotes higher AT severity. Cut-off score of 65 indicates a positive screen for ASD and need for further evaluation.</td>
</tr>
</tbody>
</table>
| Amos et al. (2018) | **Broad Autism Phenotype Questionnaire (BAPQ)**, Hurley et al. (2007). | 36-item self-report measure divided into three subscales to assess the major aspects of ATs (social aloofness, rigidity, and pragmatic language) in non-autistic relatives of individuals with ASD and in general population samples. Items are rated on a 6-point Likert scale (from 1=very rarely,
to 6=very often). Participants consider interactions they have had with people in general rather than special relationships, immediate family members or their partner, and to consider their behaviour across their entire adult life. Total score ranges from 36 to 216, with a higher score denoting higher AT severity. Originally designed to be completed by self-report and informant-report from which a best estimate score of BPAQ is calculated and interpreted in line with a cut-off score.

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Scale/Measurement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asano et al. (2014)</td>
<td><strong>Broader Phenotype Autism Symptom Scale (BPASS), Dawson et al., (2007).</strong></td>
<td>Eleven item measure of ATs assessed via face-to-face administration, comprising four domains (subscales): motivation (social interest in peers and groups), expressiveness (nonverbal social communication) conversational skills (clinical observations of conversation skills), flexibility/range of interests. Items scores on Likert-type scales ranging from 1 to 5, 1 to 4, or 1 to 3 (item-dependent). Yields a composite broad autistic phenotype score (subscale totals summed to calculate a total score), with a higher composite score denoting higher degree of impairment.</td>
</tr>
<tr>
<td>Hallett et al. (2009),</td>
<td><strong>Child Autism Screening Test (CAST), Scott et al., (2002).</strong></td>
<td>7-item parent-report rating scale of behavioural indicators of possible Asperger Syndrome and related social and communication conditions in 4-11 year olds. Items are scored as present or absent. Results are reported based on an overall score. Cut-off score of 15 or greater indicates a need for further evaluation.</td>
</tr>
<tr>
<td>Ronald et al. (2008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lundström et al. (2011a,b)</td>
<td><strong>A-TAC score, Hansson et al., (2005).</strong></td>
<td>12-item measure yielding a single DSM-IV score from the Autism-Tics And Comorbidities (A-TAC), a validated psychiatric parent-report telephone</td>
</tr>
</tbody>
</table>
### Social Responsiveness Scale (SRS)

**Wallace et al. (2012)**

65-item parent-report measure assessing severity of autism symptoms as they occur in natural social settings over the past six months in 4-18 year olds, across three DSM-IV autistic symptom domains (social impairment, communication impairment, and stereotyped/ repetitive behaviours). Primarily focuses on social behaviour traits. Items are rated on a 4-point Likert scale (from “not true” to “almost always true”). Yields a raw total and gender-normed T-score (intended to correct gender differences observed in normative samples), with a higher score denoting higher degree of impairment.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>65-item parent-report measure assessing severity of autism symptoms as they occur in natural social settings over the past six months in 4-18 year olds, across three DSM-IV autistic symptom domains (social impairment, communication impairment, and stereotyped/ repetitive behaviours). Primarily focuses on social behaviour traits. Items are rated on a 4-point Likert scale (from “not true” to “almost always true”). Yields a raw total and gender-normed T-score (intended to correct gender differences observed in normative samples), with a higher score denoting higher degree of impairment.</td>
</tr>
</tbody>
</table>

### 11-item measure derived from SRS

**DeAlwis et al. (2013), Park et al. (2017)**

11-item parent-report measure assessing ATs, modified from SRS to reduce participant burden (items were selected on the basis they had strong loadings on the first unrotated factor of a principal components analysis of the SRS in a paediatric sample; Constantino et al., 2000). Most items relate to reciprocal social interaction, but others measure stereotyped/ repetitive behaviours or communication impairment. Items are rated on a 4-point scale (0 = false, not at all true; 1=slightly true; 2=mainly true; 3=very true). Total score ranges from 0-33, with a higher score denoting higher AT severity.

<table>
<thead>
<tr>
<th>DeAlwis et al. (2013), Park et al. (2017)</th>
<th>11-item measure derived from SRS, Reiersen et al. (2008).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11-item parent-report measure assessing ATs, modified from SRS to reduce participant burden (items were selected on the basis they had strong loadings on the first unrotated factor of a principal components analysis of the SRS in a paediatric sample; Constantino et al., 2000). Most items relate to reciprocal social interaction, but others measure stereotyped/ repetitive behaviours or communication impairment. Items are rated on a 4-point scale (0 = false, not at all true; 1=slightly true; 2=mainly true; 3=very true). Total score ranges from 0-33, with a higher score denoting higher AT severity.</td>
</tr>
</tbody>
</table>

### Social Communication Questionnaire (SCQ)

**Naajen et al. (2017)**

40-item parent-report measure assessing autistic symptom severity based on the three DSM-IV ASD domains for four years and older: social relating, communication, and domains of interest.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40-item parent-report measure assessing autistic symptom severity based on the three DSM-IV ASD domains for four years and older: social relating, communication, and domains of interest.</td>
</tr>
</tbody>
</table>
Comprises two forms: the Lifetime form which focuses on developmental history and behaviour, and the Current form which focuses on behaviour during the past three months. Items are scored as yes/no. Lifetime form yields a total score, with a cut-off of 15 indicative of a positive screen for ASD and a need for further assessment.
<table>
<thead>
<tr>
<th>Study</th>
<th>Selection</th>
<th>Comparability</th>
<th>Outcome</th>
<th>Total score (stars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amos et al. (2018)</td>
<td>★★★☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
<tr>
<td>Asano et al. (2014)</td>
<td>★☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>2</td>
</tr>
<tr>
<td>Christ et al. (2010)</td>
<td>★★★☆☆☆</td>
<td>★☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>5</td>
</tr>
<tr>
<td>DeAlwis et al. (2013)</td>
<td>★★★☆☆☆</td>
<td>★☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>5</td>
</tr>
<tr>
<td>Dickter et al. (2018a,b)</td>
<td>★★★☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>2</td>
</tr>
<tr>
<td>Doherty et al. (2017)</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>0</td>
</tr>
<tr>
<td>Domes et al. (2016)</td>
<td>★★★☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>2</td>
</tr>
<tr>
<td>Ebrahimi et al. (2017)</td>
<td>★★★☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
<tr>
<td>Freeth et al. (2012)</td>
<td>★★★☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>2</td>
</tr>
<tr>
<td>Geurts et al. (2013)</td>
<td>☆☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>0</td>
</tr>
<tr>
<td>Geurts et al. (2016)</td>
<td>★★★☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>3</td>
</tr>
<tr>
<td>Hallett et al. (2009)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>3</td>
</tr>
<tr>
<td>Horibe et al. (2018)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
<tr>
<td>Jackson et al. (2016)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
<tr>
<td>Keenan et al. (2018)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
<tr>
<td>Kitzakoe et al. (2014)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
<tr>
<td>Kunihara et al., (2006)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
<tr>
<td>Lamport et al. (2014)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
<tr>
<td>Liew et al. (2015)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
<tr>
<td>Lundström et al. (2011a,b)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>2</td>
</tr>
<tr>
<td>Mealey et al. (2014)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
<tr>
<td>Naajen et al. (2017)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>2</td>
</tr>
<tr>
<td>Oliver et al. (2016)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
<tr>
<td>Panagiotidi et al. (2017)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
<tr>
<td>Park et al. (2017)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
<tr>
<td>Pelton et al. (2017)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
<tr>
<td>Polderman et al. (2013)</td>
<td>★☆☆☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>☆☆☆☆☆</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 2. Forest plot of 10 studies examining the correlation between ATs and anxiety symptoms showing random-effect model effect sizes, confidence intervals, and weight of each study.
Figure 3. Forest plot of eight studies examining the correlation between ATs and social anxiety symptoms showing random-effect model effect sizes, confidence intervals, and weight of each study.
**Figure 4.** Forest plot of 16 studies examining the correlation between ATs and depression symptoms showing random-effect model effect sizes, confidence intervals, and weight of each study.
Figure 5. Forest plot of nine studies examining the correlation between ATs and ADHD symptoms showing random-effect model effect sizes, confidence intervals, and weight of each study.
Table 5

*Effect of study design on AT and mental health outcome association*

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Studies (N)</th>
<th>Transformed ES</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cross-sectional studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>8</td>
<td>.40</td>
<td>.28-.51</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Depression</td>
<td>14</td>
<td>.37</td>
<td>.31-.43</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Longitudinal studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>2</td>
<td>.23</td>
<td>-.01-.47</td>
<td>0.07</td>
</tr>
<tr>
<td>Depression</td>
<td>2</td>
<td>.15</td>
<td>-.03-.32</td>
<td>0.10</td>
</tr>
</tbody>
</table>

*Note. Transformed ES = transformed effect size (r), 95% CI = 95% Confidence Interval.*

Table 6

*Effect of sample age on AT and mental health outcome association*

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Studies (N)</th>
<th>Transformed ES</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Childhood and adolescence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>3</td>
<td>.38</td>
<td>.17-.58</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ADHD</td>
<td>3</td>
<td>.37</td>
<td>.13-.62</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Social Anxiety</td>
<td>3</td>
<td>.48</td>
<td>.22-.74</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Depression*</td>
<td>2</td>
<td>.43</td>
<td>.25-.62</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Adulthood</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>3</td>
<td>.36</td>
<td>.22-.50</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ADHD</td>
<td>6</td>
<td>.34</td>
<td>.16-.52</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Social Anxiety</td>
<td>5</td>
<td>.46</td>
<td>.25-.66</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Depression</td>
<td>14</td>
<td>.33</td>
<td>.25-.40</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*Note. ADHD = Attention Deficit Hyperactivity Disorder, Transformed ES = transformed effect size (r), 95% CI = 95% Confidence Interval.*

*For depression symptoms, only there were no childhood studies, only adolescents.*
Figure 6. Funnel plots of back-transformed correlation effect sizes: (A) AT and anxiety symptoms, (B) AT and social anxiety symptoms, (C) AT and depression symptoms, (D) AT and ADHD symptoms.
4.0 Discussion

4.1 Summary of key findings in the context of previous research

The aim of this meta-analysis was to identify the strength and direction of the association between ATs and internalising and externalising symptoms in the general population, across the lifespan. The review included studies measuring ATs via total score on a validated self-report instrument to ensure the whole autism phenotype was captured. Similarly, internalising and externalising symptoms had to be measured by a validated self-report instrument for a specific symptom cluster. Although this excluded a few large population-based twin studies (Hallett et al., 2012; Hallett et al., 2013; Tick et al., 2016), the aim was to enhance specificity and validity of the review by reducing some heterogeneity (Sharpe et al., 1997). Thirty-five studies were included, reporting on over 80,000 participants. There was a focus on internalising symptoms, particularly depressive symptoms, with over double the number of included studies measuring an association with internalising (n=25) versus externalising (n=12), which is in line with the historic neglect of externalising disorders other than ADHD (Mandy et al., 2014). Meta-analysis was not possible for studies of conduct symptoms and there were no studies meeting inclusion criteria for oppositional symptoms.

The primary finding was an overall positive correlation between ATs and internalising and externalising symptoms in the general population, with medium to medium-to-large effect sizes. This indicates that individuals in the general population with higher levels of ATs experience higher levels of internalising and/or externalising symptoms, confirming the modest phenotypic correlations between ATs and global measures of internalising and externalising in population-based twin studies of children and adolescents (Hallett et al., 2012; Tick et al., 2016). It is also consistent with systematic and selective narrative reviews of the association.
between ATs and ADHD symptoms and/or diagnosis (Hartman et al., 2016; Posthuma & Polderman, 2013; Taurines et al., 2012; Visser et al., 2016). There was no evidence of a publication bias that would threaten the validity of the findings except for the association between ATs and ADHD, although asymmetry tests were underpowered to detect this given the presence of likely conflating high heterogeneity (Sterne et al., 2011).

For the first time, this review convincingly extends the autism comorbidity literature to quantify the co-occurrence of subthreshold ATs and internalising and externalising difficulties in the general population. This supports a dimensional conceptualisation of the co-occurrence of AT and internalising and externalising symptoms, in line with evidence that ATs and internalising and externalising traits exist on a continuum (Coghill & Sonuga-Barke, 2012).

The largest effect size was for a positive correlation between ATs and social anxiety symptoms (0.44), in line with systematic and meta-analytic reviews reporting social anxiety disorder as the most prevalent comorbid anxiety disorder in adults with autism and high prevalence in autistic children and adolescents (Lugo-Marin et al., 2019; van Steensel et al., 2011; Vasa & Mazurek, 2015). Indeed, the included studies of social anxiety symptoms only included adolescents and adults, therefore a pooled effect size for this association in children remains unknown. Interestingly, the effect sizes across all other internalising and externalising symptoms were very consistent (around .34), suggesting a possible increase across internalising and externalising symptoms is associated with increased ATs (Lundstrom, 2011a; 2011b). Correlations reported by studies included in narrative syntheses were similar, with conduct symptoms a little larger (up to $r = .39$) and findings largely consistent for other anxiety symptoms ($r = .34$).

There was a large degree of heterogeneity between studies across the meta-analyses, particularly in sample size. Although heterogeneity tests were underpowered due to the relatively small number of studies, except for depressive
symptoms (Huedo-Medina, Sánchez-Meca, Marin-Martínez & Botella, 2006), this limits the validity of findings (Del Re, 2015; Sharpe et al., 1997). Factors that might account for heterogeneity were investigated. However, due to the small and unequal studies per category and therefore insufficient power, the effect of these factors was explored by categorical analysis (Wykes et al., 2011).

The majority of included studies were cross-sectional (n=31), with four longitudinal studies included, of which only one reported on externalising symptoms. Moderation analyses were only possible for internalising symptoms of anxiety and depression. Study design significantly influenced the effect size for depression but not anxiety symptoms. There was a significant group effect for cross-sectional studies for both depression and anxiety symptoms, but not for longitudinal studies, although there was a trend towards significance for longitudinal studies reporting an association between ATs and anxiety symptoms. This suggests cross-sectional studies evidence larger associations between ATs and internalising symptoms of anxiety and depression. However, this may have been due to insufficient power to detect an effect for longitudinal studies.

Surprisingly few studies were of children and adolescents; most included studies involved adults (n= 24; defined as 20 years and above). Analyses were therefore not possible using the a-priori proposed three age categories, so childhood and adolescence was collapsed into one category (0-19.9 years). Sample age did not significantly influence any effect sizes. There was an indication of a significant group effect for both childhood/adolescence and adulthood across all meta-analyses, suggesting there may be age-related differences in the strength of association, with a larger association present in childhood and adolescence than adulthood. However, there was insufficient power to detect this effect.

Some heterogeneity is likely due to the variable measures of internalising and externalising symptoms, and the use of a range of AT measures with varying brevity and therefore comprehensiveness in capturing the Broader Autism
Phenotype (BAP). Most studies (72%, n = 25/35) used the AQ-50 to measure ATs developed by Baron-Cohen and colleagues (2001), of which two studies used the shortened version (AQ28 or AQ-S; Hoekstra et al., 2011). The AQ is a screening tool to determine whether an individual without intellectual disability (ID) in the general population is at risk for autism, and to justify formal assessment. The majority use of the AQ measures is a strength of this review as they were assessed to be two of four screening tools (out of nine) possessing mostly satisfactory or intermediate values for psychometric properties by a systematic review of assessment tools for adults of mean normal IQ (Baghdadi, Russet & Mottron, 2017).

Criticisms have been levelled at the AQ-50 however, including variable convergent validity with diagnostic tools such as the ADI-R (r =0.18; Bishop & Seltzer; 2012), unconfirmed structural validity of the five-factor structure, and variable diagnostic accuracy and validity, with unsatisfactory evidence available for the AQ-S (Baghdadi et al., 2017). More confidence could be inspired by further replication of the effects sizes found in this review with one or two other good quality AT measures to protect the findings from being an artefact of one measure.

4.2 Limitations and Research Implications

The predominantly cross-sectional literature in this review has taken the first important step in quantifying correlations between ATs and internalising and externalising traits. The questions now are what does this association represent and what explains it? The correlations could represent a number of different scenarios.

One explanation could be epiphenomena due to measurement error. For example, the effect size for social anxiety may reflect higher severity of social anxiety symptoms inflating scores on an AT measure which also captures social communication and behavioural difficulties, or vice versa. Or, scores on either AT or psychopathology symptom measures are both inflated by some third unmeasured
variable, i.e., presence of non-autistic psychopathology such as depressive symptoms (e.g., withdrawal behaviour or lack of motivation to engage with social stimuli). However, evidence suggests this is unlikely, for example, a cross-lagged panel model in a longitudinal general population cohort (the Avon Longitudinal Study of Parents and Children; ALSPAC) found ATs and social anxiety were distinct and earlier severity of ATs predicted a small but significant amount of the variance in later social anxiety symptoms but not vice versa, suggesting ATs may be a risk factor for later internalising and causally linked (Pickard, Rijsdijk, Happé & Mandy, 2017).

In line with Pickard et al.’s (2017) finding, another explanation could be that ATs are a causal risk factor for some psychopathology, via different moderating and/or mediating mechanisms, known as a phenotypic interaction. For example, socioemotional competence moderates the developmental risk between oppositionality and later serious conduct problems in in mid-childhood, such that children with higher ATs (poorer social communication skills) who show more oppositional behaviour are more likely to develop later behavioural difficulties perhaps because they experience more social failure (Mandy et al., 2013). For example, a longitudinal follow-up of Hallett et al. (2009) found that earlier AT severity influenced later internalising traits and to a weaker extent vice versa, mediated by a more stressful environment – perhaps a need for rigid routine or difficulty interacting socially causes stress, and/or children with higher ATs are more likely to experience stress at school or home (Hallett et al., 2010).

In order to answer these questions and disentangle causality, well-powered longitudinal studies of traits with adequate follow-up periods in general population samples are needed (Constantino, 2018). The prospective, longitudinal bivariate or multivariate twin model-fitting method offers the most promise; these analyses assess the degree to which two sets of developmental traits share genetic and
environmental influences (Ronald, 2014; Kreuger et al., 2016). Research should begin this investigation as young as possible to trace the origins of ATs and psychopathologic traits, as a recent general population longitudinal twin-study found social communication variation at 18 months strongly predicted ATs as young as 36 months old. However, this was independent from contemporaneous ratings of behaviour problems and not predictive of later general internalising or externalising (Hawks et al. (2019). Further, social communication variation showed a distinct genetic structure, suggesting perhaps the co-occurrence of traits occurs later in child development, perhaps by subsequent interactions (Hawks et al., 2019).

An alternatively or additional explanation for trait co-occurrence may be that ATs and psychopathologies share some underlying etiology. This most likely applies to the AT and ADHD trait overlap, with these neurodevelopmental traits the most heritable (Ronald & Hoekstra, 2011), and converging evidence from large twin studies across the Western world demonstrating they share considerable but by no means complete genetic influence at both the categorical diagnosis and trait measure level (Ronald, et al., 2008; Ronald et al., 2010; Taylor et al., 2013). Research also finds distinct genetic and environmental influences specific to autism and ADHD, suggesting an additive co-occurrence of both condition-specific differences (Dworzynski et al., 2009; Ronald et al., 2006; 2011). Interestingly, the same cannot be said for ATs and anxiety, whereby similar twin-fitting models do not find a high degree of genetic overlap in childhood (Hallett et al., 2010).

A limitation of this review is the focus on adults which limits the generalisability of findings to children and younger adolescents. One contributing factor could have been the methodological decision to apply WHO definitions of adolescence and adulthood. This resulted in a somewhat arbitrary categorisation of studies, given a large proportion were cross-sectional studies of university students whose mean age fell very close to the cut-off. This predominance of university
students in itself limits generalisability. However, a paucity of studies of children remains, necessitating further research in children with specific not global measures of internalising and externalising symptoms.

It was also striking that reviewed studies did not routinely explore gender differences in trait associations. Given the biased male-to-female gender ratio in autism, and evidence of a female phenotype, this should be a priority for future research so as to increase generalisability of the findings to females in the general population (Mandy & Lai, 2017). There is emerging, albeit currently inconsistent, evidence that current measures developed with male samples may not adequately capture the female autism phenotype, particularly in the restricted behaviours and interests domain (Frazier & Hardan, 2017), and emerging evidence that the pattern of psychopathology risk associated with autism differs for males and females (Mandy et al., 2012).

The methodological quality of the studies was assessed using the NOS scale (Wells et al., 2014). Using the full NOS, study quality was low overall, with studies scoring 1.5 stars out of nine on average, ranging from one to five stars. On reflection, the NOS was limited in capturing the small amount of variability between studies due to its categorical approach in comparing exposed and unexposed cohorts which is at odds with the dimensional approach taken by studies assessing trait associations and the theoretical stance of this review. A revised five-item version retaining appropriate items yielded a similar mean score (one star out of a maximum of five, SD=.6, range= 0 to 3), however arguably captured a little more variability given the reduced maximum range. Some of the observed heterogeneity may reflect study quality, however the overall low score would have made categorical analysis of this as a moderator meaningless.

The majority of studies were also scored down for use of self-report measures versus more gold standard interview or observation methods. Given
studies were measuring subthreshold traits (not clinical diagnoses) in large population samples where interview or other methods would be costly and time-consuming, this raises an interesting question of what the best way is to measure ATs in the general population. Arguably, the overall low quality reflects the predominance of cross-sectional studies, supporting the call for more high quality longitudinal research.

4.3 Conclusions: Scientific and Clinical Implications

This review found a consistent and moderate positive association between severity of ATs and a range of internalising and externalising phenotypic traits in the general population for the first time, with the largest effect size for ATs and social anxiety symptoms. Study design had a moderating effect on the association between ATs and depressive symptoms such that larger associations were reported by cross-sectional studies. There was tentative support for age-related differences, with a trend towards larger associations in childhood and adolescence despite a paucity of studies in these life stages. Conclusions are tentative due to the high heterogeneity between studies.

Crucially, this review indicates that even at the subthreshold level, ATs are an important part of the autism constellation with the potential to contribute to our understanding of the relationships between autistic and psychopathologic traits. Dimensional AT measurement is needed in a range of settings. A research priority should be to use population-based longitudinal twin designs with the power to unpick the developmental sequence between subthreshold autistic and psychopathologic traits. Exploration of externalising behavioural traits other than ADHD and of gender differences in trait associations are desperately needed.

This review holds important implications for clinical practice, pointing to the importance of routine screening from childhood for ATs and internalising and externalising symptoms for early detection. If ATs are a risk factor or contributing
factor in the development and maintenance of internalising and/or externalising difficulties, formulation can be better informed, and treatment more targeted towards suboptimal social skills and psychological inflexibility which may have contributed to psychopathology.

5.0 References

References marked with an asterisk indicate studies included in the meta-analysis and/or narrative review.


Posthuma, D., & Polderman, T. J. (2013). What have we learned from recent twin studies about the etiology of neurodevelopmental disorders?. *Current Opinion in Neurology, 26*(2), 111-121.


Robinson, E. B., Koenen, K. C., McCormick, M. C., Munir, K., Hallett, V., Happé, F., ... & Ronald, A. (2011b). Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). *Archives of General Psychiatry, 68*(11), 1113-1121.


Part 2: Empirical Paper

Early infant candidate markers of autistic social trait developmental trajectories in the general population.
Abstract

Aims: Research has identified nonlinear, gender-specific autistic social trait (AST) developmental trajectories from childhood to mid-adolescence in the general population (Mandy et al., 2018). Females in particular show an uptick in ASTs from early adolescence. Identifying infant antecedents of AST growth can help to understand gendered AST chronogeneity. This study aimed to identify early developmental markers of AST trajectories in the first two years of life and to explore gender differences in predictors.

Method: Multi-group Latent Growth Curve Modelling (LGCM) was applied to ASTs assessed at seven, 10, 13 and 16 years by parent-report on the Social Communication Disorders Checklist (Skuse et al., 2005). Behavioural predictors were from selected measures of social communication and language, motor development and temperament at six to 24 months. Gender differences were explored using the multi-group model. Data were from 7,773 individuals from the Avon Longitudinal Study of Parents and Children (ALSPAC).

Results: Behavioural markers of social communication, language and temperament at 15 and 24 months old predicted initial AST status at seven years. Two characteristics predicted growth rate in ASTs from seven to 16 years, however no longer remained significant post-correction. Trends towards gender differences were observed, with markers more strongly associated with male AST trajectories.

Conclusions: Early infant developmental differences from the first two years of life predict childhood AST severity and are likely to differ between genders. Markers were limited in explaining AST chronogeneity from mid-childhood. Routine screening for developmental markers of ASTs from 15 months could be useful. Further research investigating gender differences in developmental markers is needed to understand the female adolescent AST uptick. Females may present differently to males in early developmental behaviour related to ASTs.
1.0 Introduction

1.1 Autism Spectrum Disorder (ASD)

Autism Spectrum Disorder (ASD) is a life-long neurodevelopmental condition characterised by atypicalities in two domains; (1) social communication and interaction, and (2) restricted, repetitive patterns of behaviours (RRB) or topics of interest (American Psychiatric Association [APA], 2013). Differences in sensory sensitivity in RRB are a core feature, and intellectual or language disability is common (APA, 2013). ASD is hereafter referred to as ‘autism’ to reflect the common rejection of ‘disorder’ by many (but not all) autistic people (Fletcher-Watson & Happé, 2019). Autism is highly heritable with varied neurobiological developmental pathways (Robinson et al., 2016). Prevalence is estimated at 1% of the global and UK populations (Baron-Cohen et al. 2009; Brugha et al. 2011; Elsabbagh et al., 2012). Autism has a lifetime impact on health and wellbeing, constituting a high economic cost to UK services (Knapp, Romeo, & Beecham, 2009).

1.2 Autistic-like traits (ATs) and Autistic Social Traits (ASTs)

As outlined in Part 1, a paradigm-shift has progressed our conceptualisation of autism to a fundamentally dimensional condition (Constantino, 2009). Epidemiological, clinical and family studies conclude that autism represents the extreme manifestation of the Broader Autism Phenotype, a heterogenous constellation of subthreshold autistic-like behavioural traits (ATs) reflecting social communication difficulties and other cognitive, emotional, motor and sensory features (Constantino & Todd, 2000; Constantino, Przybeck, Friesen & Todd, 2000; Piven, Palmer, Jacobbi, Childress & Arndt, 1997; Spiker, Lotspeich, Dimicli, Myers & Risch, 2002; Waterhouse et al., 1996). Autistic social traits (ASTs) are subthreshold behavioural traits reflecting only the social communication and interaction diagnostic domain (Mandy et al., 2018).
Population-based studies have established that ATs are quantitatively and continuously distributed in the general population, across genders, and from childhood through adulthood (Baron-Cohen, Wheelwright, Skinner, Martin & Clubley, 2001; Constantino & Todd, 2003; Hoekstra, Bartels, Verweij, & Boomsma, 2007; Robinson et al., 2011a; 2011b; Skuse, Mandy & Scourfield, 2005; Skuse et al., 2009). Further, ASTs represent a genetic liability for autism across trait severity, with heritability of 0.74 on the Social Communication Disorders Checklist (SCDC; Skuse et al, 2005) (Bishop et al. 2004; Bolton et al., 1994; Constantino & Todd, 2003; Lundström et al., 2012; Piven et al., 1997; Robinson et al., 2011b; Robinson et al., 2016).

Identifying candidate markers in infancy that predict variability in later ATs can increase our understanding of the established ‘chronogeneity’ of autism: the heterogeneity and changeability of autistic symptoms over time (Georgiades, Bishop & Frazier, 2017). There is a plethora of evidence that even subthreshold ATs are associated with particular patterns of atypical infant behaviours or ‘red flags’, behaviours indicative of a high risk or presence of autism, from the first year of life (Bolton, Golding, Emond & Steer, 2012). ATs may be an important and understudied area with implications for understanding the relationships between early development and autism manifestation (Constantino, 2009).

1.3 Individual differences in developmental trajectories of AST

Longitudinal studies have identified within-person change in autism and ASTs over the course of child and adolescent development which shows some individuals escalate in ASTs as they enter adolescence. A forthcoming systematic review found Growth Mixture Modelling techniques such as Latent Class Analysis typically identify a four-class model of AST trajectories in clinical and non-clinical samples (Pender, in press); four distinct subgroups of AST trajectories; low stable,
high stable, increasing and decreasing (Fountain, Winter & Bearman, 2012; Gotham et al., 2012; Lord & Luyster, 2006; Lord et al., 2012a, 2012b; Szatmari et al., 2015; Venker et al., 2014).

This class-model also explains variability in AST trajectories in the general population. Robinson et al. (2011b) identified high within-person stability in parent-reported autistic social traits (ASTs) on the SCDC between seven to 13 years. However, a shallow u-shaped trajectory best characterised ASTs, whereby ASTs decreased between seven to 10 years then increased from 10 to 13 years in a large UK longitudinal cohort, the Avon Longitudinal Study of Parents and Children (ALSPAC). A more nuanced six-latent class model of the same sample was revealed between seven and 16 years, suggesting a high degree of variability in intra-individual differences in the development of ASTs from mid-childhood to mid-adolescence (Pender, 2017).

1.4 Gender difference in AST trajectories

Intriguingly, females in particular are reported by parents to show an escalation in ATs on entry to adolescence. In Pender’s (2017) work, the ‘Late Increase’ class (from 10 to 16 years) had a female majority (58%), suggesting gender-specific latent classes of AT trajectories. The findings of Pender (2017) and Robinson et al. (2011b) are corroborated by Mandy et al. (2018) who identified non-linear and gender-specific AST latent trajectories measured by parent-reported SCDC scores from seven to 16 years in the ALSPAC sample using Latent Growth Curve Analysis. Across genders, there was an initial decline in ATs during childhood (seven to 10 years), followed by an increase in adolescence (10 to 16 years) (see Figure 1). Females were more likely than males to show an escalation in parent-reported ATs during adolescence (10 to 16 years) as shown by their deeper U-shaped quadratic trajectory. This escalation accounted for the disappearance of the gender difference in ATs at 16 years (see Figure 2).
Figure 1. Social Communication Disorders Checklist (SCDC) trajectories for boys (denoted by upper curve/white circles) and girls (denoted by lower circles/black circles). Reproduced with permission from Mandy et al. (2018).

Figure 2. Developmental gender difference in SCDC across childhood and adolescence. Reproduced with permission from Mandy et al. (2018).

This adolescent gender difference is consistent with the observed lower male-to-female ratio in adolescent or adult autism clinics versus child clinics and may be a contributing factor to later than average diagnosis for autistic females.
compared to males. There is an inequitable diagnostic gender bias in autism; males are diagnosed at a rate of 3:1 in part due to the dominant male-centric definition of autism (Giarelli et al., 2010; Mandy & Lai, 2017; Rutherford et al., 2016), with females often waiting until adolescence or adulthood (Bargiela, Steward & Mandy, 2016; Loomes, Hull & Mandy, 2017).

One can speculate about what this female AST uptick represents. It may represent a true later onset of new social communication problems for females. However, the finding is consistent with the adolescent emergence hypothesis that female ASTs manifest later than males (Asperger, 1944), and has been observed in other externalising disorders such as ADHD in women during adolescence (Zahn-Waxler, Crick, Shirtcliff & Woods, 2006). Some females with subtle social difficulties in childhood may find these become more problematic and therefore observable to parents in adolescence when social demands increase, consistent with the retrospective reports of late-diagnosed females (Bargiela et al., 2016). Perhaps the uptick represents epiphenomena given the association identified between ATs and social anxiety and other internalising symptoms identified in Part 1.

1.5 Early candidate markers of AST trajectories

The development of heterogenous, gender-specific AST trajectories now needs to be understood (Mandy & Lai, 2017). Identifying early developmental signs of ATs for each gender holds potential to explain how developmental domains interact to manifest in different AST trajectories (Johnson, Gliga, Jones & Charman, 2015; Jones, Gliga, Bedford, Charman & Johnson, 2014). To date, there is no single behavioural marker (nor at any descriptive level) with a robust positive predictive value for autism (Jones et al., 2014). Despite this, diagnostic tests are based on identification of ‘red flags’ from a male-centric definition of the autism phenotype (Mandy & Lai, 2017), for example, the gold-standard Autism Diagnostic Observation
Schedule (now ADOS-2; Lord et al., 2012) and Autism Diagnostic Interview (ADI-R; LeCouteur, Lord & Rutter, 2003). Identification of early markers of atypical development related to ATs could inform earlier and more reliable gender-specific diagnosis and access to intervention, with profound implications for females.

Developmental theories propose that early, subtle differences across social and non-social domains of development interact with the infant’s environment to have downstream effects placing them on different developmental pathways to autism (Happé, 2015). For example, the social orienting and motivation hypotheses predict atypical social communicative behaviours could be early candidates (Chevallier et al., 2012; Dawson et al., 1998; Mundy & Neal, 2000). These accounts hypothesise either a deficit in attention (social orienting) or reward value of social stimuli (social motivation) meaning children cannot and/or do not engage with social content. They therefore fail to develop the opportunity to understand and learn normative social communication and language skills early on, reinforcing attentional and/or reward-based deficits. Such hypotheses are not without criticism (Johnson, 2014), however empirical reviews report some consistent relationships between early atypicalities in social communication and language and autism.

1.5.1 Social communication and language

Prospective longitudinal studies of infant siblings of children with autism followed up from birth to 36 months when diagnosis is more reliable and stable, the high-risk (HR) design (‘Baby Sibling’ studies), provide substantial evidence of associations between atypical social communication and language and autism whereby HR siblings are compared to low risk (LR) and typically developing (TD) (Woolfenden, Sarkozy, Ridley & Williams, 2012).

Two comprehensive reviews and a high-level synthesis conclude delays in receptive and expressive language, nonverbal communication (e.g., symbolic
gestures), and atypicalities in eye contact, social smiling and socially-directed vocalisations by 12 months predict HR infants diagnosed with autism (Jones et al., 2014; Johnson et al., 2015; Szatmari et al., 2016). Some studies find that fewer nonverbal communicative gestures and understanding fewer verbal phrases on the parent-report MacArthur-Bates Communicative Development Inventory (MB-CDI; Fenson et al., 1993) at 12-18 months characterises HR siblings who receive a diagnosis (Mitchell et al., 2006; Zwaigenbaum et al., 2005). Other measures similarly find a significantly reduced variety of gestures and reduced variability in types of consonants uttered at 12-14 months in HR siblings later diagnosed (e.g., Landa, Holman & Garrett-Mayer, 2007; Talbott, Nelson & Tager-Flusberg, 2015).

Reviews typically conclude that social communication and language delays are not pronounced until at least 12 months old (Szatmari et al., 2016). However longitudinal HR sibling studies of developmental trajectories of social communicative behaviour have found nuanced atypicalities before 12 months, suggesting masking of findings by cross-sectional group-level designs. For example, a slower growth in use of gestures to initiate joint attention (shared attention between infant and caregiver), repeatedly measured at 5 timepoints from as early as 8 months to 18 months, distinguished HR siblings with and without an autism diagnosis, despite overall increasing growth which would otherwise indicate neurotypical development (Ibañez, Grantz & Messinger, 2013).

1.5.2 Motor development

Candidates for early markers outside social developmental include atypical or delayed motor development, given that stereotyped and repetitive motor mannerisms characterise some children with autism and are a diagnostic feature in the RRBs domain (Jones et al., 2014). Delays in gross and fine motor skill acquisition may subtly disrupt typical developmental pathways by reducing the
infant’s opportunities for social communicative learning (Johnson et al., 2015), for example, poorer fine motor skills are associated with delayed language development in HR siblings (LeBarton & Iverson, 2013). Gross motor skills are larger movements such as grasping, crawling and running, developed before fine motor skills such as picking things up with fingers or using the tongue to taste and feel objects, with coordination required for independent living (Johnson et al., 2015).

Reviews find transient atypicalities in children as young as six months old who later receive an autism diagnosis, in gross and fine motor coordination, movement patterns during walking and goal-directed movement (Fournier, Hass, Naik, Lodha & Cauraugh, 2010; Johnson et al. 2015). For example, significant delay in head lag (lifting the head up) when pulled to sit up (Flanagan, Landa, Bhat & Bauman, 2012), spending more time in less developmentally advanced postures (e.g. lying down), and shifting posture less frequently (Nickel, Thatcher, Keller, Wozniak & Iverson, 2013). Similar delays on measures of fine and gross motor skills are found in HR siblings who later receive a diagnosis (Landa & Garrett-Mayer, 2006; Ozonoff et al. 2010; LeBarton & Iveson, 2013). Large-scale, longitudinal data needs to replicate these early findings so firmer conclusions can be drawn.

1.5.3 Temperament

A final and relatively understudied candidate is temperament, conceptualised as relatively stable individual differences in behavioural tendencies exerting bidirectional influences upon the social environment, expressed differently depending on changing behavioural repertoire (Thomas & Chess, 1977). Early abnormalities in emotion and behavioural reactivity and regulation could reduce social opportunities and may interact with ATs having a downstream effect on AT manifestation (Del Rosario, Gillespie-Lynch, Johnson, Sigman & Hutman, 2014).
Cross-sectional literature shows that temperament profiles differ between TD children and children with autism from as early six to 24 months old (Del Rosario et al., 2014). Despite variable measurement, emerging trends suggest relationships between atypicalities in approach, adaptability, activity levels, and negative emotionality and autism. Reduced approach behaviour towards social stimuli, reduced adaptability (flexibility), higher activity levels (physical activity during sleep, eating, play etc), higher levels of emotional distress and reduced effortful control (regulation) of negative emotions in the first 24 months predict diagnosis at 24-36 months in HR versus LR or TD peers (Clifford et al., 2013; Garon et al., 2009; Zwaigenbaum et al., 2005). Similarly, older children with autism (three to 12 years) show the same pattern (Adamek et al. 2011; Brock et al. 2012; Hepburn & Stone, 2006; Konstantareas & Stewart, 2006). Some studies also show atypicalities in reactivity to sensory stimuli in the first 24 months are associated with childhood diagnosis (Brock et al., 2012; Clifford et al., 2013; Hepburn & Stone, 2006).

Recent evidence suggests chronogeneity in temperament, and the importance of measurement from 18 months when continuity and stability increase (McDevitt, 1986; Pedlow, Sanson, Prior, Oberklaid, 1993). Del Rosario et al. (2014) found temperament trajectories of HR infants diagnosed at 36 months reflected previous findings: increases over time in activity level and decreasing adaptability and approach behaviours. Unexpectedly and similarly to Zwaigenbaum et al. (2005), initially higher activity levels and faster adaptation at six and 12 months in infants diagnosed was found but this effect reversed by 36 months.

1.6 Early candidate markers of ATs and ASD in the general population

The literature reviewed has investigated correlates of autism, not ATs specifically. One study attempted to identify early candidates predictive of ATs and autism diagnosis at 11 years old in the general population ALSPAC sample (Bolton
et al., 2012). Bolton and colleagues (2012) investigated 241 contemporaneous parent-report questionnaire items capturing social and non-social domains measured from six to 30 months old. Children who developed higher ATs or an autism diagnosis showed developmental differences from six months. The strongest predictors were social communication and vocabulary from six months (on the Denver Developmental Screening Test [DDST]; Frankenburg & Dodds, 1967), 'Vocabulary' and 'Combines Words' scores from the MB-CDI; Fenson et al., 1993), and a repetitive behaviour score. Differences in fine motor skills (measured by the DDST) were predictive from six months. Temperament measured by the CTS (Carey & McDevitt, 1977) was predictive from 24 months, with associations between elevated ATs and less activity and intensity at six months which reversed by 24 months, similar to Del Rosario et al. (2014). This approach, although comprehensive, may have masked effects, and the non-conventional AT measure limits comparability.

1.7 Study aims

In summary, LCGA has identified non-linear, gender-specific developmental latent AST trajectories in a longitudinal general population study (ALSPAC; Mandy et al., 2018). To date, no study has investigated early behavioural indicators of gender-specific AST developmental trajectories in the general population, nor used a theoretical and empirical approach to select candidates. It is striking that discussion of gender differences in reviews of early markers is lacking. Therefore this study uses the ALSPAC sample to extend the work of Mandy and Bolton and colleagues to investigate selected early social and non-social behavioural predictors over the first 24 months of life of Mandy et al.’s (2018) gendered AST trajectory model, and to explore gender differences. This will help to illuminate the developmental origins of AST trajectory chronogeneity between genders.
1.7.1 Research questions

1. What social communication, language, temperament and motor behaviours in the first 24 months of life predict change over time in ASTs from mid-childhood to mid-adolescence for males and females?

2. Are there gender differences in early candidates predictive of change over time in ASTs?

2.0 Method

2.1 Setting

Secondary data analysis was carried out at University College London (UCL), using data from the Avon Longitudinal Study of Parents and Children (ALSPAC). All pregnant women resident in three former urban and rural District Health Authorities (DHAs) of Frenchay, Southmead, and Bristol and Weston in the South West of England with an estimated delivery date between 1st April 1991 and 31st December 1992 inclusive, and any child born of these pregnancies, were eligible (Boyd et al., 2013).

2.2 Participants

2.2.1 Recruitment

Recruitment to ALSPAC was opportunistic across three phases (see Figure 3). In Phase I, all eligible women from the 1990 to 1992 estimated delivery date period were contacted via media, recruitment visits to the community, and study promotion during routine health service appointments. The majority (82.6%) enrolled in Phase 1 (Boyd et al., 2013).
2.2.2 Sample

Of 15,247 enrolled pregnancies (75.3% of eligible), 14,701 live-born children were alive at one year, named the ‘enrolled sample’ (see Figure 4). The majority entered the study at Phase one of whom 13,988 live-born children were alive at one year. Children were followed-up from birth to 18 years. 1,498 participants were lost to follow-up, resulting in 12,350 remaining aged 18. ASTs were measured at seven, 10, 13 and 16 years, of which 9,744 children had at least one recorded total SCDC score at a minimum of one timepoint with ≥50% of the items completed. Of these children, 7,773 children had data on at least one predictor variable and so were eligible for inclusion in this study.
Figure 3. ALSPAC recruitment phases, reproduced with permission from Boyd et al. (2013).
2.2.3 Demographic characteristics.

The enrolled child sample were 49.7% female, 96.1% White and 6.22% were categorised as having a Low Household Income (joint parental income of ≤£16,000 per annum; Boyd et al., 2013).
Demographic information for participants from Phase one included in this study (n=7773) is in Appendix I. 51.0% were male, 93.3% were White, 60.7% had a mother with O-Levels or A-Levels as their highest education level (15.8% had a university degree), 80.8% had a mother who was a homeowner, and 40.1% had a mother in a professional, managerial or technical occupation. There were no significant gender differences between any of these characteristics (all p’s > .2). One percent reported their child had an ASD diagnosed before age 16.

The odds of being excluded due to missing all SCDC data were significantly higher for children with mothers with less than A-Level education (Odds Ratio [OR] = 2.36, 95% CI [2.15, 2.58]), was not a homeowner (OR = 2.93, 95% CI [2.71, 3.18]), and for BME participants (OR = 2.22, 95% CI [1.78, 2.77]). Sample composition differed significantly from national demographics (The National Pupil Database KS4 dataset; Boyd et al., 2013); participants were more likely to be White (OR 3.85, CI [3.50-4.24]) and less likely to be of lower socioeconomic status (OR .46, CI [.43-.50]).

2.3 Power Calculation

In LGCM power must be evaluated in relation to the particulars of the data and model (Muthén & Muthén, 2002). A power calculation can detect that a parameter is different from zero. A mean and covariance matrix were determined using residuals, growth factor means, and variances from a growth curve model of the whole study population. This matrix was analysed with the mean of the slope fixed to zero (a misspecification) with 7773 observations to determine a chi-square as an approximate, non-centrality parameter. This was used to calculate the power to detect this chi-square value using α = .05 using SAS University Edition Software for Mac OS X (SAS, 2017). This revealed that the sample size had 90% power to detect the slope value was different from zero.
2.4 Measures

Measures are reproduced in Appendix II, excluding the well-known and copyright protected Wechsler Intelligence Scale for Children.

2.4.1 Autistic social traits (ASTs): Social and Communication Disorders Checklist (SCDC; Skuse et al., 2005)

The SCDC is a 12-item parent-report measure which primarily assesses social and communication skills (verbal and nonverbal). Items are rated according to whether the behaviour has been observed during the past six months, on a 3-point Likert-type scale from ‘0 (not true) to 2 (very or often true)’. Total scores range from zero to 24, where 24 represented the highest level of ASTs and a cut-off of nine or above indicated examination for autism (Skuse et al., 2005). Despite its brevity, the SCDC demonstrates high sensitivity of .90 and specificity of .69 to independently diagnosed cases of autism in the ALSPAC sample (Skuse et al., 2005). It has excellent internal consistency (α=.93), strong discriminant validity, modest correlations between SCDC total score and ADI algorithm scores, good test-retest reliability (ICC=.81) and genetic variability related to autism is associated with SCDC variability (Skuse et al., 2005).

2.4.2 Early behavioural candidates.

Hypothesised predictors were selected from mother-report measures administered in the first 24 months of life. Variable selection was informed by theoretical and empirical rationale and availability. Measures were the same as those used in Bolton et al. (2012). Table 1 summarises the predictors which are subscale scores representing four developmental domains: social communication, language, temperament, and motor skills.
Table 1.
Predictor variables, measure and timepoint of measurement by developmental domain

<table>
<thead>
<tr>
<th>Developmental domain</th>
<th>Predictor</th>
<th>Measure</th>
<th>Timepoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>Language subscale score</td>
<td>DDST</td>
<td>6 months</td>
</tr>
<tr>
<td></td>
<td>Vocabulary score</td>
<td>MB-CDIs</td>
<td>24 months</td>
</tr>
<tr>
<td>Motor</td>
<td>Fine Motor-Adaptive subscale score</td>
<td>DDST</td>
<td>6 months</td>
</tr>
<tr>
<td></td>
<td>Gross Motor subscale score</td>
<td>DDST</td>
<td>6 months</td>
</tr>
<tr>
<td>Social communication</td>
<td>Personal-social subscale score</td>
<td>DDST</td>
<td>6 months</td>
</tr>
<tr>
<td></td>
<td>Non-verbal communication score</td>
<td>MB-CDIs</td>
<td>15 months</td>
</tr>
<tr>
<td>Temperament</td>
<td>Activity subscale score</td>
<td>CTS</td>
<td>24 months</td>
</tr>
<tr>
<td></td>
<td>Adaptability subscale score</td>
<td>CTS</td>
<td>24 months</td>
</tr>
<tr>
<td></td>
<td>Approach subscale score</td>
<td>CTS</td>
<td>24 months</td>
</tr>
<tr>
<td></td>
<td>Mood subscale score</td>
<td>CTS</td>
<td>24 months</td>
</tr>
<tr>
<td></td>
<td>Sensory Threshold subscale score</td>
<td>CTS</td>
<td>24 months</td>
</tr>
</tbody>
</table>

*Note.* DDST = Denver Developmental Screening Test, CTS = Carey Temperament Scales, MB-CDIs = MacArthur Bates Communicative Development Inventories.
2.4.2.1 Denver Developmental Screening Test (DDST; Frankenburg & Dodds, 1967).

The DDST is a 105-item validated, standardised screening test for preschool children to identify risk of developmental delay and disability. Forty-two selected items from the original DDST were administered, with remaining items omitted in piloting of the ALSPAC batteries (Golding et al., 2007). It comprises four subscales which assess four areas of development via researcher/parent observation during activities of daily living: personal-social (getting along with others and caring for personal needs); fine motor-adaptive (coordination, object manipulation, problem solving); language (hearing, understanding, using language); gross motor (sitting, walking etc.). Items are scored on a 3-point Likert-type scale from ‘1 (does this often) to 3 (has not started doing this)’. Higher subscale scores indicate typical versus delayed development. The DDST was standardised on two representative samples of American children, with good concurrent validity with gold standard measures of development (.86 to .97; Frankenburg, Camp & Van Natta, 1971), and high test-retest stability (.66 to .93; Frankenberg et al., 1971).

2.4.2.2 MacArthur Communicative Development Inventories (MB-CDIs; Fenson et al., 1996).

The MB-CDIs are validated parent-report measures of receptive and expressive language and communicative skill development. Parents indicate which of a prespecified list of communicative gestures, words and/or sentences their child understands and/or produces. There are two separate forms for this age range: ‘CDI: Words and Gestures for 8-18 months’, and ‘CDI: Words and Sentences for 16-30 months’. The MB-CDIs have been extensively validated and normed worldwide, with excellent reliability (.97 to .99), construct validity with classification accuracy of
up to 92% in predicting language status, and concurrent validity with established measures of language development across ages and impairment levels (Fenson et al., 1993). The items are similar to behavioural items in the ADOS-2 (Lord et al., 2012) and ADI-R (Le Couteur et al., 2003).

Adapted versions were developed by ALSPAC (Golding et al., 2007). The non-verbal communication score from the ‘CDI: Words and Gestures’ form and vocabulary score from the ‘CDI: Words and Sentences’ form were used in this study. The non-verbal communication score is the sum of 10 items rated on a 3-point Likert-type scale from ‘1 (not yet), 2 (sometimes) to 3 (often)’ in answer to ‘When children are first learning to communicate, they often use gestures to make their wishes known. Which does your infant do?’ A higher non-verbal communication subscale score indicates a larger range of communicative and symbolic gestures. The vocabulary score is checklist of words rated on a 3-point Likert-type scale as ‘0 (neither), 1 (understands) to 2 (says)’. Responses are summed to yield a total score from zero to 246, with a higher score indicating larger vocabulary.

2.4.2.3 Carey Infant Temperament Scales (CTS; Carey & McDevitt; 1977).

The CTS is a set of five validated parent-report questionnaires measuring nine categories of temperament across different ages: activity level, regularity, adaptability, initial approach-withdrawal, intensity, mood, persistence, sensory threshold and distractibility. Items were rated on a Likert-type 5-point scale in terms of ‘how much the child’s behaviour has been like that described over the past four to six weeks, from 1 (almost never) to 5 or 6 (almost always), with some items reverse-scored. The CTS was standardized on four to eight month old infants (n=208) and showed good test-retest reliability and internal consistency (Carey & Devitt, 1978).
The Toddler Temperament Scale version for one and two year olds was used, with items omitted following non-response during piloting by ALSPAC (Golding et al., 2007). The activity, adaptability, approach, mood and sensory threshold subscale scores measured at 24 months were selected and interpreted as follows: a higher activity score indicates a more active versus inactive child during daily activities; a higher adaptability score indicates the child is more flexible in modifying behavioural reactions to social and non-social stimuli; a higher approach score indicates higher levels of initial cautiousness versus confidence in responding to new social and non-social stimuli; a higher mood score indicates higher level of negative versus positive overall affect represented by unpleasant/unfriendly behaviour; a higher sensory threshold score indicates the child needs less sensory stimulation from its environment to respond (i.e. is more sensitive).

2.5 Ethics

Ethical oversight of ALSPAC data collection was provided by the ALSPAC Ethics and Law Committee. Approval was granted in 1989 and 1990 by the three DHAs. Permission to access the ALSPAC data for this study was approved by the ALSPAC Executive Committee in 2018. The researcher agreed to abide by ALSPAC data sharing and security policies. Data from the ALSPAC cohort is anonymised, containing no identifying information. Participants have the right to withdraw their consent for elements of the study or the study entirely at any time as outlined in the ALSPAC Withdrawal of Consent Policy. Polices and supporting documentation are available on the ALSPAC website.
2.6 Data analysis

Data was analysed using Latent Growth Curve Modelling (LGCM) in MPlus 8.0 Plus Mixture Add-On for Mac OS X (Muthén & Muthén, 2017). LGCM is a special form of structural equation modelling that evaluates continuous rate of linear or non-linear change over time (McArdle & Epstein, 1987). It uses a multilevel, random-effects model assuming a latent intercept (starting point) and latent slope (rate of change or growth over time) that varies across and within individuals. This heterogeneity is captured by random effects, continuous latent variables or ‘growth factors’, at baseline and longitudinally (Byrne, 2012). LGCM can include a quadratic growth factor or other type of relationship (e.g., cubic) to capture nonlinearity and can be extended to a multi-group LGCM to examine growth rate differences across observed groups.

2.6.1 Reproduction of Mandy et al. (2018) multi-group LGCM

This study replicated the multi-group quadratic LGCM identified by Mandy et al., (2018) with a slightly reduced sample size (n=7773) and added 11 hypothesised predictors of model intercept, slope and quadratic to explore interindividual differences in AST growth trajectories. Mandy et al. (2018) selected a multi-group LGCM that permitted the mean SCDC score to change as a quadratic function of age, and captured gender differences in patterns of change over time as the best fitting representation of average change over seven, 10, 13 and 16 years. As per the default in multilevel modelling, residual variances were constrained to be longitudinally invariant within each group. The growth factor variance and covariance terms were allowed to be gender variant. Time was specified as the number of years since age seven (with loadings of 0.157, 3.221, 6.406, 9.330).
The authors reran their analyses for complete SCDC data to explore for effects of missingness at all four timepoints \((n=4,380)\) and those with one \((n=6385)\), or two \((n=8024)\) of the four timepoints and found it remained robust (Mandy et al., 2018). IQ did not moderate the effect, measured by the Wechsler Intelligence Scale for Children-III at eight and a half years (WISC-III; Wechsler, Golombok, & Rust, 1992), therefore IQ was not a covariate in this study. The model was also robust for high AT scorers (Mandy et al., 2018).

Goodness-of-model-fit in this analytic sample was assessed using Structural Equation Modelling fit indices, the chi-square test of model fit (likelihood ratio statistic; Bollen, 1989) and alternative fit statistics (Fan & Sivo, 2005; MacCallum & Austin, 2000): the Comparative Fit Index (CFI; Bentler, 1990), Root Mean Square Error of Approximation (RMSEA; Steiger, 1990) and Standardised Root Square Mean Residual (SRMR; Byrne, 1998) (Hu & Bentler, 1995; MacCallum & Austin, 2000). For CFI, a cut-off of .95 and above indicates a well-fitting model (Hu & Bentler, 1999). For RMSEA and SRMR, cut-offs below .06 for RMSEA and .8 for SRMR indicate good fit (Hu & Bentler, 1999).

2.6.2 Hypothesised covariate multi-group LGCM

Eleven predictors of intercept and slope were added into a covariate multi-group LGCM in the final model which did not include predictors of the quadratic term. All predictors were continuous, time-invariant and centred to a mean of zero. Example syntax is in Appendix V. Given the number of comparisons, a Bonferroni correction was applied to reduce the possibility of type one error. A total of 44 comparisons were calculated using the following formula:

\[
11 \text{ predictors} \times 2 \text{ growth factors (intercept and slope)} \times 2 \text{ groups (male and female gender)} = 44
\]
The correction was then calculated using the formula from Bland and Altman (1995) to result in a more stringent $p$ value of $p<.001$:

$$\frac{\alpha}{\text{number of comparisons}} = \text{adjusted } \alpha \text{ value}$$

$$0.05/44 = 0.001$$

*Figure 5.* Path diagram of quadratic LGCM showing one group as an example. *Note.* ‘Scdc1, 2, 3...’ = AST score on SCDC at the four timepoints; $i$=intercept, $s$=slope, $q$=quadratic.
Figure 6. Path diagram of quadratic LGCM with growth factor predictors for one group. Note. ‘Scdc1, 2, 3...’ = AST score on SCDC at the four time points; i=intercept, s=slope, q=quadratic; ke800b-kc968=predictors.

2.6.3 Gender differences.

Predictors of growth factor terms were then explored to see whether associations differed by gender. The Satorra-Bentler scaled (TRd) chi-square test of difference used a mean adjusted (scaled) chi-square test of difference for continuous non-normal outcomes under Maximum Likelihood-based estimation (Satorra & Bentler, 2010).

2.7 Missing data handling

Maximum Likelihood-based estimation with robust standard errors was used in MPlus to handle missing SCDC data to replicate Mandy et al. (2018) (MLR;
Muthén & Muthén, 2017). MLR permits inclusion of partial responders to maximise the amount of information retrieved from incomplete data and computes robust versions of model fit indices whose values are valid despite data non-normality, as in Mandy et al. (2018) and this study. MLR makes the Missing At Random (MAR) assumption of no systematic differences between observed and missing values for dependent variables when estimating the model conditional on remaining covariates (MAR; Little & Rubin, 2014).

3.0 Results

3.1 Descriptive statistics

3.1.1 Missing data

Table 2 shows the proportion of participants with missing data on the SCDC from the analytic sample (n=7773). Over half (52.4%; n=4701) had complete data at all four time points. Proportions of missing data increased across each time point as shown in Table 3, from 10.0% at age 7.5 years to 37.3% by age 16. Of the original sample in Mandy et al. (2018), 1,973 participants were missing data on all predictors and were therefore excluded, resulting in this analytic sample of 7,773 participants (males = n=3963; females = n=3810). Table 4 shows the proportion of participants with missing data for each predictor. Missing data was consistent, except for the MB-CDIs which had a higher percentage of missing data at 15 months for the Non-verbal communication score (28.8%).
Table 2.

*Number and percent of missing SCDC data points (n=7333)*

<table>
<thead>
<tr>
<th>Data points</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4071</td>
<td>52.4</td>
</tr>
<tr>
<td>1</td>
<td>1506</td>
<td>19.4</td>
</tr>
<tr>
<td>2</td>
<td>1143</td>
<td>14.7</td>
</tr>
<tr>
<td>3</td>
<td>1053</td>
<td>13.5</td>
</tr>
</tbody>
</table>

Table 3.

*Missing SCDC data at each time point (n=7333)*

<table>
<thead>
<tr>
<th>Time point</th>
<th>Age</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>774</td>
<td>10.0</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>1371</td>
<td>17.6</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>1904</td>
<td>24.5</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>2902</td>
<td>37.3</td>
</tr>
</tbody>
</table>

3.1.2 Correlations between predictors.

Table 5 summarises correlation coefficients and means and standard deviations for the predictors examined using SPSS Version 25 for Mac. A number of predictors very weakly or weakly correlated, with a few moderately correlated. Due to low risk of multicollinearity, all predictors were entered into the model. The MB-CDI Vocabulary score was the only variable to correlate with every variable except the Non-verbal communication score. The DDST subscale scores showed a weak to moderate positive correlation with each other, with strongest correlations for the Gross Motor and Fine Motor subscales, $r = .50$, $p < .001$, CI=.48-.51, and Gross
Motor and Personal-Social subscales, $r = .50$, $p < .001$, CI=.39-.42. These subscales were not collapsed because this study was interested in partialling out the effects of developmental domains hypothesised to be predictive of ASTs. There were moderate positive correlations for the CTS subscales: Adaptability and Mood subscales, $r = .59$, $p < .001$, CI=.57-.60, Adaptability and Activity, $r = .47$, $p < .001$, CI=.46-.49, and Approach and Mood, $r = .37$, $p < .001$, CI=.35-.39.
### Table 4.

**Number and percent with missing data for predictors from Mandy et al. (2018) sample (n=9744)**

<table>
<thead>
<tr>
<th></th>
<th>Language</th>
<th>Personal-social</th>
<th>Fine Motor</th>
<th>Gross Motor</th>
<th>Non-verbal Communication</th>
<th>Vocabulary</th>
<th>Activity</th>
<th>Approach</th>
<th>Adaptability</th>
<th>Sensory Threshold</th>
<th>Mood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>6 months</td>
<td>1013(10.40)</td>
<td>1014(10.41)</td>
<td>1012(10.39)</td>
<td>998(10.24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2810(28.84)</td>
<td></td>
</tr>
<tr>
<td>24 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>333(13.69)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1289(13.23)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1293(13.27)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1326(13.61)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1288(13.22)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1290(13.23)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5

Correlations among predictor variables, sample mean and standard deviation (n=7773)

<table>
<thead>
<tr>
<th>Predictor (Measure)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Personal-Social (DDST)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Fine Motor (DDST)</td>
<td></td>
<td>.50**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Language (DDST)</td>
<td></td>
<td></td>
<td>.36**</td>
<td>.37**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Gross Motor (DDST)</td>
<td></td>
<td></td>
<td></td>
<td>.41**</td>
<td>.50**</td>
<td>.32**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Non-verbal communication (MB-CDIs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
<td>-.02</td>
<td>-.01</td>
<td>-.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Vocabulary (MB-CDIs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.24**</td>
<td>.27**</td>
<td>.19**</td>
<td>.20**</td>
<td>-.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Activity (CTS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.03*</td>
<td>.04**</td>
<td>-.01</td>
<td>-.05**</td>
</tr>
<tr>
<td>8. Adaptability (CTS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.07**</td>
<td>-.10**</td>
<td>-.05**</td>
<td>-.06**</td>
<td>.00</td>
</tr>
<tr>
<td>9. Approach (CTS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
<td>-.06**</td>
<td>-.03**</td>
<td>-.05**</td>
</tr>
<tr>
<td>10. Sensory Threshold (CTS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.13**</td>
<td>.11**</td>
<td>.11**</td>
</tr>
<tr>
<td>11. Mood (CTS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.04**</td>
<td>-.08**</td>
</tr>
</tbody>
</table>

Note. DDST = Denver Developmental Screening Test, CTS = Carey Temperament Scales, MB-CDIs = MacArthur Bates Communicative Development Inventories. * p<.05, ** p<.01. Significant correlations are in bold.
3.2 Multigroup LGCM

3.2.1 Model identification

The multi-group quadratic LGCM identified by Mandy et al. (2018) remained robust in this sample with excellent fit to the data, CFI=.995, RSMEA=.03 (95% CI=.01-.04), and SRMR=.02. Estimated growth factor means for this model are shown in Table 6.

Table 6

<table>
<thead>
<tr>
<th>Growth Factor</th>
<th>Mean</th>
<th>S.E</th>
<th>Mean/S.E</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>3.31</td>
<td>.06</td>
<td>51.84</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Slope</td>
<td>-0.18</td>
<td>.02</td>
<td>-8.05</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Quadratic</td>
<td>0.02</td>
<td>.00</td>
<td>6.94</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.44</td>
<td>.05</td>
<td>48.05</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Slope</td>
<td>-0.19</td>
<td>.02</td>
<td>-9.42</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Quadratic</td>
<td>0.03</td>
<td>.00</td>
<td>12.64</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Mean SCDC scores at the four time points estimated by this model are shown in Table 7. These values are almost identical to that estimated by Mandy et
al. (2018). As per the pattern of AST trajectories in Mandy et al. (2018), males had a higher mean estimated SCDC score than females from seven to 13 years, however by age 16 this gender difference became non-significant ($p=1.00$). The gender gap narrowed from seven years and this reduction escalated as the sample entered adolescence at 13 years. As per Mandy et al. (2018), the absence of a gender difference at 16 years is explained by a sharper increase in estimated mean SCDC score in females compared to males (Figures 7 and 8).

*Figure 7.* AST trajectories for males and females, with circles denoting estimated sample Social Communication Disorder Checklist (SCDC) means.

*Figure 8.* Gender difference in AST (estimated mean SCDC score) across childhood and adolescence.
Table 7.

*Estimated mean SCDC scores for males and females and mean differences in gender across childhood and adolescence (n=7773)*

<table>
<thead>
<tr>
<th></th>
<th>7 years</th>
<th>10 years</th>
<th>13 years</th>
<th>16 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male mean SCDC score (SE)</td>
<td>3.22 (.07)</td>
<td>2.85 (.07)</td>
<td>2.81 (.07)</td>
<td>3.07 (.08)</td>
</tr>
<tr>
<td>Female mean SCDC score (SE)</td>
<td>2.40 (.05)</td>
<td>2.06 (.05)</td>
<td>2.29 (.05)</td>
<td>2.99 (.07)</td>
</tr>
<tr>
<td>Gender difference in means (95% CI)</td>
<td>0.86 (.70-1.03)**</td>
<td>0.80 (.64-.96)**</td>
<td>0.52 (.36-.69)**</td>
<td>0.08 (.12-.29)</td>
</tr>
</tbody>
</table>

*Note. SCDC = Social Communication Disorders Checklist; SE = Standard Error; 95 % CI = 95% Confidence Interval. *p<.05, **p<.01.
3.3 Hypothesised predictors of multi-group LGCM

A summary of predictors of initial SCDC status at seven years (intercept) and rate of change in SCDC from seven to 16 years (slope) is shown in Table 8. Initially, all growth factors were entered. However, only one variable was predictive of the quadratic term and no longer remained significant after correction. Therefore, predictors of the quadratic were removed (males; DDST language: \( p = .01 \); females; CTS Approach \( p = .04 \)). Two variables subsequently became predictive of slope due to correlations between predictors and slope and quadratic knocking out some associations in the first analysis.

All five temperament variables at 24 months and the two social communication and language variables of non-verbal communication and vocabulary scores at 15 and 24 months predicted initial SCDC score for males and females. For males, these remained significant post-correction. For females, only the temperament scores remained significant. Males and females with higher CTS adaptability and mood scores at 24 months started with a relatively higher SCDC score at seven years. Males and females with lower CTS approach and sensory threshold scores, and vocabulary scores at 24 months and non-verbal communication scores at 15 months started with a relatively lower initial SCDC score. Females with a lower score on fine motor skills at six months started with a lower initial AST score relative to peers with a higher fine motor skill score.

For males, there was a trend towards the CTS Activity subscale score at 24 month predicting rate of SCDC change from seven to 16 years; -0.01, estimated standard error (SE)= -2.20, \( p = .03 \), however, this became non-significant post-correction (see Figure 9). This suggested that males with a higher activity score at 24 months old started with a relatively higher SCDC score and rate of change decreased relatively more over time, representing a negative growth rate. For females, there was a trend towards the DDST Language subscale score at six months predicting rate of SCDC change; .01, estimated SE=2.26, \( p = .03 \), which
became non-significant post-correction (see Figure 10). This suggested that rate of SCDC change increased relatively more over time for females with a higher language score at six months, representing a positive growth rate.
Table 8

*Summary of the 11 predictors of model intercept and slope (AST score) for males and females (n=7773)*

<table>
<thead>
<tr>
<th>Age</th>
<th>Developmental Domain</th>
<th>Standardised Measure</th>
<th>Intercept</th>
<th></th>
<th>Slope</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Estimate</td>
<td>Estimated SE</td>
<td>p</td>
<td>Estimate</td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>Language</td>
<td>DDST – Language</td>
<td>0.01</td>
<td>0.44</td>
<td>.66</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Motor</td>
<td>DDST – Fine Motor</td>
<td>-0.00</td>
<td>-0.13</td>
<td>.90</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DDST – Gross Motor</td>
<td>0.10</td>
<td>0.72</td>
<td>.47</td>
<td>0.00</td>
</tr>
<tr>
<td>6 months</td>
<td>Social Communication</td>
<td>DDST – Personal Social</td>
<td>0.00</td>
<td>0.02</td>
<td>.82</td>
<td>-0.00</td>
</tr>
<tr>
<td>15 months</td>
<td>Social Communication</td>
<td>MB-CDIs Non-Verbal</td>
<td>-0.08</td>
<td>-3.37</td>
<td>.00***</td>
<td>0.01</td>
</tr>
<tr>
<td>24 months</td>
<td>Temperament</td>
<td>CTS – Activity</td>
<td>0.08</td>
<td>4.65</td>
<td>.00***</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CTS – Adaptability</td>
<td>0.10</td>
<td>4.34</td>
<td>.00***</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CTS – Approach</td>
<td>-0.03</td>
<td>-3.52</td>
<td>.00***</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CTS – Mood</td>
<td>0.08</td>
<td>-0.06</td>
<td>.00***</td>
<td>-0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CTS – Sensory Threshold</td>
<td>-0.06</td>
<td>-0.02</td>
<td>.00***</td>
<td>0.00</td>
</tr>
<tr>
<td>24 months</td>
<td>Language</td>
<td>MB-CDIs Vocabulary</td>
<td>-0.01</td>
<td>-3.60</td>
<td>.00***</td>
<td>0.00</td>
</tr>
<tr>
<td>Age</td>
<td>Developmental Domain</td>
<td>Standardised Measure</td>
<td>Intercept</td>
<td>Slope</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------</td>
<td>------------------------</td>
<td>--------------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Co-efficient</td>
<td>Estimated SE</td>
<td>p</td>
<td>Co-efficient</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>Language</td>
<td>DDST – Language</td>
<td>0.01</td>
<td>0.44</td>
<td>.66</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DDST – Fine Motor</td>
<td>-0.02</td>
<td>-2.22</td>
<td>.03*</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DDST – Gross Motor</td>
<td>0.02</td>
<td>0.20</td>
<td>.00</td>
<td>0.00</td>
</tr>
<tr>
<td>6 months</td>
<td>Social Communication</td>
<td>DDST – Personal Social</td>
<td>0.00</td>
<td>0.13</td>
<td>.90</td>
<td>-0.00</td>
</tr>
<tr>
<td>15 months</td>
<td>Social Communication</td>
<td>MB-CDIs Non-Verbal</td>
<td>-0.04</td>
<td>-2.06</td>
<td>.04*</td>
<td>0.00</td>
</tr>
<tr>
<td>24 months</td>
<td>Temperament</td>
<td>CTS – Activity</td>
<td>0.05</td>
<td>3.84</td>
<td>.00***</td>
<td>-0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CTS – Adaptability</td>
<td>0.05</td>
<td>3.31</td>
<td>.00***</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CTS – Approach</td>
<td>-0.03</td>
<td>-4.00</td>
<td>.00***</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CTS – Mood</td>
<td>0.07</td>
<td>5.50</td>
<td>.00***</td>
<td>-0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CTS – Sensory Threshold</td>
<td>-0.04</td>
<td>-3.00</td>
<td>.00***</td>
<td>0.00</td>
</tr>
<tr>
<td>24 months</td>
<td>Language</td>
<td>MB-CDIs Vocabulary</td>
<td>-0.04</td>
<td>-2.10</td>
<td>.04*</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Note. DDST = Denver Developmental Screening Test, CTS = Carey Temperament Scales, MB-CDIs = MacArthur Bates Communicative Development Inventories, SE = Standard Error.
*p<.05, **p<.01, ***p<.001. Predictors remaining significant at adjusted p<.001 are in bold.*
Figure 9. AST trajectories (estimated SCDC mean scores) for males and females scoring high and low on CTS Activity at 24 months. Note. ‘higher’ = one Standard Deviation (SD) above mean Activity score, ‘lower’ = one SD below mean Activity score.

Figure 10. AST trajectories (estimated SCDC mean scores) for males and females scoring high and low on DDST Language at six months. Note. ‘higher’ = one
Standard Deviation (SD) above the mean Language score, ‘lower’ = one SD below the mean Activity score.

3.4 Exploratory analyses.

3.4.1 Gender differences.

Gender differences in predictors of initial SCDC score at seven years and growth rate from seven to 16 years are shown in Table 9. The CTS Activity subscale score at six months was more strongly associated with the rate of change in SCDC for males than females, $\chi^2 = 4.75, p = .03$. The MB-CDIs Non-verbal communication score at 15 months and Vocabulary score at 24 months were more strongly associated with the initial SCDC score at seven years for males; $\chi^2 = 4.75, p = .03$, and $\chi^2 = 4.75, p = .03$ respectively. All temperament subscale scores at 24 months were more strongly associated with the initial SCDC score at seven years for males than females ($p < .001$). As these were exploratory analyses, Bonferroni correction was not applied.
<table>
<thead>
<tr>
<th>Age</th>
<th>Predictor</th>
<th>Intercept</th>
<th></th>
<th></th>
<th>Slope</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\chi^2$ (Trd)</td>
<td>$p$</td>
<td>$\chi^2$ (Trd)</td>
<td>$p$</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>DDST - Language</td>
<td>0.20</td>
<td>.66</td>
<td>0.55</td>
<td>.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DDST - Fine Motor</td>
<td>0.03</td>
<td>.87</td>
<td>0.24</td>
<td>.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DDST - Gross Motor</td>
<td>0.52</td>
<td>.47</td>
<td>0.01</td>
<td>.93</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DDST - Personal Social</td>
<td>0.05</td>
<td>.82</td>
<td>1.97</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>15 months</td>
<td>MB-CDIs - Non-Verbal</td>
<td>11.61</td>
<td><strong>.00</strong></td>
<td>3.38</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>24 months</td>
<td>MB-CDIs - Vocabulary</td>
<td>13.70</td>
<td><strong>.00</strong>*</td>
<td>0.16</td>
<td>.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTS - Activity</td>
<td>21.43</td>
<td><strong>.00</strong>*</td>
<td>4.75</td>
<td>.03*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTS - Adaptability</td>
<td>19.03</td>
<td><strong>.00</strong>*</td>
<td>3.28</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTS - Approach</td>
<td>12.24</td>
<td><strong>.00</strong>*</td>
<td>3.34</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTS - Mood</td>
<td>27.64</td>
<td><strong>.00</strong>*</td>
<td>0.96</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTS - Sensory Threshold</td>
<td>16.14</td>
<td><strong>.00</strong>*</td>
<td>1.10</td>
<td>.30</td>
<td></td>
</tr>
</tbody>
</table>

Note. DDST = Denver Developmental Screening Test, CTS = Carey Temperament Scales, MB-CDIs = MacArthur Bates Communicative Development Inventories, Trd = Satorra-Bentler chi-square. Degrees of freedom = 1 for all tests. * $p<.05$, ** $p<.01$, *** $p<.001$. Significant predictors at $p<.05$ are in **bold**.
4.0 Discussion

4.1 Summary of key findings.

This study was the first to investigate behavioural candidates from the first two years of life predictive of AST developmental trajectories spanning mid-childhood (seven years) through mid-adolescence (16 years) in the general population. It also explored gender differences in early candidates predictive of AST trajectory using multi-group LGCM for the first time.

The main finding was that select behavioural markers in the first two years of life explained some of the variability in where children start in AST severity in childhood but were very limited in explaining the rate of change in ASTs through later childhood and adolescence. As hypothesised, behavioural markers of social communication, language and temperament at 15 and 24 months old predicted initial AST status at seven years for males and females, and largely remained significant after Bonferroni adjustment. Only two behavioural markers predicted growth rate in ASTs from seven to 16 years and were no longer significant after adjustment so should be interpreted conservatively.

This study was the first to identify trends towards gender differences between predictors of AST trajectories, with early behavioural markers more strongly associated with male ASTs than females. Overall, eight gender differences between predictors were identified and primarily related to initial ASTs. Activity level at 24 months was more strongly predictive of the rate of change in ASTs for males, and non-verbal communication and temperament markers at 24 months were more strongly predictive of the rate of change in ASTs for males.

Regarding initial AST status, five temperament characteristics at 24 months and non-verbal communication and vocabulary development at 15 and 24 months were predictive for both genders as hypothesised. Infants who showed higher
negative affect and, surprisingly, infants who were more adaptable (flexible in modifying responses to stimuli) at 24 months started with relatively higher ASTs. Infants who were more approachable (more confident in approaching novel social/non-social stimuli) and who were less reactive to sensory input (required more environmental stimulation) started with relatively lower ASTs. Contrary to previous research (Bolton et al., 2012), males and females who had a smaller range of social communicative gestures at 15 months and smaller vocabulary at 24 months started with relatively lower ASTs. Also unexpected was that only female ASTs were predicted by fine motor skills at six months, such that females with less well developed fine motor skills at six months old started with relatively lower ASTs.

Regarding rate of change in ASTs, males who had a more active temperament (showed more motor movements during daily activities) at 24 months old started with relatively higher ASTs and their ASTs decreased relatively more over time to 16 years than less active male peers. Unexpectedly, females who had more highly developed language skills at six months increased in ASTs over time relatively more than females with less well developed language however this was no longer significant post-correction.

4.2 Findings in the context of previous research

This study extended Mandy et al. (2018) by identifying behavioural predictors of gendered AST trajectories from social and non-social developmental domains over the first 24 months of life in the ALSPAC sample. Bolton and colleagues’ (2012) work was also extended by selective entry of predictors into a developmental AST trajectory model charting AST change beyond 11 years in the same sample. The major strength of this study was the large-scale, prospective longitudinal cohort design that assessed a UK population-based sample over multiple time-points. Findings were therefore less affected by recall and
ascertainment biases and had higher power to detect effects than historically possible.

The primary finding that behavioural markers of social communication, language and temperament predicted initial severity of ASTs in mid-childhood extends Bolton et al.'s (2012) findings before 11 years with a validated AST measure. This is in line with evidence that behavioural differences in social communication, language and temperament development are associated with autism in mid-childhood emerge in the first 24 months of life (Del Rosario et al., 2014; Jones et al., 2014; Johnson et al., 2015; Szatmari et al., 2016), demonstrating early developmental differences are worth studying from 15 months old. It extends the current body of research to demonstrate these are also early markers of subthreshold ASTs. This adds to evidence supporting developmental theories that propose subtle differences in social domains of development have a downstream effect on ASTs (Happé, 2015), although the lack of predictors of rate of change in AST limit further interpretation.

A range of temperament characteristics at 24 months old were predictive of initial AST severity, highlighting more research is needed on this understudied domain. This is particularly important for females as temperament only remained significant post-correction. Temperament explained variance in childhood ASTs earlier than identified by Bolton et al. (2012), perhaps due to fewer predictors which made this study more powerful to detect effects. Higher negative affect, activity levels and (counterintuitively) higher adaptability to respond to social and non-social stimuli at 24 months predicted relatively higher AST severity at seven years consistent with previous research (Adamek et al., 2011; Brock et al., 2012; Clifford et al., 2013; Garon et al., 2009; Konstantareas & Stewart, 2006; Paterson et al., 2019; Zwaigenbaum et al., 2005). Also consistent was that infants more confident in approaching novel stimuli and less reactive to sensory input started with relatively lower ASTs (Brock et al., 2012; Clifford et al., 2013; Del Rosario et al., 2014;
Hepburn & Stone, 2006). It would be interesting to investigate the causal relationship between temperament characteristics and ASTs to unpick downstream effects they may have.

Interestingly, the only predictor of growth rate for males was activity levels. The direction of this finding was initially in line with a previous study which charted temperament trajectories from six to 36 months (Del Rosario et al., 2014). Despite this effect being reversed by 36 months in Del Rosario et al. (2014), in this study higher activity levels at 24 months old predicted a negative AST growth rate. This highlights the chronogeneity of developmental domains, and that subthreshold ASTs may vary in their relationship with atypical early developmental markers compared to clinical samples.

The direction of the relationships between the social communication markers of non-verbal communication and vocabulary development and AST severity was unexpected. Reviews conclusively find delayed verbal and non-verbal communication skills from 12 months are associated with more severe ATs in HR children and with autism diagnosis (Jones et al., 2014; Johnson et al., 2015; Szatmari et al., 2016). It may be these findings add something new; poorer verbal and non-verbal communication skills predict initially lower subthreshold ASTs in mid-childhood in the general population.

The unexpected finding that females with more highly developed early language skills at six months increased in ASTs over time relative to females with poorer language skills can be understood through the phenomenon of ‘camouflaging’ and offers support for this theory at a younger age than previously identified (Hull et al., 2017; Lai et al., 2017). Camouflaging involves behavioural modifications made by autistic individuals to appear neurotypical, for example, copying phrases or suppressing repetitive behaviours. Similarly, ‘social compensation’ theory proposes cognitive atypicalities in autistic individuals are compensated for by improving those atypicalities whilst leaving other traits.
unchanged (Livingston et al., 2019). Both genders use these mechanisms, with research suggestive that females more likely to camouflaging more often (Lai et al., 2017). It could be that the females with higher early language skills in this study are already implicitly learning to mask their ASTs via higher language skills at a young age. This fits with previous research that higher verbal intelligence quotient (IQ) is more directly associated with lower SCDC in females compared to males aged seven (Skuse et al., 2009). It also echoes qualitative reports of late-diagnosed autistic women who report using social mimicry from youth to learn socially acceptable verbal and non-verbal behaviours to mask their autism in order to fit in (Bargiela et al., 2016)

It was surprising that motor development was not particularly predictive of initial AST status nor particularly predictive of growth rate given research suggesting an association between delayed motor development and autism (Fournier et al., 2010; Johnson et al., 2015; Jones et al., 2014). Female AST severity at seven years was predicted by development of fine motor skills at six months in a direction inconsistent to previous research (Bolton et al., 2012; Landa & Garrett-Mayer, 2006; LeBarton & Iveson, 2013; Ozonoff et al. 2010). This was not the case for boys, however, nor for gross motor skills which are theoretically linked to fine motor skills (Johnson et al., 2015). Given the DDST subscales moderately correlated (Frankenburg & Dodds, 1967), combining the motor subscales via factor analysis could be worthwhile to see if more global motor atypicalities better predict ASTs.

The lack of significant predictors of AST growth rate could represent many scenarios. One speculation could be the small variability in the AST growth parameter, leaving limited variance to be explained. In a general population sample with overall low parent-reported ASTs, it may be that ASTS do not relate as strongly to early infant developmental differences compared to higher risk or clinical populations.
Another possible explanation could be the temporal nature of candidate measurement. Behaviours in the first two years may be related to childhood AST severity in mid-childhood but not to more distal timepoints. Later measures may supersede earlier measures when entered into the same model. For example, Bolton and colleagues found similar measures or versions of measures collected at a later time point replaced earlier measures as most predictive over time. The paucity of significant findings for the DDST may be explained in this way, for example the DDST Personal-Social and Language subscales may have been superseded by later non-verbal and vocabulary scores. Future research should measure candidate markers over multiple time-points to address this.

Emerging research is beginning to chart developmental candidate trajectories over time, for example, Paterson and colleagues (2019) charted trajectories of temperament characteristics from six to 24 months in HR and LR infants. Parallel Latent Growth Curve Modelling can answer how trajectories, or two or more parallel processes, are causally related to each other using any variation of LGCM techniques (Cheong, MacKinnon & Khoo, 2003; Felt, Depaoli & Tiemensma, 2017). Child development is dynamic across domains therefore this is arguably a more appropriate way to capture changeable relationships.

The trend towards gender differences in predictors indicates further investigation into early developmental atypicalities related to ATs and autism for males and females is worthwhile. The significance of findings for males compared to females is almost unsurprising given the male-centric definition of autism on which self-report and diagnostic measures are based, and highlights again the bias in autism research perpetuating a gender inequality (Giarelli et al., 2010; Mandy & Lai, 2017; Rutherford et al., 2016). One could also speculate that larger atypical developmental differences are required in early infant behaviours for females to identify a predictive relationship with female ASTs. For example, Robinson et al. (2013) found female sex protects from autistic impairments and that females may
require greater familial etiologic load than males to manifest the phenotype.
Unfortunately, we are not much closer to understanding what the uptick in female
ASTs represents in adolescence from this study.

4.3 Limitations and Research Implications

This study would be strengthened by extension of data collection beyond the
16 year time-point to later adolescence and adulthood. It is important for future
research to examine AST trajectories into adulthood to determine whether observed
trends in each gender are maintained or change over time, and whether the
relationships between early candidates and ASTs are maintained or change.

Part 1 concluded that ATs are modestly correlated with a range of
internalising and externalising traits in the general population. To more
comprehensively explore early candidates, future research should model trajectories
of internalising and externalising symptoms to see if psychopathologic traits predict
AT severity over time. This could begin to answer the questions posed by Part 1
findings; how are ATs and psychopathologic traits developmentally related and are
ATs a risk factor for later emotional difficulties? A recent study found four trajectory
classes of internalising symptoms on the SDQ (Goodman et al., 1995) best
characterised a general population sample of children aged three to nine years;
stable low, stable moderate, stable high, and increasing low to moderate (Klein et
al., 2019). This bears resemblance to the four-latent class model identified in AST
trajectory research of later childhood and adolescence (Pender, in press).

A social developmental construct that was present in Bolton et al. (2012) and
missing from this study was a marker of symbolic play. Derivation of a play score
was explored, however it was not possible to develop a theoretically and statistically
meaningful score from available measures. Symbolic play is a key social
developmental behaviour with considerable evidence for a marked difficulty specific
to autism (Jarrold, Boucher & Smith, 1993; Scambler, Jarrold, 2003; Rogers & Wehner, 2001). It provides an early environment from 12 months in which primitive mind-reading and role-playing multiple perspectives occur, promoting development of communication, problem solving, empathy and emotion regulation (Berk, Mann & Ogan, 2006; Hoffmann & Russ, 2012; Hughes, 1999). Reduced frequency and duration of spontaneous pretend play predicts autism diagnosis in HR versus TD infants (Baron-Cohen, Allen & Gillberg, 1992), and in a large general population sample at 18 months (Baron-Cohen et al., 1996).

A key limitation is that this study is limited to commenting on candidates predictive of ASTs and not the ‘non-social’ aspects of autism, including the RRB subdomain and differences in sensory sensitivity, which is a core definitional domain of ASD (APA, 2013). The SCDC measures social communication difficulties and therefore does not capture RRB, however it is a well-validated measure (Skuse et al., 2005), showing high heritability in adolescence and shared genetic risk with ASD (Robinson et al., 2016; St. Pourcain et al., 2014;). This study could have been strengthened by inclusion of a RRB measure, for example, the Repetitive Behavior Scale-Revised (RBS-R; Lam & Aman, 2007), or a measure capturing both core domains of the phenotype such as the Social Responsiveness Scale which is valid for four to 18 years (SRS; Constantino & Gruber, 2005), or of overall Autism symptoms such as the ADI-R (Le Couteur et al., 2003). This would increase comparability with Bolton et al. (2012) who included an RRB measure.

This validity of this study could have been further strengthened by inclusion of multi-method assessment of ASTs. The SCDC is a parent-report measure reliant on a single source of potentially biased information. Gold standard practice requires a multi-informant, multi-method and multi-context approach (Möricke, Buitelaar, & Rommelse, 2016). Comparison of parent-reported ASTs with a direct observational measure of behaviours such as the ADOS-2 may increase validity (Lord et al., 2012), although this is challenging in a large population study.
Another future avenue could be to have a measure of ASTs/ATs before seven years that is temporally closer to early candidates. For example, from two or three years using an age-appropriate measure such as the Autism Observation Scale for Infants (AOSI; Bryson et al., 2008), the Quantitative Checklist for Autism in Toddlers (Q-CHAT; Allison et al., 2008) or Baby and Infant Screen for Children with aUtism Traits (BISCUIT; Matson et al., 2009), or at least from five years during the transition to school where social and academic challenges increase. Recent population-based research suggests it might be possible to flag and diagnose for autism as young as 14 to 36 months old (Pierce et al., 2019). The first study to explore AT stability from five to eight years in the general population demonstrated ATs can be reliably assessed this young, finding ATs remained primarily stable for both genders (Haraguchi et al., 2019).

Interestingly, Haraguchi et al. (2019) found a gender-specific pattern in children with higher ATs, whereby intra-individual variation remained relatively stable for males, whereas female AT growth was more nuanced. Females remained highly stable in RRBs, but social communication difficulties varied intra-individually despite a stable mean score, and their total SRS score was higher than males at eight years, pointing to possible gender differences before seven years and the need to measure the whole phenotype, in particular RRBs for females.

4.4 Conclusions: Scientific implications

This study found early infant developmental behaviours are predictive of later ASTs from the first two years of life, and that these are likely to differ between genders. The present findings extend research reporting an association between early developmental markers and autism by demonstrating these markers also predict subthreshold ASTs. Several early infant behavioural markers predicted initial AST trajectory in mid-childhood in a large general population cohort, largely in the
expected direction. Social communication, language and temperament characteristics explained some of the variance in initial AST severity from as young as 15 months old. The same markers were however extremely limited in explaining the rate of change in ASTs through later childhood and adolescence. Only two behavioural markers, a temperament characteristic at 24 months for males and early language development at six months for females, predicted growth rate in ASTs from seven to 16 years but no longer remained significant post-correction so should be interpreted with caution. This study was the first to find trends towards gender differences in predictors with early markers significantly more strongly associated with male ASTs, suggesting early development may be differently related to AST development in males and females.

4.5 Conclusions: Clinical implications

This study proffers important implications for clinical practice, indicating screening for signs of higher ASTs from as young as 15 to 24 months, using measures of social communication, language and temperament, could be useful and meaningful. Assessing temperamental characteristics is an overlooked and informative early candidate that could inform more accurate screening. Clinicians and researchers should consider that young females screened or assessed for possible autism may present differently to males in early developmental behaviour. This study demonstrates it is valuable to study subthreshold ATs in large longitudinal population cohorts to identify infant developmental candidates of ATs. Future research should explore gender differences in parallel developmental behaviour and AT trajectories using AT measures capturing both autism symptom subdomain from childhood to adulthood to better understand the chronogeneity of ATs for each gender.
5.0 References


Robinson, E. B., Koenen, K. C., McCormick, M. C., Munir, K., Hallett, V., Happé, F., ... & Ronald, A. (2011b). Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). *Archives of General Psychiatry, 68*(11), 1113-1121.


Part 3: Critical Appraisal
1.0 Dealing with dilemmas and methodological decisions

1.1 Alternative statistical approaches

In approaching Part 2 of this project, I was faced with a choice of two related statistical approaches that had been applied to the Social Communication Disorders Checklist (SCDC) scores in the Avon Longitudinal Study of Parents and Children (ALSPAC) (Skuse et al., 2005); latent class growth modelling, a growth mixture modelling technique (GMM) (LCGM; Pender, 2017), and latent growth curve modelling (LGCM; Mandy et al., 2018). This represented both a methodological and theoretical decision: a) which analytic method best captured chronogeneity in autistic social trait (AST) trajectories, and b) how dimensional is ‘dimensional enough’ when investigating traits? If there had been time available, it would have been interesting to try the LCGM approach used by Pender (2017) to explore what early developmental markers predicted latent class membership, and whether this differed to the findings in this study which opted for the LGCM approach.

GMM approaches have become increasingly popular for capturing chronogeneity over time, both in developmental psychopathology and wider afield. LCGM allows for chronogeneity to be explored by investigating the presence of different latent classes or subgroups of trait trajectories over time, identifying improving and worsening phenotypes (Jung & Wickrama, 2008). This approach has identified novel groups of individuals previously overlooked by studies that applied a single growth trajectory to the ALSPAC data (Robinson et al., 2011), including the finding that a cluster of individuals, particularly females, escalate in ASTs from a low SCDC score at seven years to a more severe score at 16 years (Pender, 2017).

LCGM has been criticised however because it risks over-extraction of spurious latent trajectory classes (Bauer & Curran, 2003). Existence of multiple classes can simply reflect highly skewed data as GMM can locate both ‘real’
subgroups and represent a non-normal distribution as normally-distributed subgroups lacking ecological validity (Bauer & Curran, 2003). As with any form of structural equation modelling, there is a substantial degree of subjective judgement involved in comparing combinations of numerous model fit indices to arrive at model selection (Fan & Sivo, 2005; Hu & Bentler, 1995). Attempting to group trajectories with LCGM can therefore result in sometimes subjective and potentially biased choice of latent classes. However, perhaps we are being overly critical with this latter limitation as this also applies to growth curve analysis, and researcher judgement in balancing statistical precision with interpretability and usability will always be needed, as Sobel and Bohnstedt (1985, p. 158) originally stated,

“Scientific progress could be impeded if fit coefficients (even appropriate ones) are used as the primary criterion for judging the adequacy of a model”.

I discussed this dilemma at the start of the research design process with the authors of previous research and expert statisticians in latent growth modelling. After a lot of deliberating, we opted for the growth curve modelling approach used by Mandy et al. (2018). It was recognised that the final selection of number of latent classes in previous LGCM research with the ALSPAC dataset represented a partially arbitrary decision between a four and six-class model (Pender, 2017). LGCM is one most the most powerful and informative approaches in multiwave longitudinal analysis, enabling modelling of interindividual differences (change across individuals) in intraindividual developmental trajectory change (individual change over time) (Curran & Hussong, 2003; Fan, 2003; Muthén, 2004; Muthén & Curran, 1997; Muthén & Muthén, 2017; Nesselroade, 1991; Tomarken & Waller, 2005). Hypothesised predictors of change can be easily added to explore whether they explain observed heterogeneity in rate of change over time in individual growth trajectories (Muthén & Muthén, 2017; Willett & Sayer, 1994).
On a theoretical level, the latent class growth modelling (LCGM) approach also sat uncomfortably with me. It seemed a little at odds with the highly dimensional conceptualisation of the autism phenotype adopted by the research and autistic communities, where the terminology of ‘spectrum’ has even been superseded by “constellation” (Constantino & Todd, 2003, p.524). Although a step in the right direction, LCGM still takes a somewhat categorical approach in assuming that the best way to capture heterogeneity over time is to divide individuals into groups (classes), albeit more groups than the dichotomous diagnosed versus subthreshold. Arguably, a similar criticism can be levelled at LGCM in its reduction of individuals into a mean latent trajectory. This raises the wider question of what the best way is to approach autistic trait (AT) chronogeneity, and where do we draw the line in acknowledging the incredible neurodiversity in the autism phenotype yet still reduce it enough to enable investigation? Maybe ‘dimensional enough’ is good enough.

1.2 Selection of early candidates of AST chronogeneity

The next decision to grapple with was selection of predictors of the observed chronogeneity in SCDC score (AST trajectory). In anticipation of a criticism of the approach taken in this study, I will elaborate on the reasons for the approach taken and consider alternative methods.

A strength of this study is the selective approach taken to identification of predictors of AST trajectories, representing an extension of previous research which adopted what could be arguably described as ‘the kitchen sink’ approach (Barreto & Howland, 2005), an extensive battery of over 200 predictors entered into a regression model (Bolton et al., 2012). The ‘kitchen sink’ approach helpfully identifies variables warranting further investigation in relation to a construct of interest, however, as Bolton and colleagues acknowledged, entry of a vast number of variables can mask effects and can lead to overfitting and unwanted imprecision.
Significant relationships between variables can be misleadingly suggested because the probability of finding spurious statistically significant effects increases in line with increasing Type 1 error (Barreto & Howland, 2005). To address this, this study selected predictors on the basis of empirical and theoretical justification for inclusion, resulting in a much smaller number of variables and reduced likelihood of overfitting. Similarly to Bolton and colleagues (2012), this study also applied a Bonferroni correction to that end.

Despite the reduction in predictors in this study, there was still a large number due to their hypothesised and/or empirically-supported relationship to later AST development. A number of the predictors also correlated with each other, suggesting a small degree of multicollinearity. This was particularly the case between the Denver Developmental Screening Test subscales (DDST; Frankenburg & Dodds, 1967) and between a number of subscales of the Carey Infant Temperament Scales (CTS; Carey & McDevitt; 1977), resulting in a number of significant moderate correlations for those measures (as would be expected). Given the majority of correlations were very weak to weak, and this study was interested in exploring for the first time what domains of early infant development predicted AST trajectory, I made the decision to include each predictor as planned.

An alternative approach would have been to conduct an exploratory factor analysis of some or all of the predictors prior to model entry to explore what subscale scores best hung together to reveal underlying latent factors causing covariance (Costello & Osborne, 2005; Spearman, 1904; 1927). This method acts as a data reduction technique, reducing the number of individual variables to be entered into a model, which may have resulted in selection of fewer predictors. It may have addressed the issue of moderate correlations between certain subscales and reduced the impact of the small amount of variance in SCDC scores available to be potentially explained by predictors. More meaningful overarching developmental behavioural factors that better and/or differentially accounted for the
variation in AST trajectories may have been identified. On the other hand, collapsing variables may have masked specific effects of particular behaviours predictive of AST chronogeneity.

2.0 Measurement of ATs in population-based research

Part 1 of this project, the meta-analytic review, highlighted the challenges faced when measuring subthreshold ATs in large general population cohorts. Myself and my colleague who rated risk of bias for included studies using the Newcastle Ottawa Scale for cohort studies (NOS; Wells et al., 2004) reflected upon the limited applicability and appropriateness of this tool as introduced in Part 1, leading to development of a revised version. Other alternatives to overcome these limitations could have been to not use a risk of bias tool, as arguably such tools should only be used as and when appropriate, or to select a different tool with greater flexibility such as the Critical Appraisal Skills Checklist for cohort studies (CASP; 2018). Although this measure also takes an implicitly categorical approach, the 12 items are informed by broader questions with less direct reference to exposed versus unexposed cohorts.

The most interesting issue borne out of application of the NOS was how best to measure subthreshold ATs in the general population. Almost all included studies were rated as poor for use of self-report as opposed to more gold standard interview or observation-based methods (Wells et al., 2004). Although on the surface this seems fair, measurement of ATs using an interview or observation-based method is costly and time-consuming and therefore unlikely to be pragmatic in large longitudinal population-based studies.

There has been much debate about the ‘best’ way to measure ATs across the continuum of trait severity, with the general consensus that multi-method assessment is most appropriate, reliable and valid, comprising a combination of
self-report, observation and multiple respondents such as self, parent and schoolteacher (Möricke, Buitelaar, & Rommelse, 2016). However, as noted, this is not always practical and perhaps not necessary for subthreshold traits.

Perhaps for population-based research, a step in the right direction could be multiple respondent reporting using questionnaire tools as standard practice. The majority of included studies used self-report only for adult samples, or parent-report only for child and/or adolescent samples. Similarly, the study in Part 2 of this project was limited to only a parent-reported AST score. Use of a self-report measure of ATs designed for young people would have also supplemented results, providing an interesting comparison to parent-report. Discrepancies between child and parent reports are a robust finding in child psychopathology (Achenbach, McConaughy, & Howell, 1987; De Lose Reyes & Kazdin, 2005). Future research should adopt child or adolescent self-report in addition to parent report and chart the trajectories of self-reported ATs to explore how discrepant or similar these trajectories are. To the author’s knowledge, this has not yet been conducted.

Use of self-report may offer a powerful insight into young female’s perception of their own ATs. We still do not know what explains the finding of an uptick in female ASTs upon entry to adolescence (Mandy et al., 2018). A possible explanation could be that it reflects an increase in parental expectations of female social skills during this developmental period. Or it may reflect normative systemic changes within family relationships during adolescence in line with individuation theory, such as a normative reduction in connectedness with parents as the daughter attempts to individuate from the family system, which can lead to increased conflict (Collins, 1990; Pinquart & Silbereisen, 2002; Youniss & Smollar, 1985). It could be that mothers find their daughter and/or their daughter's ATs more problematic than they would for a son within the context of their relationship in adolescence (the SCDC was completed by mothers in the ALSPAC study).
In line with this hypothesis, De Lose Reyes and Kazdin (2006) found mother-child discrepancies in perceived child behaviour problems in six to 16 year old males and females referred for problematic behaviour (oppositional, aggressive, antisocial behaviour) were related to mother-child conflict and mediated by maternal stress. Given the established association between higher ATs and externalising behaviours identified in Part 1 of this project, it would be interesting to know if the same is true for ATs. Self-report could help to elucidate what the adolescent increase in parent-reported female ASTs represents, as well as giving a voice to young females.

3.0 We need to talk about women

“Things I hear as a woman with autism: "You don't seem autistic to me””

(Paradiz, Autism Speaks, 2018).

In reviewing evidence for this project, I was struck by the lack of investigation of gender differences across the subthreshold AT literature, inherently limited by the real and harmful gender inequality in how we currently conceptualise, assess, diagnose and treat autism (Mandy & Lai, 2017). Autism comorbidity has been briefly researched in relation to gender differences, however research is limited to clinical samples and inconsistent (Young, Oreve & Speranza, 2018). Women and girls are more likely than men and boys to be diagnosed later in life with autism or not diagnosed at all (Dworsynski et al., 2012; Russell, Steer & Golding, 2011), reducing their access to understanding, support and intervention.

Emerging research shows the female autism phenotype is different to the male phenotype (Mandy & Lai, 2017). Indeed, the most widely used direct observational diagnostic tool, the Autism Diagnostic Observation Schedule-2 (ADOS-2; Lord et al., 2012), is less sensitive to female than male autistic difficulties.
(Mussey, Ginn & Klinger, 2017). A narrative review concludes females show a number of phenotypic differences (Young et al., 2018). Females are socialised in the importance of social interaction more than males, and therefore some autistic females learn to ‘camouflage’ their difficulties more, masking them with mimicked language and behaviour until social demands increase to an unmanageable level (Young et al., 2018). Autistic females score lower on measures of restricted, repetitive patterns of behaviours (RRB) or topics of interest than males, and this is due to true differences rather than measurement of different constructs (Frazier & Hardan, 2017; Grove et al., 2017; Young, Oreve, & Speranza, 2018). RRBs are also expressed differently in females in line with broader differences in the socialisation of normative gender behaviour in western society. For example, female RRBs concern people more than objects therefore are harder to recognise (Young et al., 2018), and traditional gender stereotyped interests such as animals or dancing (Sutherland et al., 2017).

This raises an important implication for researching gender differences within their social and cultural context, given ‘normal’ and therefore ‘less normal’ behaviour differs. This again highlights a major limitation of Part 2 of this project which used the SCDC, a parent-report measure primarily assessing social communication difficulties (Skuse et al., 2005). Although a gender difference was found, this difference may have been larger if a measure of RRB was also included.

Given the dimensionality of autism, it is reasonable to assume this gender inequality extends to subthreshold ATs, however research is severely lacking. A major obstacle is the lack of female-specific AT self-report measures. This is a great opportunity for high quality, co-produced research and co-developed assessment tools. Diagnostic tools that are based on empirically-supported ‘red flags’ for males and females are also needed. By developing equitable assessment tools, we can open the door to equitable understanding and support for women.
4.0 Other limitations and directions for research

A final limitation of Part 2 of this project not previously highlighted was that participants were predominantly White (93.3%) and participants with higher socioeconomic status were more likely to participate, which limits the generalisability of findings. Future studies need to investigate AT trajectories in general population samples that are more diverse and representative, using measurement tools completed by multiple respondents that assess the whole autism constellation.

5.0 References


Dworzynski, K., Ronald, A., Bolton, P., et al. (2012) How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders?


Robinson, E. B., Koenen, K. C., McCormick, M. C., Munir, K., Hallett, V., Happé, F., ... & Ronald, A. (2011). Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). Archives of General Psychiatry, 68(11), 1113-1121.


Appendices
Appendix I: Search terms used in PsychINFO and Medline

Appendix II: Newcastle-Ottawa Quality Assessment Scale (Wells et al., 2014).
NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
COHORT STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability.

1) Representativeness of the exposed cohort
   a) truly representative of the average (describe) in the community ✪
   b) somewhat representative of the average in the community ✪
   c) selected group of users e.g. nurses, volunteers
   d) no description of the derivation of the cohort

2) Selection of the non exposed cohort
   a) drawn from the same community as the exposed cohort ✪
   b) drawn from a different source
   c) no description of the derivation of the non exposed cohort

3) Ascertainment of exposure
   a) secure record (e.g. surgical records) ✪
   b) structured interview ✪
   c) written self report
   d) no description

4) Demonstration that outcome of interest was not present at start of study
   a) yes ✪
   b) no

Comparability

1) Comparability of cohorts on the basis of the design or analysis
   a) study controls for (select the most important factor) ✪
   b) study controls for any additional factor ✪ (This criteria could be modified to indicate specific control for a second important factor.)

Outcome

1) Assessment of outcome
   a) independent blind assessment ✪
   b) record linkage ✪
   c) self report
   d) no description

2) Was follow-up long enough for outcomes to occur
   a) yes (select an adequate follow up period for outcome of interest) ✪
   b) no

3) Adequacy of follow up of cohorts
   a) complete follow up - all subjects accounted for ✪
   b) subjects lost to follow up unlikely to introduce bias - small number lost - > ___ % (select an adequate %) follow up, or description provided of those lost ✪
   c) follow up rate < ___ % (select an adequate %) and no description of those lost
   d) no statement
Appendix III: Participant demographic information
### Participant demographic information for the analytic sample (n=7773)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3810</td>
<td>49.0</td>
</tr>
<tr>
<td>Male</td>
<td>3963</td>
<td>51.0</td>
</tr>
<tr>
<td><strong>Child Ethnic Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>7251</td>
<td>96.4</td>
</tr>
<tr>
<td>Non-White</td>
<td>273</td>
<td>3.5</td>
</tr>
<tr>
<td>Unknown</td>
<td>146</td>
<td>1.9</td>
</tr>
<tr>
<td><strong>Maternal Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>7504</td>
<td>96.5</td>
</tr>
<tr>
<td>Black-Caribbean</td>
<td>23</td>
<td>0.3</td>
</tr>
<tr>
<td>Black-African</td>
<td>5</td>
<td>0.1</td>
</tr>
<tr>
<td>Black-Other</td>
<td>15</td>
<td>0.2</td>
</tr>
<tr>
<td>Chinese</td>
<td>15</td>
<td>0.2</td>
</tr>
<tr>
<td>Indian</td>
<td>20</td>
<td>0.3</td>
</tr>
<tr>
<td>Pakistani</td>
<td>7</td>
<td>0.1</td>
</tr>
<tr>
<td>Other</td>
<td>38</td>
<td>0.5</td>
</tr>
<tr>
<td>Unknown</td>
<td>38</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>Child ASD diagnosis status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother told child has an ASD</td>
<td>80</td>
<td>1.0</td>
</tr>
<tr>
<td>Child aged ≤ 3 years when mother told</td>
<td>24</td>
<td>31.0</td>
</tr>
<tr>
<td>Child aged 4-9 years when mother told</td>
<td>34</td>
<td>68.3</td>
</tr>
<tr>
<td>Unknown child age at diagnosis</td>
<td>22</td>
<td>0.7</td>
</tr>
</tbody>
</table>
### Maternal living status

<table>
<thead>
<tr>
<th>Status</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortgaged/owned</td>
<td>6281</td>
<td>80.9</td>
</tr>
<tr>
<td>Rented (private)</td>
<td>436</td>
<td>5.6</td>
</tr>
<tr>
<td>Rented (council/Housing Association)</td>
<td>704</td>
<td>9.0</td>
</tr>
<tr>
<td>Other</td>
<td>194</td>
<td>2.5</td>
</tr>
<tr>
<td>Unknown</td>
<td>158</td>
<td>2.0</td>
</tr>
</tbody>
</table>

### Maternal education level

<table>
<thead>
<tr>
<th>Level</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSE/None</td>
<td>1038</td>
<td>13.4</td>
</tr>
<tr>
<td>Vocational</td>
<td>663</td>
<td>8.5</td>
</tr>
<tr>
<td>O-Level</td>
<td>2718</td>
<td>35.0</td>
</tr>
<tr>
<td>A-Level</td>
<td>1998</td>
<td>25.7</td>
</tr>
<tr>
<td>University degree</td>
<td>1231</td>
<td>15.8</td>
</tr>
<tr>
<td>Unknown</td>
<td>125</td>
<td>1.6</td>
</tr>
</tbody>
</table>

### Maternal occupation

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional</td>
<td>471</td>
<td>6.1</td>
</tr>
<tr>
<td>Managerial and technical</td>
<td>2247</td>
<td>28.9</td>
</tr>
<tr>
<td>Skilled non-manual</td>
<td>2828</td>
<td>36.4</td>
</tr>
<tr>
<td>Skilled manual</td>
<td>409</td>
<td>5.3</td>
</tr>
<tr>
<td>Partly-skilled manual</td>
<td>532</td>
<td>6.8</td>
</tr>
<tr>
<td>Non-skilled manual</td>
<td>116</td>
<td>1.5</td>
</tr>
<tr>
<td>Armed forces</td>
<td>3</td>
<td>0.0</td>
</tr>
<tr>
<td>Unknown</td>
<td>1167</td>
<td>15.0</td>
</tr>
</tbody>
</table>

*Note. ASD = Autism Spectrum Disorder, CSE = Certificate of Secondary Education, O-Level = General Certificate of Education Ordinary-level taken between 14-16 years old (superseded by the GCSE in 1988).*
Appendix IV: Social and Communication Disorders Checklist (Skuse et al., 2005)
Social and Communication Disorders Checklist

The Social and Communication Disorders Checklist (SCDC) was devised to be simply and quickly rated, comprising just 12 questions. Nine of these serve to measure abnormalities in those aspects of the autistic triad that reflect ‘reciprocal social interaction skills’ and ‘communication skills’. Items 4, 5 and 6 measure behavioural problems in a more general sense, and reflect functional impairment. Each item on the scale is rated according to whether the behaviour has been seen over the past 6 months, and if so whether the associated statements are ‘quite or sometimes true’ or ‘very or often true’. Corresponding scores of 0, 1 and 2 apply, so the maximum possible score is 24. The instrument was originally developed to measure social-behaviour deficits in Turner’s syndrome (Skuse et al., 1997).

Checklist

For each item, please mark the box that best describes your child’s behaviour over the past 6 months.

1. Not aware of other people’s feelings
2. Does not realise when others are upset or angry
3. Does not notice the effect of his/her behaviour on other members of the family
4. Behaviour often disrupts family life
5. Very demanding of other people’s time
6. Difficult to reason with when upset
7. Does not seem to understand social skills, e.g., persistently interrupts conversations
8. Does not pick up on body language
9. Does not appear to understand how to behave when out (e.g., in shops, or other people’s homes)
10. Does not realise if s/he offends people with her/his behaviour
11. Does not respond when told to do something
12. Cannot follow a command unless it is carefully worded

Do you have any other comments or concerns? (If yes, please describe.)
Appendix V: Example syntax specifying multi-group quadratic latent growth curve model with predictors of model intercept, slope and quadratic
Variable: NAMES = ke800b ke802b ke803b ke808b ke805b kb853a kb854a kb855a kb856a ke643 scdc1 scdc2 scdc3 scdc4 male kc968;
USEVARIABLES = ke800b ke802b ke803b ke808b ke805b kb853a kb854a kb855a kb856a ke643 scdc1 scdc2 scdc3 scdc4 male kc968;
MISSING = all(999)
GROUPING = male(0=girl 1=boy)
Analysis: ESTIMATOR = MLR
Define:
CENTER ke800b ke802b ke803b ke808b ke805b kb853a kb854a kb855a kb856a ke643 kc968 (GRANDMEAN);
Model:
i s q i scdc1 @ 1.568141 scdc2 @ 3.221164 scdc3 @ 6.405596
scdc4 @ 9.330376;
scdc1 scdc2 scdc3 scdc4 (same_res);
i ON ke800b ke802b ke803b ke808b ke805b kb853a kb854a kb855a kb856a ke643 kc968;
s ON ke800b ke802b ke803b ke808b ke805b kb853a kb854a kb855a kb856a ke643 kc968;
q ON ke800b ke802b ke803b ke808b ke805b kb853a kb854a kb855a kb856a ke643 kc968;

MODEL girl:
[i] (g_int);
[s] (g_slp);
[q] (g_quad);
i (g_vint);
s (g_vslp);
q (g_vqad);
i with q;
s with q;
i with s;
scdc1 scdc2 scdc3 scdc4 (g_vres);
i ON ke800b ke802b ke803b ke808b ke805b kb853a kb854a kb855a kb856a ke643 kc968;
s ON ke800b ke802b ke803b ke808b ke805b kb853a kb854a kb855a kb856a ke643 kc968;
q ON ke800b ke802b ke803b ke808b ke805b kb853a kb854a kb855a kb856a ke643 kc968;

MODEL boy:
[i] (b_int);
[s] (b_slp);
[q] (b_quad);
i (b_vint);
s (b_vslp);
q (b_vqad);
i with q;
s with q;
i with s;
scdc1 scdc2 scdc3 scdc4 (b_vres);