Latent Trajectories of Autistic Traits in the General Population

Richard Pender

D.Clin.Psy Thesis (Volume 1), 2017

University College London

UCL Doctorate in Clinical Psychology

Thesis declaration form

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: Richard Pender

Date: 7/2017

Overview

This thesis examines developmental trajectories of Autistic Traits (ATs) in a large, general population cohort. Previous research on development of ATs over time is critically reviewed, and limitations and directions for future research are highlighted.

Part 1, the literature review, examines the development of ATs over time, and the ways that heterogeneity between individuals' trajectories has been captured. ATs showed broad improvement over time, which is linked to the social and communication subdomain rather than restricted and repetitive behaviour, which showed greater stability. Substantial heterogeneity between participants was observed, which was characterised by groups specified *a priori* as well as by using Latent Class Growth Models (LCGM). LCGM analyses consistently showed four classes (consistently high, consistently low-moderate, increasing and decreasing) characterised the data. However, no studies applied LCGM to general population samples, which would have important advantages including the ability to detect trajectories increasing from low to severe scores over time, and reducing bias against female participants.

Part 2, the empirical paper, investigates ATs in a large, general population cohort (n = 9,744) using Growth Mixture Modelling (GMM), an LCGM approach. A six-class model was observed, characterised by persistent high, persistent low, later increasing, early increasing, high decreasing and moderate decreasing trajectories. Female participants were found to be overrepresented in the later increasing trajectory, supporting the contention that ATs emerge later in women.

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

Table of Contents

Appendices

Contents of tables and figures

Part 1: Systematic literature review

Part 2: Empirical paper

Acknowledgements

I am grateful to my supervisors Will Mandy and Pasco Fearon for the opportunity to work on this project, and for their generous support, encouragement, insights, and feedback on my drafts. I am also thankful to Jon Heron at the University of Bristol for offering his invaluable thoughts and feedback to a complete beginner in Growth Mixture Modelling. I am grateful to my tutor, Nancy Pistrang, for her support and guidance through the process of developing my thesis, as well as throughout the programme. I would also like to express my thanks to the friends, colleagues and clinical supervisors whose calm reassurance, understanding, empathy and humour has been immensely helpful along the way.

I would like to thank my partner Claire for her encouragement and interest, her unwavering support and her patience while I have been buried in work. Although to be fair, this may have been a blessed relief. Finally I would like to thank my family, without whose generosity, support and belief none of this would have been possible.

Part 1. Systematic Literature Review: Developmental Trajectories of Autistic Traits

Abstract

Aims: While it is well established that Autistic traits (ATs) are distributed across the population, no recent review has examined the development of ATs over time. This review investigates studies reporting change in ATs, and considers the ways in which heterogeneity between individuals' trajectories has been captured. Critical analysis and directions for further research are offered.

Methods: A systematic literature review was conducted using PubMed, PsycINFO and EMBASE. Thirty-two papers meeting inclusion criteria were identified.

Results: Overall ATs were observed to improve over time; this was a general pattern and not restricted to one developmental period. Social communication scores improved over time, while restricted and repetitive behaviour scores remained more stable. Heterogeneity between individuals' AT trajectories was consistently observed, with identified sub-groups showing different patterns of change. Latent class growth model (LCGM) analyses suggested four classes, representing increasing, decreasing, stable severe and stable low-moderate trajectories, can help classify these diverse trajectories of ATs.

Conclusions: ATs show broad improvement over time but this masks the heterogeneity observed between participants. LCGM analyses appear to consistently show four classes can represent this diversity. Future studies should apply LCGM to general population participants to capture increasing ATs and reduce bias against females.

1. Introduction

Autism Spectrum Disorder (ASD) is one of the most commonly-diagnosed neurodevelopmental conditions in the world (Elsabbagh et al., 2012), affecting around 52 million people globally (Baxter et al., 2015). The *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM-5; American Psychiatric Association [APA], 2013) ASD category integrates previous diagnoses of Autism, Asperger Syndrome (AS), Childhood Disintegrative Disorder (CDD) and Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS), grouping symptoms into two primary areas: impairments in social communication and interaction, and restricted and repetitive interests and behaviour (APA, 2013).

Studies investigating ASD across the lifespan have noted broad stability, but also some heterogeneity of outcomes. In a systematic review of diagnostic stability of ASD, Woolfenden et al (2012) observed stability estimates ranging from 53% to 100% for Autistic Disorder and 14% to 100% for ASD over time; they concluded that a minority of participants generally no longer met criteria after a period of follow-up, and the broader ASD diagnoses were especially variable. The possibility has been raised that unstable diagnoses may reflect improving or worsening phenotypes, warranting further investigation (Ozonoff et al., 2015). Indeed, ASD is a complex and heterogeneous neurodevelopmental disorder with varied neurobiological pathways and presentations (Amaral, Schumann, & Nordahl, 2008; DiCicco-Bloom et al 2006; Herbert & Anderson, 2008; Zimmerman, 2008), and as Szatmari et al have concluded: “We have known for some time that ASD is not a homogeneous syndrome. The question is not ‘whether’ to account for heterogeneity but ‘how’” (2009, p. 1465).

Instead of categorically, Autism can be conceptualized dimensionally (Pickles & Angold, 2003), and a growing body of population-based studies has shown ASDs exist at the extreme of a social and communication impairment continuum,

extending from few to large numbers of autistic traits (ATs) exceeding diagnostic thresholds (Constantino, Przybeck, Friesen, & Todd, 2000; Constantino & Todd, 2000; Posserud, Lundervold, & Gillberg, 2006; Robinson et al., 2011a; Robinson et al., 2011b; Skuse et al., 2009; Spiker, Lotspeich, Dimiceli, Myers, & Risch, 2002). Conceptualising Autism as a continuously-distributed dimension of traits has important advantages, including charting developmental profiles and improving prognostic prediction (Baker, Seltzer, & Greenberg, 2011; Lord, Leventhal, & Cook, 2001; Lord et al., 2012a; Soke et al., 2011). Emerging evidence has shown that genetic influences on Autism risk also influence normal-range AT variability (Robinson et al., 2016; St Pourcain et al., 2010), meaning that power and understanding of genetic aetiology of Autism could be greatly increased by accounting for heterogeneity and dimensionality (Chaste et al., 2015). Furthermore, given that factor analytic studies have shown social communication (SC) and restricted and repetitive behaviour (RRB) domains can be separated (Mandy, Charman, & Skuse, 2012), dimensional and symptom-focused approaches allow these domains to be investigated separately.

Longitudinal studies of ATs offer important intellectual and clinical benefits, including improved understanding of nosology and aetiology (Kasari et al., 2014; Magiati, Tay, & Howlin, 2014). Georgiades, Bishop and Frazier (2017) introduced the portmanteau "chronogeneity", defined as heterogeneity in Autism symptomatology across time ("chrono-"). Understanding chronogeneity involves attention to individual and group level variability, together with individual deviance from group norms. Building a comprehension of Autism chronogeneity in this sense will help to clarify whether turning points in trajectories may be related to individual or contextual factors, such that targeted, adaptive intervention approaches (Almirall & Chronos-Tuscano, 2016) could base individual treatment on an understanding of their evolving status (Georgiades et al., 2017).

Only one existing systematic review of longitudinal studies of ATs was located (Seltzer, Shattuck, Abbeduto, & Greenberg, 2004). Twenty-five studies were included, of which ten were prospective designs – of these seven did not report ATs at more than one time point, one was an intervention evaluation, one a study of infantile psychosis, and one included both children and adults at the first time point. Including adults at baseline obscures developmental effects and is likely to introduce a number of key conceptual differences (Magiati, Howlin, & Tay, 2014), reducing comparability. Studies employing retrospective designs are subject to potential recall bias, and cross-sectional studies to age and cohort effects; intervention studies do not seek to capture naturalistic change across all participants and are often restricted to comparably small periods of time. The present review therefore seeks to build on this previous review by systematically searching for prospective longitudinal studies of Autism symptoms, in order to answer the following research questions:

1. How do overall ATs change over time and how do SC and RRB subdomain scores change over time?

2. Is there evidence of heterogeneity in longitudinal trajectories, and is there evidence of latent classes that can represent this heterogeneity?

3. Where latent classes are detected, is class membership consistently associated with demographic, cognitive or clinical variables?

2. Method

2.1 Search strategy

A systematic search was conducted in EMBASE, PsycINFO and PubMed from 1st January 1990 up to and including 1st January 2017. The search consisted of the

following terms in PubMed: (“trajectory” [All Fields] OR “trajectories” [All Fields] OR “developmental course” [All Fields] OR “stability” [All Fields]) AND (“autism” [All Fields] OR “autistic” [All Fields] OR “autistic disorder” [MeSH Terms] OR (“autistic” [All Fields] AND “disorder” [All Fields]) OR “autistic disorder” [All Fields] OR “autism spectrum disorder” [MeSH Terms] OR (“autism” [All Fields] AND “spectrum” [All Fields] AND “disorder” [All Fields]) OR “autism spectrum disorder” [All Fields] OR “asd” [All Fields]). The following terms were used in PsycINFO: (trajectory$.mp. OR (development$.mp. AND course.mp.) OR stability.mp.) AND (exp Autism Spectrum Disorders/ OR autism.mp. OR autistic.mp. OR autism spectrum disorder$.mp. OR ASD.mp.) The following terms were used in EMBASE: (Autism/ OR autism.mp or autistic.mp or asd.mp) AND (trajector$.mp or developmental course.mp or stability.mp). The search was limited to original research studies published in English in peer-reviewed journals.

2.2 Inclusion/exclusion criteria

The following inclusion criteria were applied:

(i) Study reports analysis of ATs

(ii) ATs assessed using a validated instrument

(iii) Autism measure is a continuous scale reflecting severity

(iv) Study is a prospective longitudinal design

(v) ATs are measured on at least two occasions

(vi) Participants are aged <18 at baseline

(vii) The same measure is applied at all time points

The following exclusion criteria were applied:

(i) Studies exclusively measuring diagnostic stability over time, without addressing variability in AT severity

(ii) Studies on specific medical syndromes for which ATs are a typical feature (e.g. Fragile X Syndrome)

2.3 Study selection process

The flow of papers through screening for inclusion in the review is shown in Figure 1**.** The search retrieved 1,847 papers from PubMed, 1,754 papers from PsycINFO and 2,264 papers from EMBASE. These were combined and duplicate papers deleted, leaving a total of 2,161 papers. Titles and abstracts of the papers were screened individually, and 2,010 papers that were clearly unrelated to developmental trajectories of ATs, were case studies, conference abstracts, employed cross-sectional designs or otherwise failed to meet inclusion criteria were excluded. The remaining 151 papers that were considered potentially relevant to the review from initial screening were then read in full. Of these, 32 papers were identified meeting inclusion criteria. Following this, the references of these papers were searched for additional potential papers not retrieved by the search, but none were included after detailed screening of seven studies.

The 32 papers were then categorised into those with two time points (*n* = 19) and those with three or more time points (*n* = 13). Studies with two time points are described in Table 1. Studies with three or more time points were categorised into those that reported analyses without groups, those that reported analyses with *a priori* groups (specified in advance according to theory and/or diagnostic convention), and those that reported analyses to find latent groups (Table 2, Table 3).

2.4 Included studies' summary / coding procedures

After selecting studies for inclusion in the review, relevant information and variables from each study were summarised in Tables 1 to 3. Table 1 reports key features of the studies with two time points, including sample size, percentage of participants who were male, key measures and key outcomes. Table 2 reports participant characteristics of the studies with three or more time points, including sample size, time points, diagnoses, percentage of participants who were male, demographic information, comorbidities and inclusion criteria. Table 3 reports the key outcomes from the studies with three or more time points, including outcome measures, results and identified predictors associated with group membership or differential trajectories. A key is also presented to symbols clarifying which outcomes papers presented (overall ATs, SC, RRB), whether participants' ages were uniform at each time point, and whether ages overlapped by time point (such that some participants were older at baseline than others at follow-up). This process was conducted by the first reviewer, and reviewed and overseen by the supervising reviewers.

3. Results

3.1 Studies with two time points: participants and measures used

Nineteen studies with two time points, meeting inclusion/exclusion criteria were identified (Table 1). Studies were published between 2003 and 2016, with twelve papers having been published prior to the release of DSM-5 (APA, 2013). Sample sizes spanned from 20 to 1,241. Two studies had sample sizes smaller than 30, and three studies had sample sizes greater than 200. Thirteen of the 19 papers focused on those with clinical diagnoses, and a further four on those with suspected diagnoses, meaning that participants lay at the extreme of the continuum of ATs. Reported diagnostic status of participants included ASD, Autism, PDD-NOS, AS,

developmental delay, possible Autism, high familial risk for Autism and low familial risk for Autism. Two studies recruited participants with the diagnosis of Pervasive Developmental Disorder (PDD); not to be confused with PDD-NOS, PDD is an *International Statistical Classification of Diseases and Related Health Problems* (10th ed.; ICD-10; World Health Organization [WHO], 1992) diagnostic category encompassing Autistic disorder, AS, CDD, PDD-NOS and Rett's Syndrome (Constantino et al., 2009; Darrou et al., 2010). The percentage of male participants ranged from 50 to 100%, with a median percentage of 79%, in line with the estimate of a 4:1 sex ratio found in passive case ascertainment studies (identifying only cases already diagnosed by services), but notably over-representing males compared with active case ascertainment studies (actively identifying cases regardless of prior diagnosis) in which the ratio is closer to 3:1 (Loomes, Hull, & Mandy, 2017).

Ages of participants at baseline ranged from 6 months to a mean of 12 years 8 months, and at follow-up from 12 months to a mean of 19 years. Studies varied between those in which all participants were of the same age at given time points, and those in which participants' ages lay within a reported range. In some of these studies, there was an overlap between ages at time points such that some participants were younger at follow-up than other participants had been at baseline.

Measures used were relatively heterogeneous across studies. Seven studies reported outcomes measured by the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1999), seven studies the Autism Diagnostic Interview (ADI; Le Couteur et al., 1989) or Autism Diagnostic Interview-Revised (ADI-R; Le Couteur, Lord & Rutter, 2003), two studies the Childhood Autism Rating Scale (CARS; Schopler, Reichler, & Renner, 2002), and one study each of the Social Responsiveness Scale (SRS; Constantino & Gruber, 2007), Social Communication

Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003) and Autism Observational Scale for Infants (AOSI; Bryson, Zwaigenbaum, McDermott, Rombough, & Brian, 2008).

3.2 Outcomes of two time-point studies

3.2.1 *Trajectories of overall ATs*

Eleven papers were located with two time points that reported trajectories of change in overall levels of ATs. All papers tracked absolute change (rise and fall in mean and/or individual scores), rather than relative change (differences in individuals' place in the overall distribution across time). Seven papers (Darrou et al., 2010; Estes et al., 2015; Lord et al., 2006; Moss, Magiati, Charman, & Howlin, 2008; Pellicano, 2012; Soke et al., 2011; Yirmiya, Seidman, Koren-Karie, Oppenheim, & Dolev, 2015) reported significant improvement in symptoms over time, and four (Eaves & Ho, 2004; Kim, Macari, Koller, & Chawarska, 2016; Louwerse et al., 2015; Postorino et al., 2015) reported no significant change and broad stability of symptom profiles. Seven papers (Darrou et al., 2010; Eaves and Ho, 2004; Estes et al., 2015; Kim et al., 2016; Louwerse et al., 2015; Moss et al., 2008; Yirmiya et al., 2015) provided evidence suggestive of heterogeneity between individuals in terms of change trajectories. This is important because in a heterogeneous population, a single statistic averaging out change across participants may miss large variation between different trajectories.

Of the papers that reported no significant change in ATs, two (Eaves and Ho, 2004; Louwerse et al., 2015) noted the heterogeneity of change between individuals, masked by the overall stability. Eaves and Ho (2004) found that change in CARS scores in participants referred with possible Autism ranged from -14 to +14, with 33% of participants changing 5 or more points up or down (a change of 8 points represents significant change at the p =.05 level; Schopler et al., 2002). Further,

Louwerse et al (2015) observed that despite the broad stability of traits in participants with PDD-NOS between age 9 and 16, 40% showed increasing symptoms and 20% decreased on the ADOS.

Two papers characterised the heterogeneity of change by comparing sub-groups of participants. Yirmiya et al (2015) found that while an overall significant improvement was shown in ATs on the ADOS between 8 and 11 years of age, an opposite change pattern was observed by sex – boys revealed a significant decrease in severity of ATs, while girls showed an increase in scores. Estes et al (2015) observed a significant decrease in AOSI scores between 6 and 12 months of age in children with high and low risk for Autism, with the most marked improvements occurring in the high-risk but no diagnosis group.

Darrou et al (2010) observed overall improved CARS scores in children with PDD, but a more nuanced picture of heterogeneity emerged from a Latent Class Analysis. High and low severity groups emerged at each time point, and 28% of participants were observed to transition between low and high level groups over time. A further study conducted a cluster analysis at baseline, finding four groups with stable ATs and one cluster with increasing ADOS scores (Kim et al., 2016). Finally Moss et al (2008) reported quartile ranges for the degree of change shown for overall and subdomain ADI-R scores, demonstrating a broadly even spread of participants between quartiles.

Overall, papers were divided between finding broad stability and broad improvement with a majority of studies reporting the latter. Studies reporting inter-individual differences noted significant heterogeneity in trait trajectories.

3.2.2 Trajectories of Social and Communication difficulties

Eleven of the included studies reported outcomes for change in SC scores. Of these studies, seven reported significant decreases in severity (Barbaro & Dissanayake, 2016; Chawarska, Klin, Paul, & Volkmar, 2007; Constantino et al., 2009; McGovern & Sigman, 2005; Moss et al., 2008; Pellicano, 2012; Soke et al., 2011), one reported no significant change (Moore & Goodson, 2003), and three reported differential outcomes between sub-groups of participants (Honey, McConachie, Randle, Shearer, & Le Couteur, 2006; Messinger et al., 2015; Starr, Szatmari, Bryson, & Zwaigenbaum, 2003). A clear majority of papers therefore suggested a pattern of significant decline in SC difficulties over time.

There were indicators of heterogeneity where it was investigated. Five papers made comparisons between *a priori* groups; two found scores in one group increased at a greater rate than in the other, and three found that scores increased for one group and decreased or remained stable for another. McGovern and Sigman (2005) compared individuals with IQ ≥ 70 with IQ < 70, and found the higher IQ group showed greater improvement on the ADI-R SC sub-domain between 12 and 19 years, while Starr et al (2003) observed larger reductions in ADI scores in participants with Autism diagnoses than AS between 6 and 8 years. Honey et al (2006) found groups with Autism and ASD diagnoses showed increased scores while those in a language delay group showed decreased scores. Messinger et al (2015) reported that between 24 and 36 months, a group of children with high risk for ASD but without diagnoses did not demonstrate the score increases found for other participants. Moore and Goodson (2003) noted that 10 of the children with Autism diagnoses showed increases in ADI-R scores while nine showed clear decreases between 2 and 4.5 years of age.

Overall, there was strong evidence for improvement over time, with seven of eleven papers reporting decreasing trajectories of scores. Consistently with the literature on overall ATs, papers also reported heterogeneity between participants, with identified groups showing differential trajectories.

3.2.3 Trajectories of Restricted and Repetitive Behaviours

Ten of the included studies reported outcomes for change in RRB scores. Of these, six (Barbaro & Dissanayake, 2016; Chawarska et al., 2007; Messinger et al., 2015; Moss et al., 2008; Soke et al., 2011; Starr et al., 2003) found no significant change over time, two found a significant decrease (McGovern & Sigman, 2005; Pellicano, 2012), one a significant increase (Moore and Goodson, 2003), and one different profile between groups (Honey, et al, 2006). This latter paper compared scores between ASD, Autistic Disorder and 'Other' groups, and found that a significant increase was shown for ASD only. McGovern and Sigman (2005) also showed that RRB scores decreased significantly more in a group with IQ ≥ 70 than in a lower IQ group. Moss et al (2008) reported quartile scores for degree of change, revealing an evenly distributed degree of change between participants. Hence overall there was evidence for broad stability of RRB scores, with heterogeneity observed again between participants.

In conclusion, the literature reporting AT trajectories between two time points suggests improvement in overall ATs over time. This appears to be linked to the SC subdomain, while RRB shows greater stability over time.

3.3 Outcomes of studies with three or more time points

3.3.1 Studies with three or more time points: participants and measures used

Thirteen studies with three or more time points, meeting inclusion/exclusion criteria were identified (Table 2, 3). Studies were published between 2005 and 2017, with eight papers having been published prior to the release of DSM-5 (APA, 2013). Sample sizes ranged from 26 to 6,539. One study had a sample size smaller than 30, and five studies had sample sizes greater than 200. Reported diagnostic status of participants included Autism, ASD, later-born siblings, AS and non-spectrum. The percentage of participants that were male ranged from 50.1% to 92.3%, with a median of 83%. Ages of participants at baseline ranged from 18 months to 7 years, and at follow-up from 36 months to 19 years. Two studies included varying time points per participant.

Seven studies reported outcomes measured by the ADOS, three studies the ADI-R, and one study each the Repetitive Behavior Scale-Revised (RBS-R; Lam & Aman, 2007), the Social and Communication Disorders Checklist (SCDC; Skuse, Mandy, & Scourfield, 2005) and the Autism Behavior Checklist (ABC; Krug, Arick, & Almond, 1988).

3.3.2 Analyses without groups

Two papers investigated change in symptom trajectories across participants, without comparisons between subgroups. Charman et al (2005) investigated changes in ADI-R scores in 26 participants with Autism diagnoses at three time points (2, 3 and 7 years) using paired *t*-tests. A significant reduction was found in both Reciprocal Social Interaction (RSI) and Non-Verbal Communication (NVC) domains; in the RRB domain, scores increased between 3 years and 4-5 years, and then decreased to the final time point. Substantial heterogeneity was observed within these

trajectories, and variances were noted to increase over time. This was not captured statistically; the authors noted an initial pattern of uniformly high scores, and a wide range from very low to near-ceiling scores by 7 years. Joseph, Thurm, Farmer and Shumway (2013) investigated change over time among 39 participants diagnosed with Autism on the RBS-R using linear mixed models, and found no significant changes between the three time points (mean ages 4, 5.4 and 6.2 years). Evidence from these papers supports the finding from studies with two time points, showing improvement in SC scores but stability of RRB scores over time.

3.3.3 Analyses with a priori groups

Five studies investigated trajectories for groups that were specified by the researchers *a priori*. This has the advantage that heterogeneity within the data can be explored, although the disadvantage relative to analyses with latent classes is that the groups do not emerge from the data, but are specified in advance according to theory and/or diagnostic convention.

Lord, Bishop and Anderson (2015) conducted growth curve analyses to examine changes in scores in ADI-R SC, repetitive sensorimotor (RSM) and insistence on sameness (IS) subscales among 85 participants with ASD at five time points (2, 3, 5, 9 and 19 years), divided between three groups (Verbal IQ [VIQ] <70, VIQ ≥ 70 and Very Positive Outcome [VPO]). RSM and IS are separable factors of RRB (Cuccaro et al., 2003), representing “higher order” (restricted interests and routines) or “lower order” (repetitive motor movements and preoccupation with parts of objects) aspects. Intercepts (baseline scores) for SC were highest for the less cognitively able group, followed by the more able group; VPO mean scores were lowest. The less cognitively able and VPO group scores declined following a quadratic (accelerating) trajectory; the more cognitively able group’s scores declined following a linear trajectory. For RSM scores, the intercepts were significantly

different following an identical pattern. Linear slopes followed a gradual decline for all groups. For IS, the only slope that changed significantly was for the lower ability group, which showed an increase over time; there were no significant differences in intercepts. The broad stability in RRB scores and improvements in SC scores is consistent with the trend of evidence reviewed above. However, the differential change between IS and RSM factors supports the finding that RRB can be characterised by these two factors (Cuccaro et al., 2003) and suggests that RRB may itself be a heterogeneous construct, and the observed stability may obscure different trajectories between participants within these separate factors.

Ozonoff et al (2015) used mixed-effects linear models (Laird & Ware, 1982) to investigate changes in ADOS SC scores among 418 later-born biological siblings of a child with ASD, 26% of whom had ASD diagnoses by 3 years of age. Analyses investigated change in symptoms at three time points (18, 24 and 36 months), divided between four groups (True Positive [TP], True Negative [TN], False Negative [FN] and False Positive [FP]), which were defined by stability of diagnosis until 36 months. The TP group showed a consistently severe trajectory; the TN group a consistently low trajectory. The FN group followed a positive (increasing) slope, while the FP group showed an initially increasing trajectory, followed by a rapid decrease, with scores at 36 months still significantly higher than the TN group, suggesting they were sub-threshold for diagnosis but scores remained atypical. Hence FP and FN diagnoses represented different trajectories of ATs, one that increased over time to reach the severe range, and one that improved from the severe range but remained more severe than a consistently low group.

Robinson et al (2011a) used growth models to investigate changes in scores on the SCDC among 6,539 participants in the Avon Longitudinal Study of Parents and Children (ALSPAC), a general population cohort study, at three time points (7, 10 and 13 years of age). Analyses were conducted on the complete sample,

participants with ASD diagnoses, participants in the 90th and 95th percentiles for IQ, and those scoring above the cut-off of 8 points (a time-invariant, receiver operating characteristic-maximising score) – each of these groups was further divided by sex. No significant difference in scores was found between baseline and 13 years in any group. All groups followed a similar pattern, in which scores decreased slightly between 7 and 10 years, and returned to baseline by 13; one exception was that when boys alone were analysed, scores at 13 were slightly lower than baseline, reaching significance largely because of the large sample size.

Szatmari et al (2009) employed hierarchical linear models to investigate changes in ABC scores for 57 participants over four time points (6-8, 10-14, 14-17 and 17-19 years), divided into two groups based on presence of Structural Language Impairment (SLI) at 6-8 years – those with an absence of SLI were categorised as an AS group, those showing SLI as an Autism group. Addition of the SLI criterion improved model fit. Over time, both groups showed decreasing trajectories in ABC scores, with scores for the AS group significantly lower at all time points than the Autism group.

Clark et al (2017) investigated ATs in 48 participants with ASD at three time points (24 months, 48 months and 7-9 years) between those with stable ASD diagnoses following baseline, and those with non-stable diagnoses, employing a mixed 2 (group) x 3 (age) repeated measures ANOVA on ADOS-G scores. A significant Group × Age interaction was identified (F (45, 2) = 6.11, p = .004, $\eta 2$ = .214). Symptoms in the stable group decreased significantly at 48 months, increasing again by 7-9 years such that there was no significant difference in symptom severity between baseline and final assessment. The non-stable group decreased significantly in severity between 24 and 48 months, maintaining decreased scores at 7-9 years.

In summary, consistently with evidence from studies with two time points, different trajectories can be discerned between groups of participants, revealing that a single mean statistic of change across participants obscures meaningful heterogeneity between trajectories of ATs. Overall, the variables associated with group membership did not appear to be linked with consistent results, with higher IQ groups showing greater improvement in SC and RRB in one study (McGovern & Sigman, 2005) but only differences in intercepts (lower IQ higher at baseline) in another (Lord et al, 2015). Similarly, Autism was linked to greater decrease in ATs than AS in one study (Starr et al, 2003) and only to differences between intercepts in another (Szatmari et al, 2009). The only studies to have divided groups *a priori* by sex found boys to improve over time, with girls showing either increasing (Yirmiya et al, 2015) or stable (Robinson et al, 2011a) ATs.

3.3.4 Analyses with latent classes

Six of the papers that included three or more time points carried out analyses to determine latent classes within the overall growth trajectories of participants. In studies that modelled development as a single trajectory across participants, it was frequently observed that notable heterogeneity was present. Analyses with latent classes are interesting in this respect, as they do not assume that a single growth trajectory is sufficient to describe the data. By relaxing the assumption that all participants are drawn from a single, homogeneous population, latent class analyses are able to determine multiple growth trajectories. These designs have the advantage relative to *a priori* group designs that these groups emerge from the data, rather than being specified by theory or convention.

Gotham et al (2012) applied a latent class growth curve model framework to examine latent growth trajectories in ADOS scores over two to eight time points per participant (various times) in 335 children with ASD diagnoses and 10 children

without ASD diagnoses. The authors investigated linear and quadratic models from three to six classes, and selected that which had the lowest Bayesian Information Criteria (BIC) value. BIC reflects the log-likelihood of a model, whilst penalising for parameters and observations, and performs well as a basis for optimal model selection (Nylund et al., 2007). The authors selected a linear five-class model on this basis. One class included six participants and was dropped from further analysis for this reason; an arguably preferable approach would have been to have selected the four-class model without excluding participants (Berlin, Parra & Williams, 2013).

The four classes that emerged included a persistent high-severity class (46%), a persistent moderate-severity class (38%), an increasing-severity class (9%), and a decreasing-severity class (7%). The average probability of class assignment (reflecting confidence that the participants should be allocated to their chosen classes) was good for the high, moderate and decreasing classes (p =. 79 to .82) but moderate (p =.68) for the moderately severe class (Muthén, 2004). The model was limited by intra-class variability, with 30% of participants in the worsening class for example showing wide variability across time, with final scores improved relative to baseline in some cases. Furthermore, ADI-R scores were reported between classes, and it was observed that within the worsening class mean Nonverbal Communication and Social subdomain scores actually improved relative to baseline on average across the class (significance tests not applied).

Lord and Luyster (2006) applied a latent growth curve analysis to 297 children with Autism or PDD-NOS diagnoses at four time points (2, 3, 5 and 9 years old) on Pre-Linguistic ADOS (PL-ADOS) and ADOS SC subscale severity scores. Four profiles emerged, however no fit indices were reported, along with no information about how the four-class model was selected. The classes reported were a consistently severe class, a consistently low-scoring class, an increasing class and a decreasing class.

Mean scores for the increasing class rose between baseline and age 9, with a more pronounced rise between ages 2 and 5; mean scores for the decreasing class dropped dramatically between 2 and 5, and then remained stable from 5 to 9 years. However, numbers of members in each class, diagnostic composition and class membership probabilities were not reported, and so the validity of the model is difficult to independently evaluate. Covariates were not investigated.

Lord et al (2012a) employed a latent age-related growth modelling approach to ADOS scores for 78 participants characterised as ASD, PDD-NOS and non-spectrum (n = 6) over two to 10 assessments per individual at varying time points. Initial appointments were at 12 to 19 months, and participants were seen an average of six times over 20 months. A linear four-class model was selected, yielding a consistently severe class (21%), a worsening class (21%), an improving class (19%), and a non-spectrum class (40%). It is notable that this model is consistent with the emerging overall pattern, and that the greater proportion of non-spectrum participants may have contributed to the larger consistently low-scoring class. The average probabilities for being assigned to the most likely class were very high (p =.84 to p = .99; Muthén, 2004). The non-spectrum class included all children who had never received an ASD diagnosis and one child who had received an ASD diagnosis only once. Examining subscales by class, there was no significant evidence of change over time in social affect (SA) or RRB scores in the consistently high class, while in the worsening class, both SA and RRB increased significantly. In the improving class, while SA decreased, the RRB score did not significantly improve. In the non-spectrum class, there was a marginal decrease in SA but not RRB. Therefore overall, classes appeared to capture meaningful and coherent trajectories.

Venker et al (2014) applied a Latent Class Growth Analysis (LCGA; Muthén & Muthén, 2000) to 129 participants with diagnoses of Autism and PDD-NOS over four

time points (2.5, 3.5, 4.5 and 5.5 years of age) on ADOS scores. A LCGA assumes no variance within latent classes, so variances for intercept, slope and quadratic growth factors are constrained to zero. A linear four-class model was selected, revealing a persistent high-severity class (36%), a persistent moderate-severity class (42%), a worsening class (8%) and an improving class (14%). Quality of reporting was high, and sample size-adjusted BIC values were shown to support selection of the four-class model.

Szatmari et al (2015) employed a semi-parametric latent growth modelling approach (Nagin, 2005) to data from 421 participants with ASD at three time points (2-5 years old, 6 months later and 6 years old) on the ADOS. A two-class model was selected, with one group showing less severe symptoms and a statistically significant improving trajectory (11.4%), and another group (88.6%) showing a severe and stable trajectory. BIC values continued to decrease for three classes despite the two-class model having been selected. Fit indices for quadratic models were not reported. Therefore it is unclear on what basis the two-class linear model was selected and we must infer that quadratic models either failed to converge or produced inferior fit indices, and that superior class probabilities formed the basis for this selection. It is notable that other papers included in this review found four latent classes of similar qualities, and the decision to use a two-class model despite continually improving BIC values could help to explain this discrepancy.

Richler et al (2010) investigated latent classes within RRB subscales of the ADI-R using Growth Curve Analyses in 192 participants with diagnoses of Autism and PDD-NOS at four time points (2, 3, 5 and 9 years old). RRB was divided into Repetitive Sensorimotor (RSM) and Insistence on Sameness (IS) factors. For RSM, a three-class linear solution emerged, including a mild group (25%), a slightly decreasing group (50%), and a consistently severe group (25%). For IS, a three-class quadratic model provided the best fit, yielding a mild group (13%), an

increasing group (71%), and a moderate group (16%). This supports the separation of RSM and IS factors (Cuccaro et al., 2003) within RRB, and suggests that over time RSM showed a wider range with half of participants showing improving scores, and IS a narrower range with a majority of participants showing worsening scores over time.

In summary, there was no convincing evidence to contradict the consistent finding that four latent trajectories (consistently high, consistently low-moderate, improving and worsening) represent change in overall ATs over time (Gotham et al., 2012; Lord et al., 2012a; Szatmari et al., 2015; Venker et al., 2014), while a study on RRB (Richler et al., 2010) suggested three classes captured trajectories in IS (mild, increasing, moderate) and RSM (mild, decreasing, severe). The observation of different change profiles in IS and RSM factors replicates the findings of Lord et al (2015), and supports the suggestion that RRB is itself a heterogeneous construct (Cuccaro et al., 2003). Both studies observed declining RSM scores, while IS scores were observed to be stable (Lord et al., 2015) or mostly worsening (Richler et al., 2010). This latter paper also suggested heterogeneity within these factors can be observed, which may account for the differences in findings. The stability in overall RRB scores found in single-trajectory studies may obscure a possible tendency for IS scores to worsen and RSM to improve for some participants over time.

3.3.5 Predictors of latent class membership

Four studies used multivariate regression techniques to investigate whether a range of cognitive, clinical and demographic variables (sex, ethnicity, VIQ, non-verbal IQ [NVIQ], daily living skills, receptive and expressive language and language loss) predicted membership of the observed latent trajectory classes (Gotham et al., 2012; Lord et al., 2012; Szatmari et al., 2015; Venker et al., 2014). Ethnicity and maternal education were consistently observed to be unrelated to class membership

(Gotham et al., 2012; Lord et al., 2012; Venker et al., 2012). Three studies found sex did not significantly membership of latent classes (Gotham et al., 2012; Lord et al, 2012; Venker et al, 2014), while one study (Venker et al, 2015) observed that boys were more likely to be classified in the stable and severe class, and girls to be included in the improving class. This appears to contradict the finding from studies using *a priori* groups (Yirmiya et al., 2015; Robinson et al., 2011a), which suggested boys showed improving trajectories and girls worsening or stable scores over time. Hence overall no consistent relationship between sex and AT trajectories was observed. Only one study found VIQ to significantly predict membership in an improving class (Gotham et al., 2012), and one found NVIQ to predict membership in a stable severe class. This suggests an inconsistent relationship between cognitive ability and group, as each study observed no significant relationship to support the other. Daily living skills were significantly related to membership of an improving trajectory class (Gotham et al., 2012) and slower increases in receptive and expressive language to membership of a stable severe trajectory class (Venker et al., 2014); however, these particular analyses were unique to these specific studies and therefore need replication. Overall, there was no consistent evidence of any variable significantly predicting AT trajectories.

4. Discussion

This paper systematically reviewed the literature on AT trajectories, dividing papers into those with two timepoints, and those with three or more timepoints. The latter were further subdivided into those that conducted analyses across all participants, those that investigated trajectories between groups specified *a priori*, and those that investigated and reported trajectories of latent classes within the data.

There was large variability in sample size, recruitment and measurement, although the majority of papers employed the ADOS or ADI-R. Visual analysis of the

characteristics of studies alongside findings did not suggest that time between measurements, overlapping ages at time points, sample size, measure, diagnostic category or developmental stage appeared to be consistently related to outcome.

Measures used were heterogeneous, including instruments relying on self, parent or caregiver report (e.g. SCQ, SRS, SCDC, CARS, ADI-R) and observation by trained practitioners (e.g. ADOS). Parental report has a number of potential disadvantages including its subjectivity, and due to varying degrees of concordance between raters it has been argued that a multi-method, multi-informant approach is essential for gold standard diagnostic procedure (Möricke, Buitelaar, & Rommelse, 2016). However, in this review measures used did not appear to consistently be related to outcomes, with representative spreads occurring within each type of measure. Within use of the ADOS, employment of updated algorithms such as the CSS that controls for VIQ and age effects (Gotham et al., 2009), also did not appear to be related to a consistently unique pattern of outcomes.

Amongst papers with two time points, a majority of studies (seven out of 11) showed decreases in mean AT scores over time, with four studies finding no significant change and no studies showing an increase over time. There appeared to be good evidence that this represents a general tendency towards improvement over time, rather than being limited to any specific developmental period identified. However, no papers continued investigating ATs long into adulthood, and so it is unclear from this review whether this tendency towards improvement extends across the lifespan. A similar pattern emerged within the SC symptom subdomain, with seven out of 11 papers reporting decreases over time, one finding no significant change, and three showing different outcomes for different subgroups. Within the RRB domain, the weight of evidence appeared to suggest a greater stability over time, with six of 10 papers finding no significant change, two a significant decrease, one a significant increase and one different outcomes between subgroups. Two of the studies with

three or more time points conducted analyses across all participants – the finding from these of broad stability of RRB, but a significant decrease in SC scores over time is consistent with the observed tendency in studies with two time points. Therefore as a broad conclusion, it would appear that the tendency for improvement in overall scores over time seems to be more clearly related to the SC subdomain than RRB, supporting the fractionation of autism symptoms hypothesis (Mandy et al., 2012). Moreover, two studies (Lord et al., 2015; Richler et al., 2010) supported the finding that the RRB domain can itself be separated into IS and RSM factors (Cuccaro et al., 2003), with preliminary suggestions that an overall pattern of stability in IS and improvement in RSM (Lord et al., 2015) may be a broad characterisation of around half of participants showing decreased RSM scores and a majority showing increased IS scores over time (Richler et al., 2010). Further studies are needed to replicate and clarify these findings.

It was notable that the quality of studies overall was very high, possibly as a result of the stringent inclusion criteria, together with a range of well-validated measures and well-established diagnostic practices for ASD. One area in which studies diverged was in generalisability – those studies which recruited participants in the 1990s, when diagnostic criteria were more stringent and early diagnosis was less common, may be less representative of more recent cohorts. Further limitations amongst papers with two time points included one study (Postorino et al., 2015) failing to report the ages of participants at baseline and follow-up, making interpretation in terms of developmental trajectories impossible. Interpretation of findings in these studies was further complicated by the fact that five studies showed clear overlaps of ages, and two studies showed probable overlaps, between baseline and follow-up, such that some participants were older at time one than others at time two (Table 1). These studies did not show a clear difference from the overall pattern of findings. However, only a general trend towards improvement over time can be

inferred from these studies, rather than a clear developmental trajectory between ages.

Of the papers investigating *a priori* groups, two compared stable diagnostic groups with non-stable diagnostic groups. These studies were useful in demonstrating that the consistent minority of participants with non-stable diagnoses (Woolfenden et al., 2012) do indeed represent an intermediate phenotype whose ATs change over time. However the broader participant group recruited by Ozonoff et al (2015), composed of later-born siblings of children with ASD, usefully permitted an increasing phenotype to be charted, who did not meet diagnostic criteria at the first time point, but whose scores increase over time. This particular phenotype is more likely to be missed by studies limited only to participants who already have received ASD diagnoses.

Two further studies investigated groups based on cognitive factors – Lord et al (2015) separated groups on the basis of VIQ level, Szatmari et al (2009) based on the presence or absence of SLI. On the whole both of these studies produced parallel trajectories between groups, revealing differences were most meaningfully located in differences between intercepts rather than slopes, suggesting the groups had rather limited utility in conceptualising change over time.

Only two studies (Constantino et al., 2009; Robinson et al., 2011a) included a substantial proportion of general population participants. Papers that investigate trajectories mostly in participants with diagnoses are important, but this reduces power to detect any subgroup that may move from non-clinical status into the clinical range over time. Furthermore, the possibility cannot be excluded that diagnostic procedures may bias against girls, meaning that the trajectories of change detected in these studies are weighted heavily towards a male phenotype (Loomes et al., 2017). Indeed, the median proportion of males in the studies

reviewed above was 79%, according with the sex ratio of passive recruitment studies. In the study by Constantino et al (2009), this opportunity was not available due to exclusively male participants. In the study by Robinson et al (2011a), a broad pattern of stability was reflected within all the *a priori* groups, but it is possible that latent class analyses could reveal that the pattern of stability observed by Robinson et al does not hold for all participants.

Of the six papers investigating latent classes in the data, four selected models with four classes, which were characterised in highly-similar ways, including a consistently high class, a consistently low or moderate class, an increasing class and a decreasing class (Gotham et al., 2012; Lord & Luyster, 2006; Lord et al., 2012a; Venker et al., 2014). This therefore emerged as a relatively robust finding from the scope of the papers identified. One further paper found a two-class model for total calibrated severity ADOS scores (Szatmari et al., 2015). However, the authors reported that they selected this model on the basis of BIC values and posterior probabilities, although BIC values provided did not support selection of the two-class model, and continued to decline (improved fit) for three classes. Further numbers of classes were also not reported, and therefore the optimal number of latent classes according to this index of model fit was not possible to determine. One further paper (Richler et al., 2010) found three-class solutions for modelling repetitive behaviours. In short, there were no convincing models of overall ATs to contradict the four-class model.

Aspects of the overall reporting and methodology of the latent class analyses were problematic. Lord and Luyster (2006) did not report on what basis the four-class model was selected, nor were numbers of group members or class posterior probabilities reported. Richler et al (2010) also did not report BIC values, which they declared to be the basis for their model selection, and Szatmari et al (2015) selected a two-class model despite improving BIC (also stated to be the basis for their model

selection) for greater numbers of classes. However overall the four-class model appears (with the above caveats) to emerge as the best-supported model from the evidence available so far. While class memberships were not reported by Lord and Luyster (2006), it could be observed that there was a high degree of similarity between the relative class percentages reported by Gotham et al (2012) and Venker et al (2014). On the other hand, Lord et al (2012a) recruited non-spectrum participants into their study alongside participants with ASD diagnoses, and it is interesting to note that while the consistent moderate class observed in the other two studies became a consistent low class, the membership of the increasing trajectory also substantially increased (21% vs. 8%, 9%). This further supports the interpretation that the observation of a phenotype with increasing symptoms over time depends very much on the inclusion of participants who do not currently have diagnoses.

A study by Fountain et al (2012) did not meet full criteria for inclusion in this review because symptom scores were measured using the Californian Client Development Evaluation report (CDER), which is not a validated diagnostic instrument. Group based latent trajectory modelling applied to 6,938 children with Autism Disorder diagnoses across various age-related time points revealed a six-class model for SC and RRB scores. For communication and social domains respectively, groups identified were high (12.7%, 7.0%), bloomers (7.5%, 10.7%), medium-high (25.8%, 19.6%), medium (24.5%, 29.7%), low-medium (18.4%, 24.5%) and low (11.1%, 8.4%), representing substantial development across groups, with variation in its extent. For RRB, groups were never (21.4%), improving (8.1%), declining (7.1%), when stressed (28.0%), daily (27.6%) and usually (7.8%), representing substantial stability, with one increasing and one decreasing class. It is interesting that these patterns represent more nuanced breakdowns of the broader pictures emerging from the two-time point studies, while the absence of any worsening group in social

and communication trajectories differs from the general findings from other latent class studies. However, the use of a non-validated measure, taken together with the failure to report any fit indices to support the basis for model selection, severely limits the quality of this study relative to these reported above.

One key limitation is that the intermediate worsening and improving trajectories could represent an intermediate phenotype that is variable over time, rather than a definitively improving or worsening trajectory – these labels may reflect separating out those whose regular score fluctuations have changed in one direction or the other on a particular occasion. Indeed, further follow-up into adulthood would be needed to demonstrate whether these changes in scores are sustained. Furthermore, Gotham et al (2012) noted heterogeneity within the worsening class, with 30% having variable outcomes and some improving outcomes, and showing improved mean ADI-R scores for nonverbal communication and social subdomains. Additionally, Venker et al (2014) reported relatively moderate posterior class probabilities for improving and worsening classes (p = .73, .77 respectively), and noted that 20% of the worsening class and 39% of the improving class did not significantly change over time. However, in the study that included non-spectrum participants, Lord et al (2012a) observed better posterior probabilities (p =.87 for worsening class), and did not report such fluctuations. Again, it could be that the recruitment of non-spectrum participants made an important addition to the validity of the intermediate phenotypes, and further research is needed to clarify this.

No convincing evidence emerged of any variable consistently and significantly predicting latent class membership. One study found VIQ to be linked to class membership (Gotham et al 2012), one found NVIQ, daily living skills, receptive and expressive language to predict membership (Venker et al., 2014), and one paper found sex to be a significant predictor (Szatmari et al., 2015). No findings were replicated between studies. Further, findings from *a priori* groups were similarly

contradictory, differing on how IQ (McGovern & Sigman, 2005; Lord et al, 2015) or Autism versus AS (Starr et al, 2003; Szatmari et al, 2009) are related to AT trajectories. While female sex was found to relate to a decreasing trajectory latent class (Szatmari et al, 2015), another study found females to show increasing (Yirmiya et al, 2015) scores, and another stable (Robinson et al, 2011a) ATs.

In summary, the evidence reviewed shows a tendency for a general improvement in ATs over time, which is more pronounced in the SC than in the RRB subdomain. However within this picture, there was evidence for heterogeneity between individual trajectories, and there is some support for this best being characterised by four trajectories, one consistently high, one consistently low, and an improving and a worsening trajectory over time.

4.1 Limitations of the present review

The present review was limited in its ability to explore why some studies found positive results and others did not – a visual inspection of various important factors (sample size, time spaces between observations, whether observations overlapped, measures used and percentage of male participants) appeared to show that within each level of each predictor, a similar spread of results emerged. It would be helpful if meta-analytic techniques could be applied, but this would require a far more homogeneous literature. It was also beyond the scope of this review to investigate trajectories relating to language, daily living skills, VIQ and NVIQ, which were also reported in a number of studies. Future reviews focusing on development of these characteristics over time would be useful. Finally, it was also not possible for this study to systematically rate studies using a quality appraisal tool, as piloting of tools revealed problematic ceiling effects with almost all studies scoring very highly. This may have resulted from the stringent inclusion criteria in this review, the high-quality, well-validated measures available, and the rigorous clinical practice and

recruitment procedures employed. The task therefore fell on more conceptual critiques relating to diagnostic versus dimensional approaches to ATs, the power of studies to include increasing score trajectories, and the risk of bias against females.

4.2 Directions for future research

Based on the observations of this review, it will be helpful for future research to use general population participants. ATs have been shown to be continuously distributed across the population, and recruitment of participants that are not limited to those with diagnoses would allow potential increasing AT phenotypes to be detected, and would also remove the potential for diagnostic bias against female participants. It has been argued that girls with Autism may sometimes go undetected (Gould & Ashton-Smith, 2011) due to bias towards an existing male phenotype – recruiting general population participants with an equal number of female participants and investigating longitudinal trajectories would help to control for this possibility. Furthermore, more latent class analyses are needed to replicate and extend the studies reported here. In particular, it will be essential for authors to report in full the fit indices and model characteristics alongside a clear rationale for the model selected. It would be helpful to provide some syntax or a clear model description in supplemental information, as even small modifications to models can produce substantial changes to model fit. Finally, it will be important for authors to investigate the validity of the latent classes identified by replicating this on a number of datasets. Where possible, extension of multiple time points from childhood through into adulthood would be fruitful.

It has been noted that attempts to increase genetic homogeneity via reduced phenotypic heterogeneity have been less successful than originally hoped (Chaste et al., 2015), and it is possible that investigating longitudinal phenotypes in relation

to genetic markers may help provide more clarity about the links between genes and behaviour.

It will help for researchers to consider the three levels of chronogeneity – individual-level, group (or latent trajectory) level, and individual-level deviance from group trajectories. Studies reviewed mostly reported group-level outcomes, and where individual-level deviance was reported it was not captured statistically to any great extent. Techniques such as Mixed Membership Trajectory Models (Manrique-Vallier, 2014) can be used to examine both latent trajectory classes in data, *and* individuals' probability of changing between these trajectories over time. For example, it may be that at crucial junctures or "turning points", individuals may be able to deviate from their previous trajectory to "catapult" onto a new trajectory (Georgiades et al., 2017), and therefore close attention to individuals' movement between trajectories and possible individual or contextual predictors of this change would be extremely important for future research.

Finally, further studies are needed to investigate latent classes in RRB trajectories, as only one study was located in the present review (Richler et al., 2010). Two studies were located (Lord et al., 2015; Richler et al., 2010) that supported the separation of IS and RSM factors within RRB (Cucarro et al., 2003). These appeared to suggest that stability in IS and decline in RSM across participants can be represented by latent trajectories with around half of participants showing decreasing RSM scores and a majority increasing IS scores. Further studies are needed to clarify and replicate these findings.

4.3 Clinical recommendations

Clinicians should note that there appears to be some broad reduction in AT severity over time, which is more clearly linked to the SC than the RRB domain. However,

within overall ATs and within subdomains there appears to be variation, and distinct groups of individuals who show improvement, worsening or stability of ATs. A developmental perspective is essential when working with ASD (Bolton, 2011), and charting change over time can help improve prognostic information given to service users and their families.

Evidence has accumulated demonstrating that ATs are distributed throughout the population (Constantino, 2009), and yet in clinical practice categorical and diagnostic thinking is still the norm. A dimensional approach to Autism improves our ability to think about individual differences, to understand trajectories that worsen or improve over time, and therefore to investigate aetiology and possible treatment approaches. In particular, it has been noted that a "one-size-fits-all approach" to treatment may serve the heterogeneous ASD population especially poorly, and individually tailored approaches based on predictive data are urgently needed (Shih, Patterson, & Kasari, 2016, p.470). To this end, an understanding not only of latent trajectories but of individual and contextual predictors of individuals' ability to "jump" into improving trajectories (Georgiades et al., 2017) could help develop adaptive intervention approaches that take account of an individual's development (Almirall & Chronis-Tuscano, 2016).

It was noted in this review that the proportion of male participants reproduced the 4:1 ratio typical of "passive recruitment" studies, rather than the 3:1 ratio observed when active recruitment techniques are employed. Services reliant on continuing diagnostic practice should actively work to ensure that females with ASD are not under-represented.

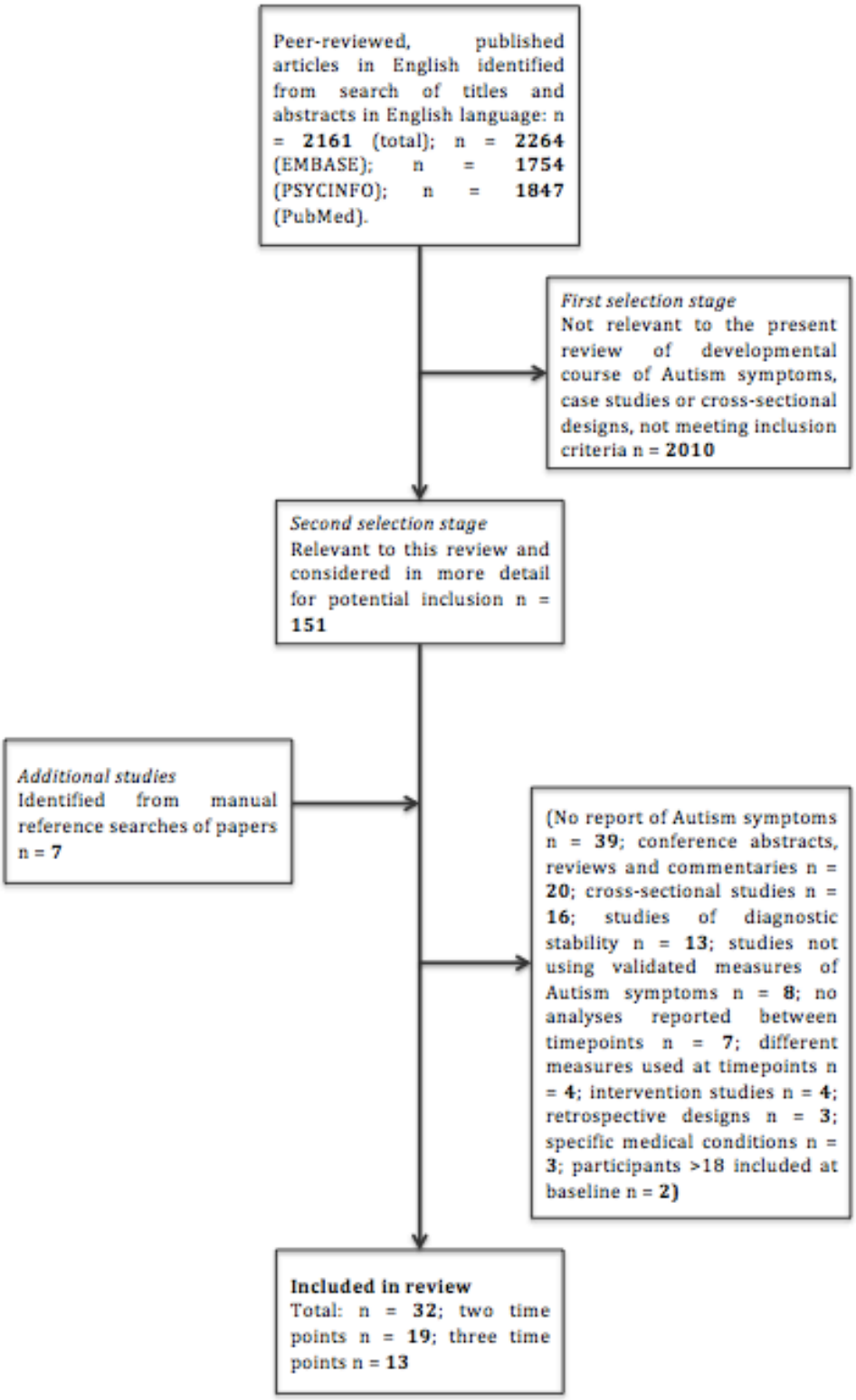


Figure 1: Flow diagram for study sele

Table 1.

Summary of two-time point studies (n = 26)

Study number	Authors	Year	N	Time points	Participants	Male (%)	Key measure	Key outcome
1	Barbaro and Dissanayake (2016)	2016	77	24m, 48m	"At risk" of ASD	76	ADOS-G Module 1 (Lord et al., 2012b) ● ●	Significant improvement in ADOS SA in both crossover and stable diagnostic groups. Change accounted for by controlling for non-verbal development. Both groups showed small improvement in RRB over time.
2	Chawarska, Klin, Paul and Volkmar (2007)	2007	31	14-25m, 15m later †	Autism, PDD-NOS, developmental delay	71	ADOS-G Module 1 (Raw scores) ● ●	Significant improvements in Communication, Social Interaction and Play and Imagination scores on ADOS in both Autism and PDD-NOS groups. No significant change in Stereotyped and Repetitive Behaviour scores.
3	Constantino et al (2009)	2009	95	8-15y, 5-6 y later *†	Twin pairs (general population); PDD clinical sample.	100	SRS ●	Modest significant improvements in mean scores over time, on the order of 0.5 SD over the five-year follow-up.
4	Darrou et al (2010)	2010	208	Median 5y, 3 y later †	PDD	80	CARS ●	Decrease in median scores between times one and two. High-level (HL) and low-level (LL) groups found using Latent Class Analysis: 72% had stable outcome, 27% from LL to HL, and 1% from HL to LL.
5	Eaves and Ho (2004)	2004	49	Mean 2y 9m, mean 4y	"Possible autism" at screening	80	CARS ●	Change in CARS scores ranged from +14 to -14 despite no mean change, and 33% changed more than five points up or down.

				11m †				
6	Estes et al (2015)	2015	308	6m, 12m	High and low family risk for Autism	60	AOSI ●	AOSI scores decreased for all groups (High Risk-ASD-High Scorers, High Risk-ASD-Moderate Scorers, High Risk-Non-spectrum and Low Risk-Non-spectrum). This was especially marked for the High Risk-Non-spectrum group.
7	Honey, McConachie, Randle, Shearer and Le Couteur (2006)	2006	89	24-48m, 13m later *†	Suspected ASD, diagnosed ASD, language disorder	80	ADI-R ● ●	Social Interaction: Small and non-significant increase for Autism and ASD groups; significant increase for "Other" diagnoses group. Communication: Small and non-significant decrease for Other group; small and non-significant increase for Autism group. RRB: Small and non-significant increases for Autism and Other groups; significant increase for ASD group.
8	Kim, Macari, Koller and Chawarska (2016)	2016	100	14-27m, 1-2y later *†	ASD	84	ADOS-G (CSS) ●	Hierarchical Clustering analysis at time one on ADOS and adaptive measures yielded four groups. Generalised Linear Mixed Models between times one and two on symptom severity yielded a significant cluster x time interaction. Stability shown in three clusters, but increase over time in one.
9	Lord et al (2006)	2006	172	2y, 9y	Referrals for possible Autism: Autism, PDD-NOS, non-spectrum	80	ADI-R, ADOS ●	Mean ADI-R score, excluding verbal items, indicated a marked reduction. Corrected ADOS scores also fell. No significant associations with sex, ethnicity, maternal education, baseline verbal or nonverbal IQs

								or adaptive behaviour.
10	Louwerse et al (2015)	2015	72	Mean 9.2y, mean 16.1y †	PDD-NOS	88	ADOS (CSS) ●	Increase in symptom severity was shown for 40% of participants, and 20% a decrease. Correlation between time points indicated broad stability.
11	McGovern and Sigman (2005)	2005	48	Mean 12y8m, mean 19y †	Autism	88	ADI-R ● ●	Significant reduction in ADI-R Social and RRB domains between mid-school and adolescent time points.
12	Messinger et al (2015)	2015	1241	24m, 36m	High-risk and low-risk for ASD	77	ADOS (CSS) ● ●	Significant interaction between sex and ADOS domains – males had significantly higher RRB scores. SA severity scores increased significantly for both ASD and Low Risk-non ASD groups, but not High Risk-non ASD. No change in RRB for any group.
13	Moore and Goodson (2003)	2003	20	Mean 2y 10m, mean 4y 5m †	Severe communication and interaction problems	80	ADI-R ● ●	Scores did not change significantly for the group as a whole in social and communication domains, but significant increase for RRB domain. Tendency for decrease in communication domain.
14	Moss, Magiati, Charman and Howlin (2008)	2008	35	Mean 3.5y, mean 10.5y	ASD	91	ADI-R ● ● ●	Mean scores for reciprocal social interaction and NVC domains significantly decreased between baseline and follow-up. Decrease in Verbal Communication (VC) did not reach significance, likely due to small *n*. No significant difference in RRB. Significant decrease in overall ADI-R scores.
15	Pellicano (2012)	2012	37	Mean 5y	ASD	89	SCQ	Significant reduction in overall symptom

				8m, mean 8y 4m †			● ● ●	scores, and in social, communication and repetitive behaviours over time. Reductions in the social domain were significantly larger than for the other domains.
16	Postorino et al (2015)	2015	60	NR	ASD	50	ADOS-G ●	No significant difference between time points and no effect of gender.
17	Soke et al (2011)	2011	28	Mean 2y 9m, mean 4y 10m †	Autism	79	ADI-R ● ● ●	Total, Social Interaction and Communication domains reduced significantly over time; no significant change in RRB.
18	Starr, Szatmari, Bryson and Zwaigenbaum (2003)	2003	58	6-8y, 2 y later *†	Autism, AS	88	ADI ● ●	Significant reduction in Communication domain – large reduction for Autism group, more stable for AS group. Significant increase in Social Interaction scores, with larger increase in AS group. No significant change in Repetitive Activities domain. Autism group significantly higher at all time points, but closing gap for Communication and Social Interaction.
19	Yirmiya, Seidman, Koren-Karie, Oppenheim and Dolev (2015)	2015	39	Mean 8y, mean 11y *†	ASD	56	ADOS (CSS) ●	Significant decrease in ADOS severity scores over time. An opposite change pattern was found – a significant decrease in boys' scores and an increase in girls' scores.

Table 2.
Summary of three-plus time point studies (n = 13)

Study no.	Authors	N	Eligible N	Time points	Participants	Male (%)	Demographic information	Comorbidities	Inclusion criteria
20	Charman et al (2005)	26	29	**3:** 2, 3, 7 y.	ICD-10 Autism or Atypical Autism	84.6%	NR	46% mental age below 18m, 6.6% epilepsy, 2.6% congenital or chromosomal abnormalities, 4.6% perinatal or prenatal condition, 48.6% self injuries.	Aged under 7; written parental informed consent; ICD-10 diagnosis of Autism or Atypical Autism
21	Joseph, Thurm, Farmer and Shumway (2013)	39	128	**3:** Mean 4, 5.4, 6.2 y	Autism: ADOS, ADI-R and clinical judgement on DSM-IV	92.3% (84.4% in n=128 group)	NR	NR	DSM-IV diagnosis of Autism.
22	Lord, Bishop and Anderson (2015)	85	213	**5:** 2, 3, 5, 9, 19 y	Best estimate consensus diagnosis of ASD: ADI-R and PL-ADOS scores.	92%	Ethnic minorities: 24%; North Carolina: 49%; Chicago: 51%.	NR	Received ASD diagnosis in early childhood and seen at age 19. Participants with profound cognitive impairment were

									excluded.
23	Ozonoff et al (2015)	418	NR	**3:** 18, 24, 36 m	Later-born biological siblings of a child with ASD (99% full siblings). 26% had ASD by 36 m.	59%	83% Caucasian, 5% Hispanic	11% Multiplex status	Older sibling has diagnosis of DSM-IV Autistic Disorder, AS or PDD-NOS; sibling and participant have no identified neurological or genetic condition that could account for ASD diagnosis (e.g., Fragile X Syndrome); <18 m old.
24	Robinson et al (2011a)	6539	7173	**3:** 7,10,13 y	ALSPAC general population cohort study participants. 86 children diagnosis of ASD, high dropout rate (*n* = 55 by T3).	50.1%	Families more likely to be homeowners; mothers more likely to have completed A levels compared to full cohort.	NR	ALSPAC guidelines.
25	Szatmari et al (2009)	57	57	**4:** 6-8, 10-14, 14-17,	AS (*n* = 21) or Autism (*n* = 36).	87.7%	NR	NR	One or more DSM-III-R criteria in

				17-19 y	Diagnosed with ADI-R algorithm. AD identified on absence of structural language impairment using Grammatic Understanding and Grammatic Completion subtests of the Test of Oral Language Development (TOLD) at 6–8 y.				social reciprocity and verbal and non-verbal communication; one or more RRB; meet criteria for 'high-functioning preschool children with ASD' based on either the Leiter Scales or the revised Stanford–Binet and using the IQ cut-offs for intellectual disability (68 and 70, respectively)
26	Clark, Barbaro and Dissanayake (2017)	48	79	**3:** 24m, 48m, 7-9y	Best estimate clinical diagnosis using ADOS, MSEL and ADI-R.	75%	48.9% mothers had completed tertiary education	NR	NR
27	Gotham, Pickles and Lord (2012)	345	NR	**2-8 per participant:** Various	Best estimate clinical diagnosis of ASD. 10 children	81.7%	78.8% white, 2% Asian American, 18.0% black, 1.2% multiracial/other;	For the 10 children without ASD: n = 5 language	ASD diagnosis, repeated ADOS administrations with best-estimate clinical

					contributing 3% of data received non-spectrum diagnoses finally.		maternal education: graduate/professional 26.5%, bachelors/some college 57.2%, high school/less 22.5%	disorders, n = 2 Intellectual Disability, n =1 Tourette's Disorder, n = 1 Mood Disorder, n = 1 Oppositional Defiant Disorder	diagnoses, Verbal and Nonverbal IQ scores, and gender and race data
28	Lord and Luyster (2006)	297	NR	**4:** 2, 3, 5, 9 y	Best estimate diagnosis: ADI-R, PL-ADOS and clinical impression for PDD-NOS or Autism.	NR	NR	NR	Inclusion: Diagnosed Autism or PDD-NOS. Exclusion: Moderate to severe sensory impairments, cerebral palsy, poorly controlled seizures.
29	Lord et al (2012a)	78	NR	**2-10+ per participant:** Various. T1 12-19m; mean 6 times over 20m.	Autism, PDD-NOS and non-spectrum participants (proportions NR).	76.9%	9.5% African-American, 2% Hispanic, 2% Asian American, 73% Caucasian, 13.5% multi-racial. Maternal education: 37%	Non-spectrum (*n* = 6), expressive language disorder (*n* = 5), general developmental delay (*n* = 3),	NR

							completed university, 37% some college, 25% high school, one mother less than high school.	non-specific Intellectual Disability (*n* = 3), Down Syndrome (*n* = 1), mild cerebral palsy (*n* = 1).	
30	Richler, Huerta, Bishop and Lord (2010)	192	214	2, 3, 5, 9 y	Autism and PDD-NOS. Consensus best estimate clinical diagnosis: clinical observation, ADI-R, ADOS and DSM.	80%	67% Caucasian, 31% African American. Maternal education: 15% graduate/professional degree, 48% college education, 26% high school, 7% less than high school.	NR	Exclusion: moderate to severe sensory impairments or cerebral palsy, known genetic abnormalities, and poorly controlled seizures
31	Szatmari et al (2015)	421	723	**3:** 2-5 yo, 6m later, 6 y	ASD: ADOS and ADI-R, confirmed clinical diagnosis using DSM-IV	84.3%	NR	NR	Aged 2 - 4 years 11 months; within 4 months received clinical diagnosis of ASD confirmed by ADOS and ADI-R; clinical diagnosis assigned by

									clinician using DSM-IV criteria
32	Venker, Ray-Subramanian, Bolt and Weismer (2014)	129	NR	**4:** 2.5, 3.5, 4.5, 5.5 y	Best estimate clinical diagnosis: ADOS and ADI-R. 91% Autism, 9 % PDD-NOS.	87%	86% Caucasian, 2 % African American, 3% Hispanic, 9% Multiracial/ other.	28% language loss	NR

Table 3.

Summary of outcomes of three-plus time point studies (n = 13)

Study no	Outcome measure	Analysis	Results	Predictors
No groups				
20	ADI-R ● ●	Paired t-tests	Significant reduction in RSI and NVC subdomains. In the RSB domain, scores increased between 3 years and 4–5 years and then decreased from 4–5 years to 7 years. Excluding the retrospective time point shows a trend for increase. Substantial heterogeneity in trajectories observed and variances increased over time.	No significant associations between NVIQ, expressive language and receptive language, and ADI-R at 3 or 7 years. Significant negative association between NVIQ measured at age 3 years and ADOS Socialisation score at age 7 years.
21	RBS-R ●	Linear mixed models	No significant differences between time points. This precluded analysis of predictors of change.	Analysis precluded by stability of scores.
Analysis with A Priori Groups				
22	ADI-R ● ●	Growth curve analysis for 3 a priori groups: ASD and VIQ<70 (n=53), ASD and VIQ≥70 (n=24), and Very Positive Outcome (VPO) (n=8).	Social communication: The intercept for the VIQ<70 group was significantly lower than for VPO and VIQ≥70. Scores for the VIQ<70 and VPO groups decreased following a quadratic pattern; scores for the VIQ≥70 group followed a linear pattern. Scores for VPO and VIQ≥70 groups diverged after age 9. RSM: The intercept was significantly higher in the VIQ<70 group than the VPO group. All scores declined gradually, following a linear pattern. IS: There was a slight increase in scores for the VIQ<70 group; no significant change for the other groups.	VIQ correlated with all except IS, but did not change slopes or intercepts
23	ADOS (social	Mixed effects linear	The TP group showed a stable, high trajectory of scores,	NR

	interaction + communication algorithm, 2002) ●	models between a priori diagnostic stability groups: True Positive (TN, n=38), True Negative (TN, n=293), False Positive (FP, n=15) and False Negative (FN, n=72) according to diagnosis at 36m.	and the TN group a stable and low trajectory. FN and FP groups began with intermediate scores, but diverged over time. The FN group demonstrated an increasing trajectory, and by 36m still showed significantly higher scores than the TN group, indicating they were sub-threshold for diagnoses but still atypical. The FP showed an initially increasing trajectory, followed by a rapid decrease in scores.	
24	SCDC ●	Growth models in 5 a priori groups: complete analytic sample (n=6,539), G8 high scorers (n=871), those above 90th percentile (n=1372), those above 95th percentile (n=687), and ASD (n=60).	SCDC scores showed little variability from ages 7 through 13 in the full sample or any of the high-scoring groups. In the full sample, SCDC scores decreased from 7-10. By 13, girls' average scores returned to their original value at baseline; boys' scores increased from 10-13 but remained slightly lower than baseline. Changes were statistically significant but small. An identical pattern was seen in each of the high-scoring groups - scores at 13 were not significantly different from baseline.	The full sample and high-scoring groups showed an inverse relationship between IQ and SCDC score. Female sex was protective, and the effect was stronger in the high-scoring groups. Higher scores on the SDQ predicted higher scores on the SCDC in each group at all measurement points. Significant interaction between behaviour problems and change in SCDC scores in all groups: The interaction resulted in a small positive predicted change in SCDC scores from ages 7 through 13 for individuals with minimal behavioural problems but small, negative predicted change for individuals with high parent-rated behavioural problems scores - this eliminated the U-shaped trend for participants with SDQ scores of 12 or above.
25	ABC	Hierarchical linear	Symptoms decreased over time in both groups at a	Non-verbal IQ and change in language scores

	•	modelling between a priori groups: Autism group (with StrLI) and AS group (without StrLI).	similar rate, with AS lower at all time points.	were added to the model, resulting in improvement in fit.
26	ADOS-G (CSS) •	ASD-Stable and ASD-Non-Stable groups were defined based on diagnostic stability. A mixed 2x3 ANOVA was conducted between groups over the 3 time points.	Significant main effects of Group, and a significant Group × Age interaction. The ASD stable group had significantly higher autism severity at each age compared to the non-stable group. Symptoms in the stable group decreased significantly at pre-school age, but increased again at school age such that there was no difference in symptom severity between toddlerhood and school age. The ASD non-stable group decreased in severity between toddlerhood and school age, with the difference in severity between these times being significantly different.	NR
Analysis with latent classes				
27	ADOS (CSS) •	Latent class growth curve models; Multinomial logistic regression to examine predictors of class membership.	4 latent trajectory classes. No evidence of an overall age trend for ADOS scores. One class n=6 had been dropped. **Groups**: Persistent high (46%), persistent moderate (38%), worsening (9%), improving (7%). A trend for improvement over time in social domains and worsening over time in RRB. Worsening class significantly greater severity over time in both domains.	Gender, class and initial NVQ did not significantly predict latent class membership. Higher NVIQ significantly predicted membership in worsening, improving or moderate rather than high group.
28	PL-ADOS, scored as Module 1 of	Latent growth curves	Four distinct profiles. One high and steady scores, one worsening (especially steep 2-5), one improving, and one low and steady	NR

	the ADOS, proxy for social communication symptoms ●			
29	ADOS ●	Generalised Linear Latent and Mixed Models	Four class model selected as the best overall. Severe persistent (21%), worsening (21%), improving (19%), non-spectrum (40%). Among worsening class, both SA and RRB increased over time. Ever ASD trajectories showed more heterogeneity than Never ASD.	The classes did not differ by gender, maternal education or referral status. There were no differences among ASD trajectory groups in hours of treatment received.
30	ADI-R ●	Growth curve analysis	For RSM subscale scores, a 3 group linear solution provided the best fit, yielding 3 groups: consistently mild (n=41), slightly decreasing (n=80) and consistently severe (n=40). For IS subscale scores, a 3 group quadratic solution provided the best fit, yielding 3 groups: mild (n=21), increasing (n=115) and moderate (n=25)..	Higher NVIQ was associated with greater improvement in RSM, but not IS.
31	ADOS (CSS) ●	Semi-parametric group-based approach (Nagin, 2005). Age at diagnosis, sex, baseline IQ, and language scores were directly included in the trajectory models as predictors.	Two groups emerged: Less severe and improving (n=48) and more severe and stable (n=373).	No one-to-one correspondence between symptom severity and adaptive functioning trajectories. Sex was the only significant predictor of group membership . Boys were more likely to be in the group with more severe symptoms and a stable trajectory than girls, who were more likely to be in the group with less severe symptoms and improving trajectory (controlling for age at diagnosis, cognitive and language scores, and site).

32	ADOS (CSS) ●	Latent Growth Curve Models. Demographic variables and experiential factors were investigated as predictors of class membership using multinomial logistic regression.	The LCGM yielded 4 classes: Persistent high (n=47), persistent moderate (n=54), worsening (n=10) and improving (n=18). At all time points, the mean CSS stable for the persistent classes. Worsening and improving classes were both characterised by means in the mild to moderate range.	Baseline non-verbal cognition was significantly lower in the Persistent High class than in the Improving class; there were no significant differences in slope of nonverbal cognition across the four classes. Baseline daily living skills were significantly lower in the Persistent High class compared to the other classes; there were no significant differences in slopes across the four classes. The Improving and Worsening classes showed significantly higher rates of growth in receptive language than the Persistent High class. The Persistent High class showed significantly slower rates of growth in expressive language than children in all other classes; children in the Persistent Moderate class also showed significantly slower expressive language growth than children in the Improving class. The Persistent High Class tended to have lower functional skills than children in the other classes, either in baseline level (intercept) or in rate of growth over time (slope).

Key:

● = total Autism scores reported

● = social communication domain scores reported

● = repetitive behaviours domain scores reported

* = overlapping ages at timepoints † = variable ages at timepoints

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*(3), 317-332.

Almirall, D., & Chronis-Tuscano, A. (2016). Adaptive interventions in child and adolescent mental health. *Journal of Clinical Child & Adolescent Psychology*, *45*(4), 383-395.Almirall & Chronis-Tuscano, 2016

Amaral, D. G., Schumann, C. M., & Nordahl, C. W. (2008). Neuroanatomy of autism. *Trends in neurosciences*, *31*(3), 137-145.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Baker, J. K., Seltzer, M. M., & Greenberg, J. S. (2011). Longitudinal effects of adaptability on behavior problems and maternal depression in families of adolescents with autism. *Journal of Family Psychology*, *25*(4), 601.

Barbaro, J., & Dissanayake, C. (2016). Diagnostic stability of autism spectrum disorder in toddlers prospectively identified in a community-based setting: Behavioural characteristics and predictors of change over time. *Autism*, 1362361316654084.

Baxter, A. J., Brugha, T. S., Erskine, H. E., Scheurer, R. W., Vos, T., & Scott, J. G. (2015). The epidemiology and global burden of autism spectrum disorders. *Psychological medicine*, *45*(3), 601-613.

Berlin, K. S., Parra, G. R., & Williams, N. A. (2013). An introduction to latent variable mixture modeling (part 2): Longitudinal latent class growth analysis and growth mixture models. *Journal of Pediatric Psychology*, *39*(2), 188-203.

Bolton, P. F. (2011). Commentary: Issues in the Classification of Pervasive Developmental Disorders. In D. Amaral, D. Geschwind, & G. Dawson (Eds.), *Autism Spectrum Disorders* (pp. 137–145). Oxford: Oxford University Press.

Bryson, S. E., Zwaigenbaum, L., McDermott, C., Rombough, V., & Brian, J. (2008). The Autism Observation Scale for Infants: scale development and reliability data. *Journal of autism and developmental disorders*, *38*(4), 731-738.

Charman, T., Taylor, E., Drew, A., Cockerill, H., Brown, J. A., & Baird, G. (2005). Outcome at 7 years of children diagnosed with autism at age 2: Predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of Child Psychology and Psychiatry*, *46*(5), 500-513.

Chaste, P., Klei, L., Sanders, S. J., Hus, V., Murtha, M. T., Lowe, J. K., ... & Geschwind, D. (2015). A genome-wide association study of autism using the Simons Simplex Collection: Does reducing phenotypic heterogeneity in autism increase genetic homogeneity?. *Biological psychiatry*, *77*(9), 775-784.

Chawarska, K., Klin, A., Paul, R., & Volkmar, F. (2007). Autism spectrum disorder in the second year: Stability and change in syndrome expression. *Journal of Child Psychology and Psychiatry*, *48*(2), 128-138.

Clark, M. L., Barbaro, J., & Dissanayake, C. (2017). Continuity and Change in Cognition and Autism Severity from Toddlerhood to School Age. *Journal of autism and developmental disorders*, *47*(2), 328-339.

Constantino, J. N. (2009). How continua converge in nature: cognition, social competence, and autistic syndromes. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*(2), 97-98.

Constantino, J. N., Abbacchi, A. M., Lavesser, P. D., Reed, H., Givens, L., Chiang, L., ... & Todd, R. D. (2009). Developmental course of autistic social impairment in males. *Development and psychopathology*, *21*(1), 127-138.

Constantino, J. N., & Gruber, C. P. (2007). Social responsiveness scale (SRS): Western Psychological Services Los Angeles.

Constantino, J. N., Przybeck, T., Friesen, D., & Todd, R. D. (2000). Reciprocal social behavior in children with and without pervasive developmental disorders. *Journal of Developmental and Behavioral Pediatrics*.

Constantino, J. N., & Todd, R. D. (2000). Genetic structure of reciprocal social behavior. *American Journal of Psychiatry*, *157*(12), 2043-2045.

Cuccaro, M. L., Shao, Y., Grubber, J., Slifer, M., Wolpert, C. M., Donnelly, S. L., ... & Pericak-Vance, M. A. (2003). Factor analysis of restricted and repetitive behaviors in autism using the Autism Diagnostic Interview-R. *Child Psychiatry & Human Development*, *34*(1), 3-17.

Darrou, C., Pry, R., Pernon, E., Michelon, C., Aussilloux, C., & Baghdadli, A. (2010). Outcome of young children with autism: does the amount of intervention influence developmental trajectories?. *Autism*, *14*(6), 663-677.

DiCicco-Bloom, E., Lord, C., Zwaigenbaum, L., Courchesne, E., Dager, S. R., Schmitz, C., ... & Young, L. J. (2006). The developmental neurobiology of autism spectrum disorder. *Journal of Neuroscience*, *26*(26), 6897-6906.

Eaves, L. C., & Ho, H. H. (2004). The very early identification of autism: Outcome to age 41/2–5. *Journal of autism and developmental disorders*, *34*(4), 367-378.

Elsabbagh, M., Divan, G., Koh, Y. J., Kim, Y. S., Kauchali, S., Marcín, C., ... & Yasamy, M. T. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Research*, *5*(3), 160-179.

Estes, A., Zwaigenbaum, L., Gu, H., John, T. S., Paterson, S., Elison, J. T., ... & Kostopoulos, P. (2015). Behavioral, cognitive, and adaptive development in infants with autism spectrum disorder in the first 2 years of life. *Journal of neurodevelopmental disorders*, *7*(1), 24.

Fountain, C., Winter, A. S., & Bearman, P. S. (2012). Six developmental trajectories characterize children with autism. *Pediatrics*, *129*(5), e1112-e1120.

Georgiades, S., Bishop, S. L., & Frazier, T. (2017). Editorial Perspective: Longitudinal research in autism–introducing the concept of ‘chronogeneity’. *Journal of Child Psychology and Psychiatry*, *58*(5), 634-636.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of autism and developmental disorders*, *39*(5), 693-705.

Gotham, K., Pickles, A., & Lord, C. (2012). Trajectories of autism severity in children using standardized ADOS scores. *Pediatrics*, *130*(5), e1278-e1284.

Gould, J., & Ashton-Smith, J. (2011). Missed diagnosis or misdiagnosis? Girls and women on the autism spectrum. *Good Autism Practice (GAP)*, *12*(1), 34-41.

Herbert, M. R., & Anderson, M. P. (2008). An Expanding Spectrum of Autism Models. In *Autism* (pp. 429-463). Humana Press.

Honey, E., McConachie, H., Randle, V., Shearer, H., & Couteur, A. S. L. (2008). One-year change in repetitive behaviours in young children with communication disorders including autism. *Journal of autism and developmental disorders*, *38*(8), 1439-1450.

Joseph, L., Thurm, A., Farmer, C., & Shumway, S. (2013). Repetitive behavior and restricted interests in young children with autism: Comparisons with controls and stability over 2 years. *Autism Research*, *6*(6), 584-595.

Kasari, C., Kaiser, A., Goods, K., Nietfeld, J., Mathy, P., Landa, R., ... & Almirall, D. (2014). Communication interventions for minimally verbal children with autism: A sequential multiple assignment randomized trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(6), 635-646.

Kim, S. H., Macari, S., Koller, J., & Chawarska, K. (2016). Examining the phenotypic heterogeneity of early autism spectrum disorder: subtypes and short-term outcomes. *Journal of Child Psychology and Psychiatry*, *57*(1), 93-102.

Krug, D. A., Arick, J. R., & Almond, P. J. (1988). Autism behavior checklist. *Austin, TX: Pro-Ed*.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 963-974.

Lam, K. S., & Aman, M. G. (2007). The Repetitive Behavior Scale-Revised: independent validation in individuals with autism spectrum disorders. *Journal of autism and developmental disorders*, *37*(5), 855-866.

Le Couteur, A., Rutter, M., Lord, C., Rios, P., Robertson, S., Holdgrafer, M., & McLennan, J. (1989). Autism diagnostic interview: a standardized investigator-based instrument. *Journal of autism and developmental disorders*, *19*(3), 363-387.

Le Couteur, A. N. N. E., Lord, C., & Rutter, M. I. C. H. A. E. L. (2003). The autism diagnostic interview-revised (ADI-R). *Los Angeles, CA: Western Psychological Services*.

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*.

Lord, C., Bishop, S., & Anderson, D. (2015). Developmental trajectories as autism phenotypes. In *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* (Vol. 169, No. 2, pp. 198-208).

Lord, C., Leventhal, B. L., & Cook, E. H. (2001). Quantifying the phenotype in autism spectrum disorders. *American Journal of Medical Genetics Part A*, *105*(1), 36-38.

Lord, C., & Luyster, R. (2006). Early diagnosis of children with autism spectrum disorders. *Clinical Neuroscience Research*, *6*(3), 189-194.

Lord, C., Luyster, R., Guthrie, W., & Pickles, A. (2012a). Patterns of developmental trajectories in toddlers with autism spectrum disorder. *Journal of consulting and clinical psychology*, *80*(3), 477.

Lord, C., Luyster, R., Gotham, K., & Guthrie, W. (2012b). Autism Diagnostic Observation Schedule–Toddler Module. *Los Angeles: Western Psychological Services*.

Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from 2 to 9 years of age. *Archives of general psychiatry*, *63*(6), 694-701.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999). Autism diagnostic observation schedule (ADOS) manual. *Los Angeles, CA: Western Psychological Services*.

Louwerse, A., Eussen, M. L. J. M., Van der Ende, J., de Nijs, P. F. A., Van Gool, A. R., Dekker, L. P., ... & Greaves-Lord, K. (2015). ASD symptom severity in adolescence of individuals diagnosed with PDD-NOS in childhood: Stability and the relation with psychiatric comorbidity and societal participation. *Journal of autism and developmental disorders*, *45*(12), 3908.

Magiati, I., Tay, X. W., & Howlin, P. (2014). Cognitive, language, social and behavioural outcomes in adults with autism spectrum disorders: a systematic review of longitudinal follow-up studies in adulthood. *Clinical Psychology Review*, *34*(1), 73-86.

Mandy, W. P., Charman, T., & Skuse, D. H. (2012). Testing the construct validity of proposed criteria for DSM-5 autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(1), 41-50.

Manrique-Vallier, D. (2014). Longitudinal mixed membership trajectory models for disability survey data. *The annals of applied statistics*, *8*(4), 2268.

McGovern, C. W., & Sigman, M. (2005). Continuity and change from early childhood to adolescence in autism. *Journal of Child Psychology and Psychiatry*, *46*(4), 401-408.

Messinger, D. S., Young, G. S., Webb, S. J., Ozonoff, S., Bryson, S. E., Carter, A., ... & Dobkins, K. (2015). Early sex differences are not autism-specific: a Baby Siblings Research Consortium (BSRC) study. *Molecular autism*, *6*(1), 32.

Moore, V., & Goodson, S. (2003). How well does early diagnosis of autism stand the test of time? Follow-up study of children assessed for autism at age 2 and development of an early diagnostic service. *Autism*, *7*(1), 47-63.

Möricke, E., Buitelaar, J. K., & Rommelse, N. N. (2016). Do we need multiple informants when assessing autistic traits? The degree of report bias on offspring, self, and spouse ratings. *Journal of autism and developmental disorders*, *46*, 164.

Moss, J., Magiati, I., Charman, T., & Howlin, P. (2008). Stability of the Autism Diagnostic Interview—Revised from pre-school to elementary school age in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(6), 1081-1091.

Muthén, B. (2004). Latent variable analysis. *The Sage handbook of quantitative methodology for the social sciences*, 345-368.

Muthén, B., & Muthén, L. K. (2000). Integrating person‐centered and variable‐centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and experimental research*, *24*(6), 882-891.

Nagin, D. (2005). *Group-based modeling of development*. Harvard University Press.

Newcomer, P. L., & Hammill, D. D. (1988). *Test of language development-2: primary*. Austin, TX: Pro-ed.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling*, *14*(4), 535-569.

Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T., ... & Zwaigenbaum, L. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: a baby siblings research consortium study. *Journal of Child Psychology and Psychiatry*, *56*(9), 988-998.

Pellicano, E. (2012). Do autistic symptoms persist across time? Evidence of substantial change in symptomatology over a 3-year period in cognitively able children with autism. *American Journal on Intellectual and developmental disabilities*, *117*(2), 156-166.

Pickles, A., & Angold, A. (2003). Natural categories or fundamental dimensions: On carving nature at the joints and the rearticulation of psychopathology. *Development and psychopathology*, *15*(3), 529-551.

Posserud, M. B., Lundervold, A. J., & Gillberg, C. (2006). Autistic features in a total population of 7–9-year-old children assessed by the ASSQ (Autism Spectrum Screening Questionnaire). *Journal of Child Psychology and Psychiatry*, *47*(2), 167-175.

Postorino, V., Fatta, L. M., De Peppo, L., Giovagnoli, G., Armando, M., Vicari, S., & Mazzone, L. (2015). Longitudinal comparison between male and female preschool children with autism spectrum disorder. *Journal of autism and developmental disorders*, *45*(7), 2046.

Richler, J., Huerta, M., Bishop, S. L., & Lord, C. (2010). Developmental trajectories of restricted and repetitive behaviors and interests in children with autism spectrum disorders. *Development and psychopathology*, *22*(1), 55-69.

Robinson, E. B., Munir, K., Munafò, M. R., Hughes, M., McCormick, M. C., & Koenen, K. C. (2011a). Stability of autistic traits in the general population: further evidence for a continuum of impairment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(4), 376-384.

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.

Robinson, E. B., St Pourcain, B., Anttila, V., Kosmicki, J. A., Bulik-Sullivan, B., Grove, J., ... & Martin, J. (2016). Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nature genetics*, *48*(5), 552-555.

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*. Western Psychological Services.

Schopler, E., Reichler, R. J., & Renner, B. R. (2002). *The childhood autism rating scale (CARS)*. Los Angeles, CA: Western Psychological Services.

Seltzer, M. M., Shattuck, P., Abbeduto, L., & Greenberg, J. S. (2004). Trajectory of development in adolescents and adults with autism. *Developmental Disabilities Research Reviews*, *10*(4), 234-247.

Shih, W., Patterson, S. Y., & Kasari, C. (2016). Developing an adaptive treatment strategy for peer-related social skills for children with autism spectrum disorders. *Journal of Clinical Child & Adolescent Psychology*, *45*(4), 469-479.

Skuse, D. H., Mandy, W. P., & Scourfield, J. (2005). Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. *The British Journal of Psychiatry*, *187*(6), 568-572.

Skuse, D. H., Mandy, W., Steer, C., Miller, L. L., Goodman, R., Lawrence, K., ... & Golding, J. (2009). Social communication competence and functional adaptation in a general population of children: preliminary evidence for sex-by-verbal IQ differential risk. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(2), 128-137.

Soke, G. N., Philofsky, A., Diguiseppi, C., Lezotte, D., Rogers, S., & Hepburn, S. (2011). Longitudinal changes in scores on the Autism Diagnostic Interview—Revised (ADI-R) in pre-school children with autism: Implications for diagnostic classification and symptom stability. *Autism*, *15*(5), 545-562.

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2006). Vineland Adaptive Behavior Scales – Second Edition. Minneapolis, MN: NCS Pearson Inc.

Spiker, D., Lotspeich, L. J., Dimiceli, S., Myers, R. M., & Risch, N. (2002). Behavioral phenotypic variation in autism multiplex families: evidence for a continuous severity gradient. *American Journal of Medical Genetics Part A*, *114*(2), 129-136.

St. Pourcain, B., Wang, K., Glessner, J. T., Golding, J., Steer, C., Ring, S. M., ... & Davey Smith, G. (2010). Association between a high-risk autism locus on 5p14 and social communication spectrum phenotypes in the general population. *American Journal of Psychiatry*, *167*(11), 1364-1372.

Starr, E., Szatmari, P., Bryson, S., & Zwaigenbaum, L. (2003). Stability and change among high-functioning children with pervasive developmental disorders: A 2-year outcome study. *Journal of Autism and Developmental Disorders*, *33*(1), 15-22.

Szatmari, P., Bryson, S., Duku, E., Vaccarella, L., Zwaigenbaum, L., Bennett, T., & Boyle, M. H. (2009). Similar developmental trajectories in autism and Asperger syndrome:

from early childhood to adolescence. *Journal of Child Psychology and Psychiatry*, *50*(12), 1459-1467.

Szatmari, P., Georgiades, S., Duku, E., Bennett, T. A., Bryson, S., Fombonne, E., ... & Volden, J. (2015). Developmental trajectories of symptom severity and adaptive functioning in an inception cohort of preschool children with autism spectrum disorder. *JAMA psychiatry*, *72*(3), 276-283.

Venker, C. E., Ray-Subramanian, C. E., Bolt, D. M., & Weismer, S. E. (2014). Trajectories of autism severity in early childhood. *Journal of autism and developmental disorders*, *44*(3), 546.

Woolfenden, S., Sarkozy, V., Ridley, G., & Williams, K. (2012). A systematic review of the diagnostic stability of autism spectrum disorder. *Research in Autism Spectrum Disorders*, *6*(1), 345-354.

Yirmiya, N., Seidman, I., Koren-Karie, N., Oppenheim, D., & Dolev, S. (2015). Stability and change in resolution of diagnosis among parents of children with autism spectrum disorder: Child and parental contributions. *Development and psychopathology*, *27*(4pt1), 1045-1057.

Zimmerman, A. W. (Ed.). (2008). *Autism: Current theories and evidence*. Springer Science & Business Media.

Part 2. Empirical Paper: Latent Trajectories of Autistic Traits in the General Population

Abstract

Aim

Research has shown that Latent Class Growth Modelling (LCGM) techniques can be applied to discover heterogeneous trajectories of autistic traits (ATs) within diagnosed samples. However, no studies have yet applied LCGM to general population participants. This study aimed to use LCGM to examine latent trajectories of ATs in a large, general population cohort.

Methods

Growth Mixture Modelling (GMM) was applied to Social and Communication Difficulties Checklist (Skuse et al, 2005) scores at ages 7, 10, 13 and 16. Data were from 9,744 participants in the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort. Sex, IQ and general difficulties were investigated in relation to latent classes.

Results

Six latent classes of ATs over time were observed, characterised by persistent high, persistent low, early increasing, later increasing, high decreasing and moderate decreasing trajectories. Females were overrepresented in the later developing trajectory. General difficulties followed similar growth patterns to ATs within groups.

Conclusions

A six-class model characterised development of ATs. Evidence was found to support the observation that girls display symptoms later than boys and are likely to be under-diagnosed. Further characterising latent trajectory classes of ATs in the general population could improve understanding of aetiology, genetics and treatment.

1. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterised by difficulties in social communication and interaction, together with a restricted focus on topics of interest and repetitive behaviours (APA, 2013). Prevalence in the UK has been estimated at 157 per 10,000 (Baron-Cohen et al., 2009), and is associated with a lifelong impact on health, wellbeing, social integration and quality of life for individuals and their families (Knapp, Romeo, & Beecham, 2009). However, ASD is a complex and heterogeneous neurodevelopmental disorder with varied neurobiological pathways and presentations (Amaral, Schumann, & Nordahl, 2008; DiCicco-Bloom et al., 2006; Herbert & Anderson, 2008; Zimmerman, 2008) and the phenotype encompasses a range of features beyond the core symptoms, including cognitive, behavioural, emotional, motor and sensory aspects (Mandy et al., 2012; Volkmar & Klin, 2005).

Research has indicated that Autism is not a discrete disorder, but rather that ASDs represent the upper extreme of a constellation of deficits in social and communicative behavior that are continuously distributed in nature (Constantino, Przybeck, Friesen, & Todd, 2000; Constantino & Todd, 2000; Posserud, Lundervold, & Gillberg, 2006; Robinson et al., 2011a; Robinson et al., 2011b; Skuse et al., 2009; Spiker, Lotspeich, Dimiceli, Myers, & Risch, 2002). Large independent studies across cultures and rating systems have confirmed that it is arbitrary where a cut-point is drawn between the respective conditions of being affected versus unaffected by ASD (Constantino, 2009). Indeed, autistic traits (ATs) have been shown to be associated with variance in genetic risk across the clinical spectrum (Lundström et al., 2012; Robinson et al., 2011; Robinson et al., 2016; St Pourcain et al., 2014), and an association has been found across the full range of AT severity with emotional, behavioural, social and functional difficulties (Lundström et al., 2011; Oliver, Berg, Moss, Arron, & Burbidge, 2011; Mandy et al., 2013).

A paradigm shift towards researching ATs rather than categorical diagnoses has profound implications. Firstly, studying ATs enhances statistical power and exploration of neural mechanisms underlying behaviour (Constantino, 2009). Even subclinical traits are meaningful in terms of conferring risk; when employing a categorical approach, mild and moderate deficits in social and communicative competence are therefore likely to be missed, especially if in the context of comorbid psychiatric disorders such as Conduct Disorder (CD; Gilmour et al., 2004) and Attention-Deficit / Hyperactivity Disorder (ADHD; Banaschewski et al., 2005). Furthermore, longitudinal studies of ATs can help to explore "chronogeneity" (Georgiades, Bishop, & Frazier, 2017), referring to heterogeneity over time. Understanding the different trajectories individuals take over time in development of ATs can help characterise differing phenotypes, aetiologies, outcomes and to adapt treatments for individuals' evolving status (Alimirall & Chronis-Tuscano, 2016; Kasari et al., 2014; Magiati, Tay, & Howlin, 2014; McGovern & Sigman, 2005; Seltzer et al., 2004).

Few studies have examined developmental trajectories of ATs, and only one study has examined traits in the general population rather than participants with ASD diagnoses (Robinson et al., 2011b). Robinson et al (2011b) examined the developmental course of ATs measured by the Social and Communication Disorders Checklist (SCDC; Skuse et al., 2005) in 6,539 population participants drawn from the Avon Longitudinal Study of Parents and Children (ALSPAC) at ages 7, 10 and 13. The study showed high levels of stability of SCDC scores in the whole sample and amongst high-scoring participants. Scores for males improved a small but significant amount over the course of the study, and for all participants a pattern was found in which scores improved between 7 and 10 years of age, returning to baseline at 13. While the study was useful in charting this development, the methodology employed implicitly assumes that all participants represent a relatively

homogeneous population that can be represented by a single growth trajectory. It is crucial that research captures and measures the heterogeneity of Autism (Ozonoff et al., 2015), and a single group trajectory approach oversimplifies the complex growth patterns that describe continuity and change in heterogeneous populations (Jung & Wickrama, 2008).

Latent class growth models (LCGMs; Muthén & Muthén, 2000) represent a constellation of techniques that do not require the assumption that a single growth curve over time can adequately represent change across an entire population. Instead, they take into account unobserved heterogeneity within the larger population, defining latent classes that reflect different groups with distinct trajectories. LCGM techniques have been applied in six prior studies examining ATs across time, albeit to date these have focused on participants with ASD diagnoses, reflecting the extreme upper end of the AT continuum. Four of these studies discovered that four latent classes best captured the different trajectories of ATs over time: a consistently high trajectory, a consistently low or moderate trajectory, an increasing (worsening) trajectory and a decreasing (improving) trajectory (Gotham et al., 2012; Lord & Luyster, 2006; Lord et al., 2012; Venker et al., 2014). Another study reported a two-class model with a mild and improving trajectory and a stable and severe trajectory (Szatmari et al., 2015), however further classes should have been added as this actually improved the fit of the model. Finally, Fountain et al (2012) reported a six-class model of change in ATs, with stable trajectories at varying degrees of severity and one improving class; however, this used a non-validated measure of ATs and therefore should be interpreted cautiously. Lower verbal IQ and more impaired daily living skills (Gotham et al., 2012) and male gender (Szatmari et al., 2015) were found to predict membership in the stable and severe trajectory classes, although two studies found that sex did not significantly predict class membership (Gotham et al., 2012; Lord et al., 2012).

While these papers are useful in highlighting the way ATs develop over time, and the consistent finding of four similar trajectories was particularly noteworthy, they also suffer from some methodological limitations. The restriction to participants with ASD diagnoses means that by definition any trajectory of increasing ATs that start at a mild or moderate level but become more severe over time cannot be captured. Furthermore, any understanding of how ATs develop and change over time across the spectrum of severity is excluded. In particular, females with high ASD traits are likely to be underrepresented, as they are less likely than equivalent males to receive a diagnosis (Dworzynski et al., 2012; Loomes, Hull, & Mandy, 2017; Russell, Steer, & Golding, 2011).

Studies of development of AT traits over time in females can help to disprove or support one of two competing hypotheses about female ASD development. A "female compensation hypothesis" has suggested that females compensate more effectively than males over time, learning social skills and increasingly adapting to the social world, such that by adulthood their traits are more attenuated (Lai, Lombardo, Auyeung, Chakrabarti, & Baron-Cohen, 2015); this would predict a female phenotype to be associated with a *decline* in ATs over time. By contrast, the "adolescent emergence hypothesis" suggests that females' social and communication difficulties only become overt in adolescence, when navigating the social world becomes more a more complex task (Bargiela, Steward, & Mandy, 2016; Kopp & Gillberg, 2011). As Hsiao, Tseng, Huang and Gau argue, "the impact of [ASD] traits on social functioning may not present until pressure from demands for new social adaptation arises across major developmental transition" (2013, p.256). This hypothesis would predict a female ASD phenotype is associated with an *increase* in ATs over time.

In summary, LCGM approaches have usefully been applied to ATs, to highlight differing groups with different trajectories over time. However, to date no studies have yet applied LCGM to ATs in the general population. This study aims to extend the work of Robinson et al (2011b) by using Growth Mixture Modelling (GMM), a LCGM technique, to examine latent trajectories of SCDC scores from the ALSPAC data set across four time points (ages 7, 10, 13 and 16). This will help to clarify whether the previously observed stable profile across participants masks sub-populations with heterogeneous AT trajectories over time. The use of general population participants will allow for the possibility of increasing trajectories to be observed, should any participants have low AT scores at baseline but increase to severe ATs over time. It will also allow for investigation into whether female participants are over- or under- represented in any particular trajectory group, and thus lending support to either adolescent emergence or female compensation hypotheses. Finally, investigating group membership relative to both IQ and psychological strengths and difficulties in peer relationships, conduct, hyperactivity/inattention and pro-social behaviour using the Strengths and Difficulties Questionnaire (Goodman, 2001) will help to test the validity of the observed groups, as well as highlight any important relations between SCDC trajectories and these variables.

Research questions / hypotheses

- Are developmental trajectories of ATs explained by latent trajectory classes?
- Are there trajectories that show an over-representation of females versus males?
- Are trajectories differentially related to IQ and SDQ scores?

2. Method

2.1 Setting

Data analysis for this study was carried out at UCL, using data collected by the Avon Longitudinal Study of Parents and Children (ALSPAC). ALSPAC data were collected in the former administrative county of Avon, comprising Health Districts of Southmead, Frenchay and Bristol and Weston. The catchment area covered by these Health Districts has now become the "Bristol and District Health Authority," and includes the City of Bristol as well as surrounding urban and rural areas including towns, villages and farming communities. This area had a population circa 0.9 million in 1991 (Boyd et al., 2012). All pregnant women resident in these health districts were eligible to participate if their estimated delivery date fell between 1st April 1991 and 31st December 1992 inclusive. Any child born of these pregnancies was considered eligible.

2.2 Participants

2.2.1 Recruitment

Recruitment to ALSPAC was carried out in three phases. Phase I was opportunistic and made contact with eligible women throughout the 1990-1992 period via media information and visits to community locations by recruitment staff. The majority (82.6%) of women recruited enrolled during this phase. Phase II was carried out when eligible children were 7 years old. Children who would have fitted the original eligibility criteria, excluding those who refused enrolment, were systematically identified and invited to participate. During Phase III, further opportunistic contact was made with eligible families via community outreach and promotion activities. Data collected during Phases II and III lack the data on pregnancy, infancy and early childhood collected during Phase I (Boyd et al., 2012).

2.2.2 Sample size

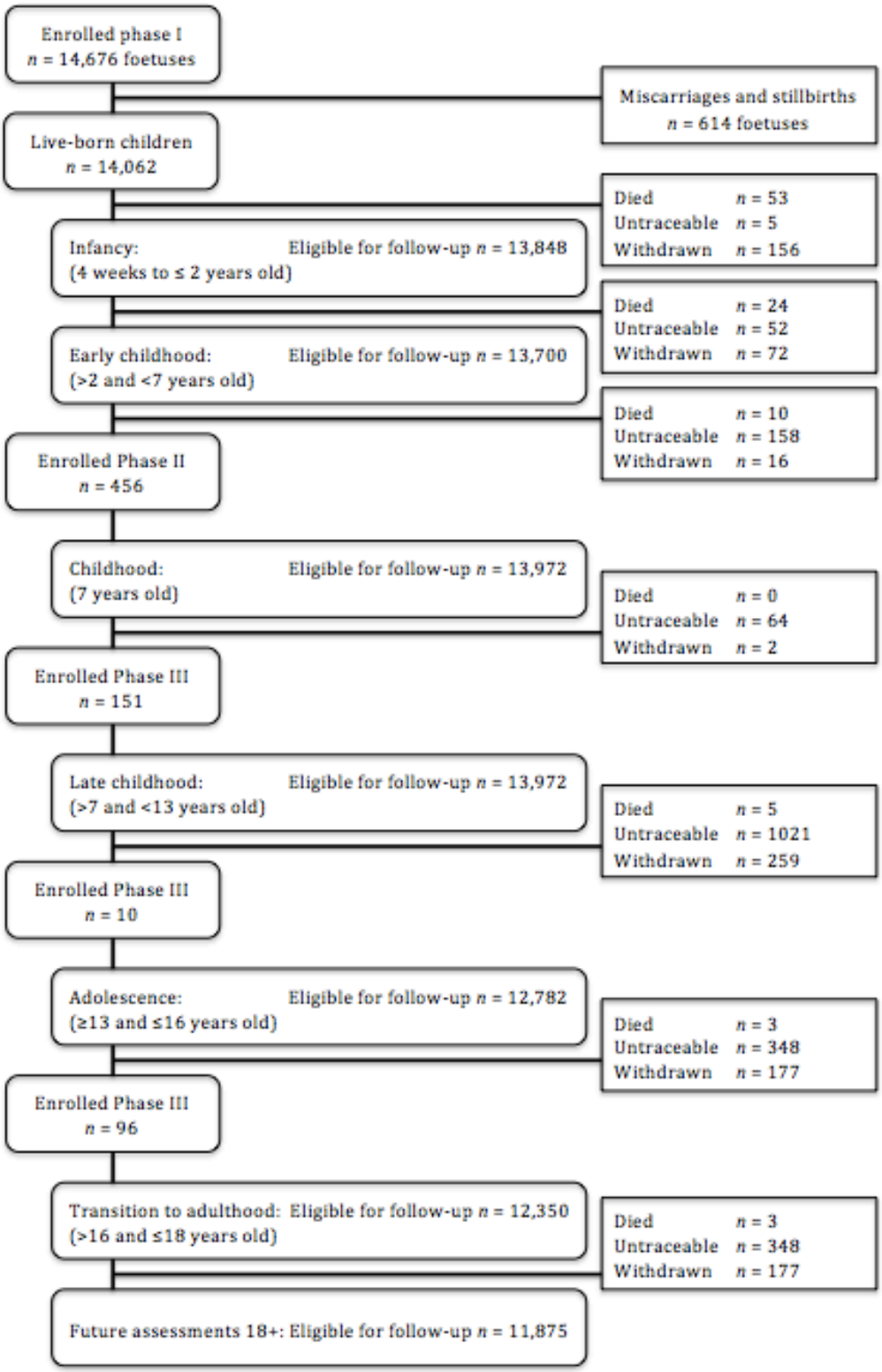


Figure 1. Study attrition flow diagram

As Figure 1 shows, 14,062 live-born children entered the study. Throughout the progression of the study, 1,498 participants were lost to follow-up, resulting in a total of 12,350 remaining in the study at the age of 18. The average response rate during the "adolescence" phase was 48.2%, and 75% responded at least once (Bray, 2012). Of the data available, 9,744 had at least one recorded SCDC score and so were eligible for inclusion in the present study.

2.2.3 Demographic characteristics

The recruited ALSPAC sample were 49.69% female, 96.09% White and 6.22% had a joint parental income of less than or equal to £16,000 per annum. This differs significantly from comparable national demographics, with participants more likely to be White (OR 3.85, 3.50 - 4.24) and less likely to be of lower socioeconomic status (OR 0.46, 0.43 - 0.50; Boyd et al., 2012). Demographic information for participants whose data was used in the study (*n* = 9,744) is shown in Appendix 1. Participants in this study were 49.1% female, 98% White, 15.2% had a mother with a University degree, 79.5% had parents who were homeowners and 40.1% had parents in professional, managerial or technical occupations. Odds of missing all SCDC data and therefore being excluded from the present study were significantly higher in participants whose mothers had less than A-Level education (Odds Ratio [OR] = 2.36, 95% CI [2.15, 2.58]), and for BME participants (OR = 2.22, 95% CI [1.78, 2.77]).

2.3 Power calculation

In GMM there is no general rule regarding how large a sample is necessary (Ram & Grimm, 1987). Identification and description of latent groups depends on the extent of between-group differences, homogeneity of the change process, relative group

sizes, reliability of measurement and other factors, and hence in GMM as in power calculations in growth modelling more generally, each application must be evaluated in relation to the particulars of the data and model (Muthén & Muthén, 2002; Ram & Grimm, 2009). A power calculation can be performed however to detect that a parameter is different from zero. A mean and covariance matrix were determined using residuals, growth factor means and variances generated from a growth curve model of the whole study population. This matrix was then analysed whilst the mean of the slope growth factor was fixed to zero (a misspecification), with 9,744 observations, to determine a chi-square as an approximate non-centrality parameter. This was then used to calculate the power to detect this chi-square value using with an α value of .05 using SAS University Edition Software for Mac OS X (SAS, 2017). This revealed that the sample size had 90% power to detect that the slope value was different from zero.

2.4 Measures

Measures are reproduced in full in Appendix 3.

Social and Communication Disorders Checklist (SCDC; Skuse et al., 2005)

The SCDC was administered with ALPSAC participants at ages 7, 10, 13 and 16. Its questions are designed to assess social and communication skills over the past six months. Total scores range from 0 to 24, where 24 indicates the highest level of ATs. The SCDC displays excellent internal consistency (α = .93) and strong discriminant validity in the ALSPAC sample (Skuse et al., 2005). Test-retest reliability was good (ICC = .81). The SCDC showed sensitivity of .9 and specificity of .69. The SCDC does not assess restricted and repetitive behaviours and interests, another definitional domain of ASD.

Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001)

The SDQ was administered with ALSPAC participants at ages 9, 11, 13 and 16. It is part of the Development and Well-Being Assessment (DAWBA), a package of interviews, questionnaires and rating techniques designed to generate psychiatric diagnoses for children aged five to seventeen (Goodman, Ford, Richards, Gatward, & Meltzer, 2000). The SDQ's questions are designed to assess psychological attributes over the past six months both positive and negative, across subscales of emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and pro-social behaviour. An overall difficulties score ranges from 0-40, and the first four subscales range from 0-10. The prosocial behaviour subscale yields a score from 0-4. The SDQ has shown acceptable internal consistency for the difficulties score ((α = .77; Mieloo et al., 2012) and acceptable concurrent and divergent validity (Mieloo et al., 2012).

Wechsler Intelligence Scale for Children III (WISC; Wechsler, Golombok, & Rust, 1992)

The WISC was administered with ALSPAC participants at age 8 in an abbreviated form that included randomly selected items from each of the 10 subtests. The WISC is a commonly used general measure of intelligence in children aged six to sixteen. It is comprised of three IQ scores – full-scale, verbal and performance. Each of the IQ scores has a mean of 100 and standard deviation (SD) of 15. Split-half reliability coefficients are .96 for full-scale IQ, .95 for verbal IQ and .91 for performance IQ (Kaufman & Lichtenberger, 2000). It was standardized on a representative sample of US children (n=2,200; Wechsler et al., 1992).

2.5 Ethics

Data from the ALSPAC cohort do not contain any identifying information. Ethical oversight of all ALSPAC data collection has been provided by the ALSPAC Ethics and Law Committee (ALEC), comprised of clinicians, researchers, lawyers and study participants. Initial ethical approval was granted by the Bristol and Weston Health Authority, Southmead Health Authority and Frenchay Health Authority.

Permission to access the ALSPAC data for this study was approved by the ALSPAC Executive Committee. As part of this process, the researcher agreed to abide by the security policy for the ALSPAC data, to only share data with those working on the project, and to securely destroy any datasets upon completion. Approval will be sought by the ALSPAC Study Executive before submitting any manuscripts for publication.

2.6 Data analysis

Data were analysed using GMM in MPlus 8.0 Plus Mixture Add-On for Mac OS X (Muthén & Muthén, 2017a). Data analyses were carried out following the guidelines for conducting and reporting LCGMs provided by Ram and Grimm (2009). Firstly, a baseline growth model was obtained (a single trajectory for all participants), fitting linear and quadratic versions of the model to obtain the best single representation of change in SCDC scores at 7, 10, 13 and 16 years. Overall goodness of fit at this stage was assessed using Structural Equation Modelling Fit Indices including the Comparative Fit Index (CFI; Bentler, 1990), the Root Mean Square Error of Approximation (RMSEA; Steiger, 1990) and the Standardized Root Mean Square Residual (SRMR; Byrne, 1998). Cut-offs of greater than .95 on the CFI and below

.06 on the RMSEA and .08 on the SRMR have been deemed indicative of well-fitting models (Hu & Bentler, 1999).

Next, a Latent Class Growth Analysis (LCGA) was carried out. LCGA assumes no within-class variance for growth factors, meaning that slopes, intercepts and quadratic terms are identical across individuals within classes. It is recommended as a first model-building step prior to GMM due to the clearer identification of classes and reduced computational burden (Jung & Wickrama, 2008). Additional classes were added to the LCGA, obtaining parameter estimates and fit statistics for all models. Custom starts and iterations were set for all analyses to avoid local maxima. To model linear change, slope loadings were fixed to represent time following the zero-point (7 years), with factor loadings of times two to four fixed at 1, 2 and 3 respectively (see Appendix 2 for LCGA and GMM example syntax).

Next, GMM was carried out. In this approach, the assumption of zero within-class variance is relaxed so that variances are obtained around intercept, slope and quadratic growth factors. By default in GMM, variances are constrained to be equal across classes; this can be relaxed to allow different variances to be obtained in each class. Residuals (error) can be constrained to be equal across classes, or estimated for each class. Hence prior to running the models, three basic GMM specifications were designed: one fully-variant (variances and residuals allowed to vary across classes), one fully-invariant (variances and residuals held equal across classes), and one mixed (variances allowed to vary, but residuals fixed; see Appendix 2). GMM also allows time points to be “freed” to create a Latent Basis Model; this was deemed inappropriate for the SCDC data due to the fact the measurements were taken at specific and fixed time points. The three basic GMM specifications were compared for model fit and viability up to seven classes, and customised accordingly.

GMM does not assume a normal distribution in the modelled data, and only normal distributions within the latent classes that it discovers (Muthén & Muthén, 2000). Nonetheless, some specialised modifications of GMM can be used to explicitly account for highly non-normal data if required. A “skewnormal” or “skewt” modelling command can be used to specify skewed data distributions, but these are not appropriate if a large proportion of zeroes are present (floor effects; Muthén & Asparouhov, 2015). When floor effects are present in the data, a two-part GMM can be specified (Olsen & Schafer, 2001; Muthén, 2001). This separates the data into two parts, each modelled by separate growth functions. The first part models binary outcomes (representing whether zero or non-zero scores are present), while the second part uses a standard GMM on the continuous outcomes of participants with non-zero scores. Additionally, the second part of the two-part model can be logarithmically transformed, to further approach normality within the data.

Count data can also be modelled using GMM, and one of the specific applications of this in the presence of floor effects is a Zero-Inflated Poisson GMM. Like the two-part GMM, in this approach two models are specified, one describing the growth model for the count part of the outcome for individuals for whom we are able to assume values greater than zero, and one specifying a growth model for the inflation part of the outcome – the probability of being unable to assume any value except zero. Models are fit separately for the two components, and then combined into a single model.

2.7 Model selection

Use of a combination of substantive theory and statistical criteria has been recommended for selecting the optimal number of classes in GMM (Muthén, 2003).

Model selection criteria can generally be grouped into four categories: information based criteria (IC), nested likelihood ratio tests (LRTs), goodness-of-fit measures and classification-based statistics (Henson, Reise, & Kim, 2007; Tofighi & Enders, 2008; Vermunt & Magidson, 2002). Of these, IC and LRTs are the most useful for selecting an optimal number of classes (Chen et al., 2017). IC indices are based on the log-likelihood value of a fitted model, and also penalise based on numbers of parameters and/or sample-size. Most prominent of the ICs are the Akaike IC (AIC; Akaike, 1987), Bayesian IC (BIC) and sample-size adjusted BIC (SABIC). Lower values on these indices reflect better-fitting models (Muthén, 2003; Nylund, Asparouhov, & Muthén, 2007). The general principle is to select the model with the lowest absolute value on the chosen IC (Ram & Grimm, 2009). LRTs include the adjusted Lo–Mendell–Rubin likelihood ratio test (ALMR; Lo, Mendell, & Rubin, 2001), and the bootstrap likelihood ratio test (BLRT; McLachlan & Peel, 2000). LRTs compare the ratio of log-likelihoods for the restricted model with *k*-1 classes to the less restricted model with *k* classes (Agresti, 1996). Significance tests are reported with LRTs, whereby $p>.05$ suggests that the *k* class model is not a significant improvement over the *k*-1 class model. Of these statistics, the BIC and the BLRT have been found to perform the best in simulation studies (Nylund et al., 2007; Peugh & Fan, 2012), while the AIC and ALMR have been shown to be relatively less reliable – in particular, the ALMR tends to cycle between significant and non-significant findings, and is cautioned against with large sample sizes (Nylund et al., 2007; Chen et al., 2017). Of the classification-based statistics, Entropy is perhaps the most prominently used (Ram & Grimm, 2009) and is a summary measure of the conditional probabilities of individuals' group membership (Jedidi, Ramaswamy, & Desarbo, 1993). High Entropy values (>.80) indicate good confidence of classification (that individually really should belong to their class), and that there is good separation between latent classes (Greenbaum, Del Boca, Darkes, Wang, & Goldman, 2005; Muthén, 2004).

2.8 Missing data handling

Missing data was handled in MPlus using Maximum Likelihood estimation with robust standard errors under the assumption of data being Missing At Random (MAR; Little & Rubin, 2014). Data are said to be MAR when the probability of missing data is not related to the missing data itself. A special case of MAR is Missing Completely at Random (MCAR; Little & Rubin, 2014), which imposes the further restriction that the probability of missing data is also not related to any observed data. Data are considered Missing Not at Random (MNAR) or to be non-ignorable (Schafer & Graham, 2002), when the probability of missing data depends on the observed variable itself.

Data were investigated to assess whether the missing data mechanism could be assumed to be MAR, MCAR or NMAR using SPSS 24 for Mac. First, Little's MCAR test was employed. This is a chi-square test of the null hypothesis that data are MCAR, which evaluates mean differences across subgroups of missing cases that share the same missing data pattern (Little, 1988). In order to decide whether data could be considered MAR as opposed to MNAR, the pattern of missing data was first investigated in relation to the construct under investigation by testing the non-parametric correlation between mean SCDC scores and the number of missing data points (one to three). Further, additional observed variables of sex, maternal education and ethnicity were compared with both SCDC scores and missing data. A connection between an observed variable and both SCDC scores and missing data does not exclude considering data to be MAR, however it should be noted (Schafer & Graham, 2002). This finding would however exclude data from being considered to be MCAR.

3. Results

3.1 Descriptive statistics

Distributions for the SCDC scores at all four time points (age 7, age 10, age 13 and age 16) were examined using SPSS Version 24 for Mac (see Appendix 1 for descriptive statistics). Data were heavily skewed and kurtotic, with skewness ranging from 1.90 (SE = .03) at time 4 to 2.46 (SE = .03) at time 2, and kurtosis ranging from 4.08 (SE = .07) to 7.51 (SE = .06) at time 2. Kolmogorov-Smirnov tests for normality were highly significant (p<.001). A floor effect in the data was observed, with proportions of zeroes ranging from 34.2% at time 1 to 42.4% at time 2.

Mean scores compared closely to the means previously observed in typically developing children and adolescents (Skuse et al., 2005; Appendix 1). However, male mean scores were observed to decrease significantly by age 16 (M = 2.75, SD = 3.79; t = 3.41, p = .001), while female scores increased significantly over the same time frame (M = 2.90, SD = 3.77; t = 7.45, p<.001), such that differences between male and female scores were no longer significant by age 16 (t = 0.95, p = .35). This pattern of change has not previously been observed in this dataset (Robinson et al., 2011) due to the unavailability of the 16 years time point.

3.2 Missing data

Table 1.

Number of missing data points

Data points	N	%
0	4380	28.4
1	2005	13.0
2	1639	10.6
3	1720	11.1
4	5701	36.9

As Table 1 shows, 5,701 participants were missing all four data points, and these were dropped from further analyses. Twenty-eight percent of participants (*n* = 4,380) had complete data. Proportions of missing data increased across each time point, with attrition ranging from 16.7% at age 7 to 41.7% by age 16 (Table 2).

Table 2.

Missing data at each time point

Time	N	%	Total
1	1,629	16.7	8115
2	2,017	20.7	7719
3	2,734	28.1	7006
4	4,063	41.7	5674

Little's MCAR test was found to be highly significant and so the null hypothesis that data were MCAR was rejected (χ^2 = 119.70, d.f = 28, p<.001).

Mean SCDC scores were compared with missing data, to investigate whether missing data were systematically related to the construct under investigation. Non-parametric correlations were carried out between mean SCDC score per participant,

and number of missing data points (maximum 3), which were found to be non-significant (Kendall's τ <.001, p = .99).

Odds of having any missing data were higher for male participants (OR = 1.12, 95% CI [1.04, 1.20]). Given the observation that male participants generally score more highly than females (Skuse et al., 2005), this could introduce bias. For example, the increasing mean scores of females across time points could be explained by male attrition from the study. To investigate this, data were coded into a binary variable using the time-invariant, Receiver Operating Curve (ROC)-maximising value of 8 (Robinson et al., 2011) to determine whether high-scorers at 13 were differentially related by sex to having missing data at age 16. It was observed that 36.17% of high-scoring males at age 13 were missing data at 16, compared to 36.16% of high-scoring females (χ^2 = .01, d.f = 1, p = .94). Hence it did not appear to be the case that male missingness was systematically related to higher SCDC scores.

Participants with less highly educated mothers and from BME backgrounds were more likely to have missing data (see *Methods*). Mean SCDC scores were not significantly different by ethnicity (t = -0.06, d.f = 8987, p = .95), however mean scores for children whose mothers had less education were significantly higher (M = 2.88 SD = 3.53) than for those with A-Levels or degrees (M = 2.54, SD = 3.05; t = 4.77, d.f = 8942.07, p<.001). Therefore missing data was shown to be related to gender, maternal education and ethnicity, and maternal education was related to both missing data and SCDC scores. There was no evidence that overall SCDC scores were related to probability of missing data. Hence while an MCAR missing data mechanism was not supported, the data were deemed to be appropriate for MAR.

3.3 Growth curve analysis

An initial model included intercept and slope growth factors, and model fit was compared once a quadratic term was added.

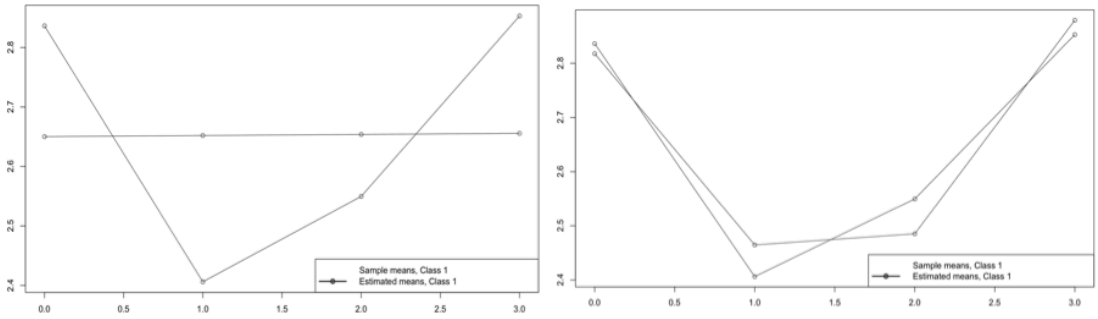


Figure 2. Linear and quadratic growth curve models

As Figure 2 shows, the quadratic growth model is a better fit to the data than the linear model. Growth factor means are shown in Table 3.

Table 3.

Growth factor means for quadratic single-class growth model

Growth Factor	Mean	S.E	Mean / S.E	Sig.
Intercept	2.86	.04	71.59	<.001
Slope	-0.55	.04	-12.91	<.001
Quadratic	0.21	.01	14.48	<.001

An improvement in model fit was observed, with the BIC value decreasing from 155698.49 to 155635.87. Fit improved from mediocre to good according to the RMSEA (.81 to .48), and from good to excellent according to the SRMR (.039 to .009) and CFI (.968 to .998).

3.4 Latent Class Growth Analysis

Fitting the single growth curve model revealed that a quadratic model fit the data well. Next, a quadratic LCGA was specified (see Appendix 2 for syntax). Classes were added up to a seven-class model, and fit indices and class compositions are shown in Appendix 4. Entropy statistics were high (.85 to .94), suggesting good fit. BIC values continued to decrease (improve) as additional classes were added. For exploratory purposes, classes were added beyond the seven-class model and BIC did not reach a minimum until a twelve-class model was reached – it was also at this point that BLRT became non-significant (indicating no improvement from adding further classes). AIC and SSABIC also continued to decrease until this point. By contrast, the ALMR became non-significant from the four-class model onward, indicating that the three-class model was the optimal model according to this index. However, due to the contrast with the rest of the fit indices and the relatively superior performance of the BIC and BLRT (see Methods) this was not deemed sufficient basis for model selection.

3.5 Growth Mixture Model

Three GMMs were specified prior to analyses: fully-variant, fully-invariant and mixed (see Methods for details and Appendix 2 for syntax). Both the fully-variant and the fully-invariant models failed to yield viable solutions once three classes were reached, and the mixed approach failed for a two-class model. The fully-variant model produced errors in the Fisher information matrix, while the fully-invariant and mixed models returned psi-matrix errors caused by the estimation of negative variances for quadratic and slope growth factors. It was observed that variances for the quadratic growth factor were small and non-significant, and the decision was

taken to run the models again with the quadratic variance constrained to zero across all classes and models.

The fully-variant and mixed models also failed to yield viable solutions with the variance-constrained quadratic factor, replicating the errors from the previous models. However, the fully-invariant (most conservative) model converged. Entropy statistics revealed that the model fit the data very well (.95 for two classes, through to .86 for seven classes). BIC, AIC and SSABIC values continued to decrease with the addition of more classes, replicating the pattern observed in the LCGA analyses. This was observed to continue until a twelve-class model was reached. The ALMR statistic was non-significant for all classes except the two- and five-class models, which diverged from the finding shown in the LCGA.

Consistently decreasing BIC values and significant BLRTs have been observed to arise frequently in GMM (Berlin, Parra and Williams, 2013). However, in order to investigate whether a model misspecification might be the cause of the continually improving fit indices, a number of other model specifications were tested. Firstly, residual covariances were added to the model, leading to psi-matrix error messages even when a single covariance was added. Next, gender was modelled as a known latent class. This model included a second latent class aspect, fixed at two classes, with binary logistic regression of the categorical latent class on the continuous unknown class aspect (Muthén & Shedden, 1999; Muthén, 2004; Muthén & Asparouhov, 2009). The improving fit indices were again replicated (Appendix 4). Following this, models were fitted separately for male and female participants. Fit indices again continued to improve to seven classes for females, and five classes for males, with eight- and six-class models respectively returning psi-matrix errors.

Following this, consideration was given to the possibility that the positively-skewed data with a high proportion of zeroes at each time point could require a different modelling approach (see Methods). A two-part GMM was specified, and the continuous part of this model showed continually-improving fit indices up to 11 classes, even following the application of natural logarithm transformations. Next, data were recoded into count outcomes, by summing positively-endorsed SCDC items at each time point. A GMM using count data also produced continually improving fit indices. This count data was then used in a zero-inflated Poisson GMM. Models were fit separately for the two components and then combined into a single model, which also showed improving fit indices until 11 classes were reached. Entropy statistics were moderate to poor (.64 to .70), and class posterior probabilities suggested low confidence in allocating participants to latent groups. It was therefore concluded that although floor effects and skewness were present in the data, approaches that are specifically modified to accommodate this distribution did not produce an improved model.

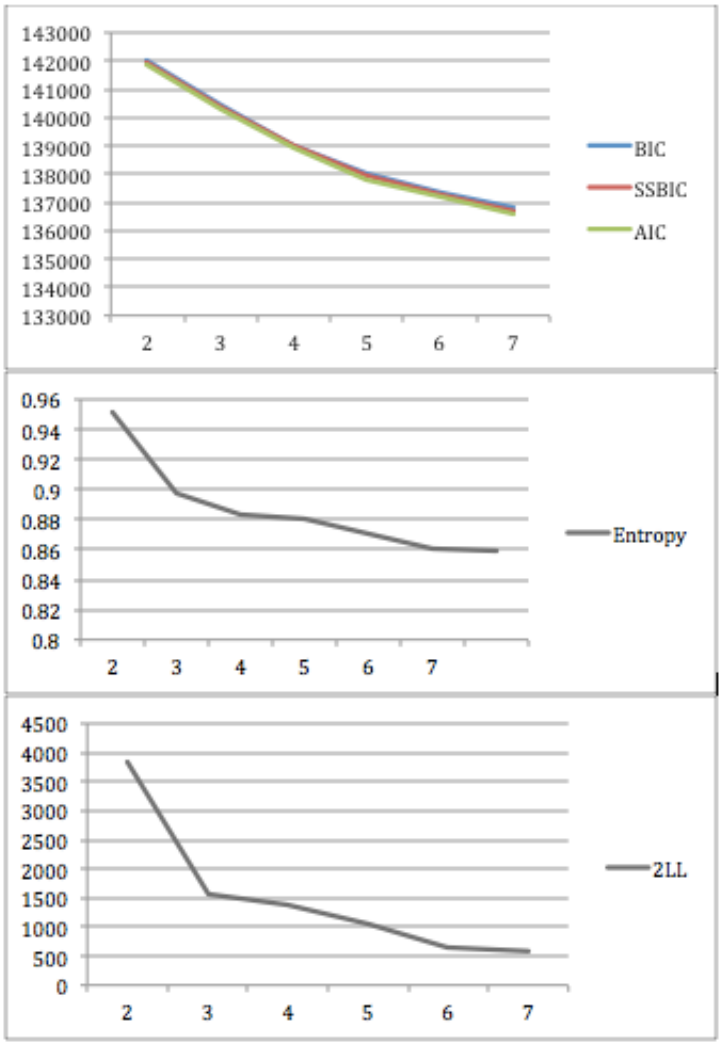


Figure 3. Fit indices for GMM with increasing classes

3.6 Model selection

Given that alternative approaches to handling the skewness and floor effects in the data showed similar patterns of improving fit indices as classes were added, the decision was taken to return to the more conservative approach of the restricted GMM, which produced superior entropy and class posterior probability statistics.

From the restricted GMM, a six-class solution was selected as the optimal model to present. As Figure 3 shows, an 'elbow' in 2-Log Likelihood values was observed once six classes were reached, following which improvements in model fit were substantially attenuated. Further, once the seven-class model was reached, one of the classes became very small (n=92), representing less than 1% of the data, which can determine an appropriate cut-off for adding further classes (Wickrama, Lee, Walker O'Neal, & Lorenz, 2016). Finally, for the seven-class model the additional class was largely drawn from a moderate increasing class from the six-class model, and appeared to be a variation on a theme rather than a substantively new trajectory. The final six-class model is shown in Figure 4; means and proportions of participants exceeding the cut-off of 8 are reported in Table 4. Additional details including variances and class posterior probabilities are reported in Appendix 5.

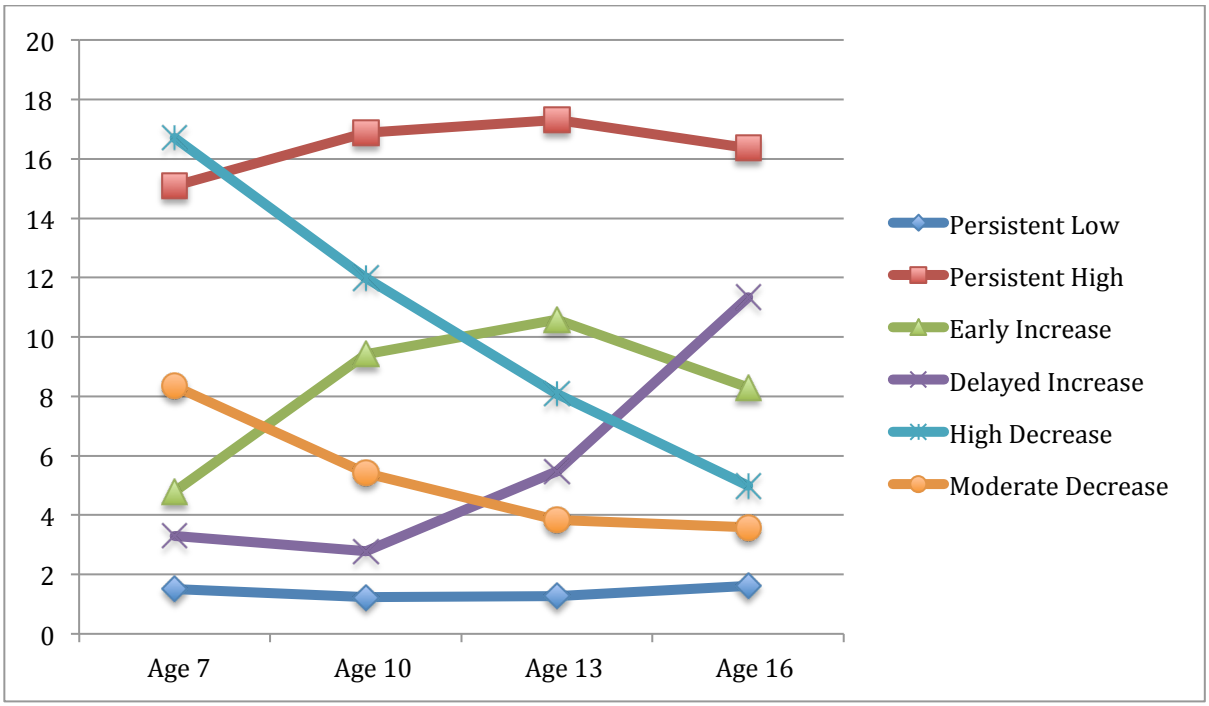


Figure 4. Six-class Growth Mixture Model for SCDC

Table 4.
Mean SCDC scores across time for latent classes

	Age 7	Age 10	Age 13	Age 16
High decrease (n=168)				
Mean	16.70	11.98	8.08	4.99
Residual	-0.07	0.19	-0.42	0.11
≥ 8 (n, %)	138 (100%)	114 (89%)	56 (53.3%)	20 (24.4%)
Delayed Increase (n = 475)				
Mean	3.29	2.78	5.47	11.34
Residual	0.02	-0.02	0.05	-0.15
≥ 8 (%)	29 (7.5%)	19 (4.8%)	152 (35.6%)	355 (98.3%)
Moderate Decrease (n = 781)				
Mean	8.35	5.41	3.83	3.59
Residual	0.02	-0.13	0.10	-0.12
≥ 8 (%)	449 (66.7%)	155 (25.5%)	55 (10.6%)	45 (10.8%)
Persistent Low (n = 7763)				
Mean	1.51	1.24	1.28	1.62
Residual	0.02	-0.06	0.07	-0.03
≥ 8 (%)	3 (<.01%)	33 (.01%)	36 (.01%)	48 (.01%)
Early Increase (n = 400)				
Mean	4.80	9.41	10.58	8.29
Residual	0.14	-0.45	0.40	-0.30
≥ 8 (%)	62 (19.3%)	224 (66.3%)	257 (87.4%)	132 (57.4%)
Persistent High (n = 157)				
Mean	15.09	16.88	17.30	16.36
Residual	-0.09	-0.07	-0.22	-0.53
≥ 8 (%)	119 (99.2%)	119 (97.5%)	88 (100%)	61 (100%)

Based on the growth patterns, the largest class (71.6%) was named “Persistent Low”. Members of this class began with an average SCDC score of 1.51

(significantly different from zero, $p< 0.001$), that decreased (linear slope), on average, by 0.04 every 3 years ($p<.001$) and accelerated with time (quadratic slope 0.15, $p<.001$) to a very small degree, marking a largely stable trajectory with no significant difference between mean scores at 7 and 16 ($t = 1.31$, d.f = 3959, $p = .14$). Class 3, called “Moderate Decrease”, had the second-highest proportion of members (8%). The mean initial SCDC score was 8.35 ($p<.001$) with a mean decrease of 3.61 every 3 years, accelerating with time (0.67, $p<.001$). The third largest class (4.9%) named “Delayed Increase” had an average initial SCDC score of 3.29 ($p < .001$), and the slope (-2.10, $p<.001$) and quadratic slope (1.60, $p<.001$) define a u-shaped trajectory with an initial significant decrease in mean scores between 7 and 10 (M=2.78; $t = 3.90$, d.f = 340, $p<.001$) followed by a sharp increase to a mean of 11.34 by age 16. The “Early Increase” group (4.1%) began with a mean of 4.80 at age 7, significantly higher than the delayed increase class ($t = 7.04$, d.f = 706, $p<.001$), and followed a positive (slope = 6.35, $p<.001$) and decelerating (quadratic slope = -1.73, $p=.01$) trajectory, rising to a peak of 10.58 at age 13 but then finishing lower at 8.29 at 16 years, such that there was no longer a significant difference from mean scores at age 10 ($t = 1.76$, d.f = 201, $p = .80$). The second-smallest class (1.7%) was named “High Decrease”, and began with a significantly higher mean score (16.70) than the “Persistent High” (15.09; 1.6%, smallest) group at age 7 ($t = 4.40$, d.f = 256, $p<.001$), following a significantly decreasing linear trajectory (slope = -5.13, $p<.001$) with no significant acceleration over time (quadratic slope = 0.41, $p = .15$). The Persistent High class showed a decreasing (2.47, $p = .03$) and decelerating trajectory (-0.68, $p = .041$), but despite increasing mean scores up to age 13, there was no significant difference between mean scores between ages 7 and 16 (16.35; $t = 1.26$, d.f = 51, $p = .21$) indicating stability over time for both the “persistent” groups.

The proportions of participants across time exceeding the clinical cut-off of 8 reinforced the interpretability of the classes. At age 7, 100% of participants in the High Decrease class exceeded the cut-off, reducing to 24.4% of participants by age 16; conversely, only 7.5% of participants in the Delayed Increase class exceeded the 8-threshold at age 7, rising to 98.3% of the class at age 16. Proportions remained at .01% or less in the Persistent Low class, and 97.5% to 100% in the Persistent High class.

3.7 Post-hoc tests

3.7.1 Gender

Table 5.

Gender composition of trajectory classes

Class	Male (n, %)	Female (n, %)
High Decrease	119, 70.8%	49, 29.2%
Delayed Increase	202, 42.5%	273, 57.5%
Moderate Decrease	456, 58.4%	325, 41.6%
Persistent Low	3843, 49.5%	3920, 50.5%
Early Increase	222, 55.5%	178, 44.5%
Persistent High	118, 75.2%	39, 24.8%

Each latent class was examined in terms of its gender composition, to determine whether female participants were overrepresented in any classes (Table 5). A multinomial logistic regression with the Persistent Low class as the reference group revealed male participants had an increased odds of being in the Persistently High Group (OR = 3.09, 95% CI [2.14, 4.45]), but female participants had increased odds of being in the Delayed Increase class (OR = 1.32, 95% CI [0.63 - 0.91]). It is notable that as theory would predict, the Persistent Low class showed a relatively equal gender balance, while the Persistent High class showed a ratio of 3:1

between men and women (Loomes et al., 2017). The Delayed Increase class was the only class in which females represented the significant majority.

3.7.2 Wechsler Intelligence Scale for Children

Data from the WISC, measured at 8 years for all participants, were investigated relative to class membership. WISC total IQ was observed to follow the expected normal distribution. A one-way ANOVA revealed a significant main effect of trajectory group ($F = 5375.78$, $d.f = 5$, $p<.001$) on IQ score. Post-hoc tests revealed that all classes significantly differed from one another after Bonferroni corrections were applied. This pattern of difference was replicated between verbal and performance IQ.

Table 6.

IQ scores by trajectory group

Class	Total IQ Mean (S.E)	Verbal IQ Mean (S.E)	Performance IQ (S.E)
High Decrease	98.629 (1.650)	102.052 (1.683)	94.979 (1.717)
Delayed Increase	102.856 (.833)	106.164 (.846)	98.029 (.865)
Moderate Decrease	101.029 (.689)	104.106 (.702)	97.215 (.715)
Persistent Low	105.606 (.220)	108.399 (.224)	100.932 (.228)
Early Increase	99.996 (.964)	103.545 (.980)	95.737 (1.002)
Persistent High	93.193 (1.733)	96.551 (1.757)	90.124 (1.793)

The highest IQ scores were observed for the Persistent Low group, and the lowest scores for the Persistent High group, with the increasing and decreasing classes demonstrating intermediate scores. No significant difference in IQ between sexes was observed within classes (all $p>.2$).

3.7.3 Strengths and Difficulties Questionnaire

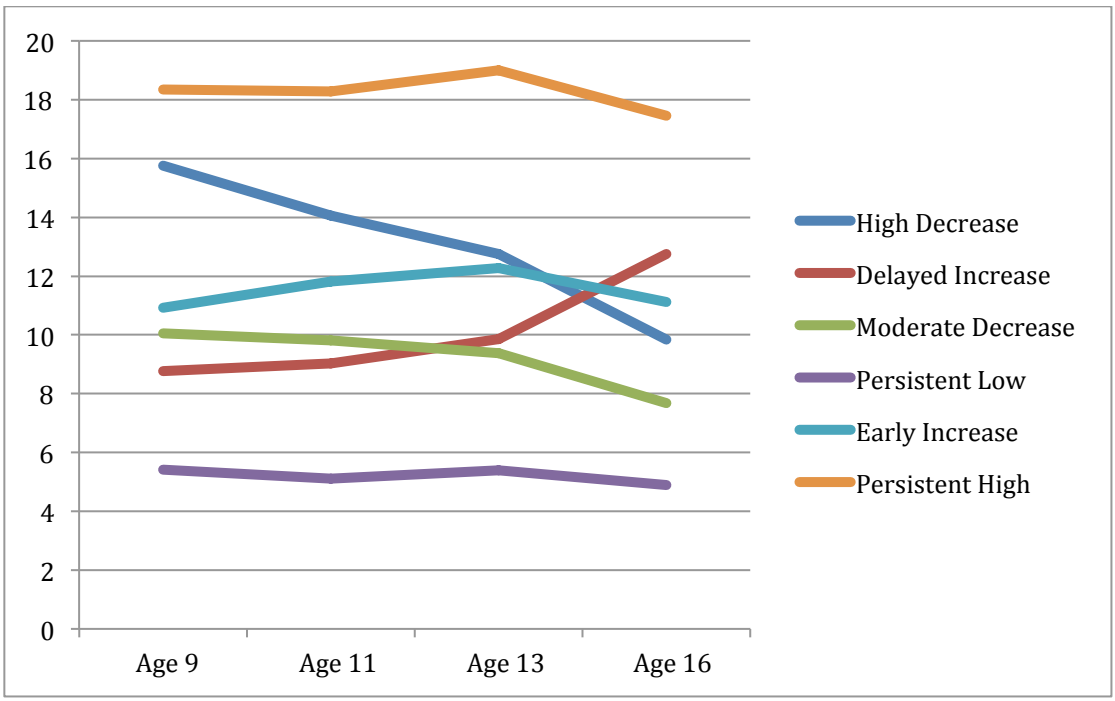


Figure 5. SDQ means across time by group

Due to the large sample size, no multivariate normality assumption was required for a Repeated-Measures ANOVA on SDQ scores with latent class as the grouping variable. Sphericity assumptions were not met (Mauchley's W = .915, d.f = 5, p<.001), and Greenhouse-Geisser's ε was large (.944), so Huynh-Feldt corrected statistics were used. A significant main effect of group (F = 456.496, d.f=5, p<.001) and a significant group-by-time interaction (F = 46.789, d.f=14.179, p<.001) were observed. Post-hoc tests with Bonferroni corrections applied revealed significant differences between all groups, with the Persistent High group maintaining a particularly large separation in elevated scores maintained until 16 years of age. As the plot in Figure 5 reveals, each class followed comparable SDQ trajectories over time to their SCDC trajectories.

Interestingly, while a significant increase was observed between first (M = 8.78, SD = 4.84) and last (M = 12.74, SD = 5.47) time points for the Delayed Increase group (t = 12.74, p<.001), there was no significant change between baseline and final measurement for the Early Increase group (M = 11.12, 11.41, SD = 5.15, 5.77, t = .703, p = .483). This makes sense conceptually, given that the first measurement for the SDQ occurred at age 9, whereas the first SCDC score was taken at age 7. By contrast the Delayed Increase group's scores did not change significantly between nine (M = 8.79, SD = 4.89) and eleven (M = 9.09, SD = 4.93) years old (t = 1.45, p = .147), but increased significantly by 16 years (M = 9.09, 12.84, SD = 4.93, 5.64, t = 11.73, p<.001). A small but significant decrease occurred in scores for the Persistent Low group between 9 (M = 5.46, SD = 3.79) and 16 (M = 4.96, SD = 3.71; t = 7.93, p<.001), whereas no significant change was observed between 9 (M = 18.21, SD = 5.85) and 16 (M = 17.29, SD 5.37) in the Persistent High group (t = 1.04, p =.305). The High Decrease group's scores dropped significantly between age 9 (M = 15.58, SD = 5.20) and 16 (M = 9.94, SD = 5.08; t = 7.16, p<.001), as did those for the Moderate Decrease group (M = 10.13, 7.72, SD = 5.00, 4.63, t = 9.18, p<.001).

The differential association of latent group classes with SDQ and IQ scores supports the conclusion that these are meaningful groups, rather than artefacts of a non-normal distribution (Bauer & Curran, 2003).

4. Discussion

4.1 Summary of key findings

This study represented the first application of LCGM approaches to modelling ATs in a general population sample. LCGA and GMM were employed, and a quadratic

GMM was selected as the optimal model. It was observed that six different trajectories best characterised the patterns of change between 7 and 16 years of age. Two of these trajectories were stable, one a Persistent Low class comprising 72% of the sample, and one a Persistent High class representing less than 2% of participants. Two increasing classes were found – one showing an early increase between 7 and 10 years representing 4% of the sample, and another showing a later increase between 10 and 16 that was especially marked between 13 and 16, characterising 5% of participants. Two decreasing trajectories marked a decline from particularly severe mean score (16.7, significantly higher than even the Persistent High class at 15.1) in the High Decrease class, which represented less than 2% of the sample, and from a more intermediate score of 8.3 in the Moderate Decrease class, which included 8% of participants.

In line with theory, the sexes were evenly balanced in the Persistent Low group, and the Persistent High group contained a 3:1 ratio of males to females, closely matching the ratio observed in active recruitment studies (those which actively sought cases in the community regardless of prior diagnosis; Loomes et al., 2017). There were relatively even greater numbers of males (71%) in the High Decrease group, which showed a dramatic decline in severity of symptoms in a linear fashion between 7 and 16. Interestingly, the absolute numbers of boys in the Persistent High and High Decrease group were similar, suggesting that around half of boys with very severe SCDC scores age 7 improved markedly, while the other half retained severe symptoms. The Moderate Decrease group showed a larger relative number of girls, which appeared to reflect the fact that fewer girls had very severe scores. Strikingly, the Delayed Increase group actually showed a majority of girls, forming 58% of the group, while the Early Increase group also had a higher relative proportion of females (45%) than any of the groups that began with severe scores. This mirrors the finding that girls are diagnosed later (Dworzynski, Ronald, Bolton, &

Happé, 2012; Loomes et al., 2017; Russell et al., 2011), and suggests that this may be linked to later symptom onset.

IQ was observed to be significantly higher in the Persistent Low class than all other classes, and lower in the Persistent High class than all other classes; increasing and decreasing trajectories showed intermediate scores. SDQ scores closely reflected the expected trajectories predicted by classes, with a particularly marked increase occurring in the Delayed Increase group between 13 and 16. This helped support the validity of the classes that were observed, by providing evidence that they were differentially related to other outcomes (Wickrama et al., 2016), whilst also suggesting that ATs and psychological strengths and difficulties are closely linked developmentally.

4.2 Findings in the context of previous research

As Chapter 1 summarises in a review of the literature on development of ATs over time, it has been consistently observed that ATs tend to decrease over time in participants with ASD diagnoses. This is linked to improvement in social and communication sub-domains, rather than restricted interests and repetitive behaviours (RRB), which appear to remain broadly stable over time. This study built on the previous literature in two important ways – ATs were examined in general population participants, and latent subgroups representing distinct trajectories were examined. This revealed that a single trajectory for all participants masks significant complexity, and different subgroups of participants follow stable, improving and worsening trajectories over time.

This study extended the work of Robinson et al (2011b), which noted broad stability of SCDC scores between the ages of 7 and 13. Firstly, this study extended the time period under consideration until mid-adolescence (age 16), and it was observed that by the age of 16 the previously observed difference between male and female scores was no longer significant – males' scores remained stable, while females' scores increased between 13 and 16. Secondly, it was observed that these trends across all participants masked notable heterogeneity within trajectories, and six different patterns of development could represent meaningful sub-groups within participants.

Previous applications of LCGM techniques to ATs have focused on participants with ASD diagnoses, representing the extreme upper end of the AT range. Fountain et al (2012) reported a six-class model of CDER scores, and observed an improving trajectory representing 7.5% of the sample, which the authors named "Bloomers". They did not observe any increasing trajectories, which is likely to reflect the fact that participants all had ASD diagnoses. Similarly, those in the consistently-low trajectory class represented a small proportion of the sample (8.4%), while in this study 71.6% of the sample followed this trajectory, again reflecting the use of the general population participants in which a greater percentage of low ATs would be expected.

Increasing trajectories have been observed in four other studies using LCGMs with participants with majority ASD-diagnosed samples, each of which presented a four-class model (Gotham et al., 2012; Lord & Luyster, 2006; Lord et al., 2012; Venker et al., 2014). This was interesting given that studies restricted to those with diagnoses by definition exclude participants who begin with low scores and only develop more severe symptoms over time. However, close examination reveals that these trajectories were questionable and differed greatly from the clearer and more

dramatic increases observed in the present study. Gotham et al (2012) selected a four-class model, in which a worsening (increasing) trajectory represented 9% of the sample. Mean ADOS scores in this class showed a small increase between the ages of 3 and 7, but this was not significant. Furthermore, Autism Diagnostic Interview-Revised (ADI-R; Le Couteur, Lord, & Rutter, 2003) Social and Non-Verbal Communication scores were observed to follow a decreasing mean trend within this group, and 70% of the increasing class was composed of participants with PDD-NOS or non-ASD diagnoses. Similarly Lord et al (2012) reported a worsening trajectory (n=16), but increase in ADOS scores was non-significant, 93% of the sample did not show regression according to ADI-R scores, and 87.5% of the group had PDD-NOS or non-spectrum diagnoses. The worsening trajectory reported by Venker et al (2014) included only 10 participants and showed a gentle slope within the mild to moderate range between the ages of 2 and 7. Characteristics of classes were not provided by Lord and Luyster (2006). Therefore, while increasing trajectories have been observed in the literature previously, this study has been the first to use entirely general population participants and the first to observe a significantly increasing trajectory from a low to a high severity range.

Previous studies have included a majority of male participants (median score 79%; Chapter 1), which reflects the male-to-female ratio of 4:1 observed in passive recruitment studies (i.e. using participants who already have clinical diagnoses) by Loomes et al (2017). By contrast, the 3:1 ratio observed in the Consistent High group in this study is consistent with the sex ratio of severe scores observed in active recruitment studies. This study was able to address the problematic underrepresentation of females that has limited previous studies (Lai et al., 2015; Loomes et al., 2017) by including an equal representation of both sexes. Under these circumstances, it was possible to observe that both of the increasing trajectories had overrepresentations of females compared to the other higher-

scoring classes. Previous studies have revealed inconsistent relationships between sex and ATs, with three studies having found sex to not significantly predict class membership (Gotham et al., 2012; Lord et al., 2012; Venker et al., 2014), and one study having found female sex predicted membership of an improving class, and male sex predicted membership of a stable and severe class (Szatmari et al., 2015). The difference in finding could be explained by the use of general population participants in this study, which has reduced bias against females and allowed the emerging class to be captured. Further replication of LCGM with general population participants is needed to confirm this.

As noted above, there are two competing theories of female underrepresentation in diagnostically defined Autism – the "female compensation hypothesis" predicting severity scores would decrease over time and the "adolescent onset hypothesis" predicting severity scores would increase over time. This study provides evidence that contradicts the female compensation hypothesis, and supports the adolescent onset hypothesis. Firstly, mean scores across all participants increased between ages 13 and 16 such that the mean difference between sexes was not significant by mid-adolescence. Further, females were overrepresented in increasing trajectory groups over time, and particularly so in the group showing its first increase after the age of 13, in which females were the majority.

This study observed two decreasing trajectories, one beginning at an especially severe score and dramatically increasing, and one showing a more moderate decrease from intermediate scores. Previous studies using LCGM techniques have consistently discovered subgroups of participants whose scores improve over time from those who had ASD diagnoses at baseline (Fountain et al., 2012; Lord & Luyster, 2006; Lord et al., 2012; Szatmari et al., 2015; Venker et al., 2014), representing proportions of participants ranging from 7 to 14%. This study not only

supports the finding of notably improving trajectories over time, but also suggests that these may have been underestimated in samples restricted largely to those with diagnoses – equal absolute numbers of boys were found in the Persistent High and High Decreasing groups, suggesting around half of those with very high initial scores followed improving trajectories.

The observation of higher IQ in the stable low class, and lower IQ in the stable severe class accords with theory (Mayes & Calhoun, 2011), but the finding that the increasing and decreasing trajectories were related to intermediate IQ scores appears to be a novel finding. Previous studies using LCGM have found inconsistent results, with one study finding higher verbal IQ predicted membership of an improving class (Gotham et al., 2012), and another finding lower non-verbal IQ predicted membership of a severe class (Venker et al., 2014). However, these studies recruited a majority of participants with ASD diagnoses, and hence trajectories marking improvement from low to high scores, and consistently low scores (relative to general population) could not be observed. Further LCGM examining ATs in general population participants will help to clarify whether this finding is replicated.

The comparison of SDQ scores between latent classes, which were in turn defined using SCDC scores, provided interesting results. Trajectories of difficulties were observed to follow similar trajectories within groups to those shown in ATs. There are several possible interpretations of this finding. Firstly, the SCDC and SDQ are moderately correlated (Skuse et al., 2009; Kendall's $\tau = .46$, $p<.001$). Hence, the finding of similar trajectories in SDQ scores within classes defined by the SCDC could reflect this correlation. Nonetheless, the relationship between these constructs, which holds within classes, requires explanation. It could indicate that the SCDC itself is simply measuring general psychopathology, or social impairment

driven by this psychopathology, rather than ATs. However, this appears unlikely in the context of the finding that SCDC predicts anxiety, but causation does not run in the other direction (Pickard, Rijsdijk, Happé, & Mandy, 2017). It could be the case that changes in ATs measured by the SCDC increase risk for general psychopathology, or that an unknown exogenous variable drives risk for both social and communication difficulties and general psychopathology. This finding should therefore be interpreted cautiously, and is perhaps more usefully viewed as a spur to further research than as a conclusive finding.

4.3 Limitations and future research directions

A key limitation of this study is that the SCDC is a measure of social and communication difficulties, and therefore does not capture RRB, another core definitional domain of ASD (American Psychiatric Association, 2013). However, the SCDC is a well-validated measure (Skuse et al., 2005), showing high heritability in adolescence (St Pourcain et al., 2014) and shared genetic risk with ASD (Robinson et al., 2016). Nonetheless, it could have strengthened the study if a measure of RRB such as the Repetitive Behavior Scale-Revised (RBS-R; Lam & Aman, 2007) or a measure of overall Autism symptoms such as the ADI-R (Le Couteur et al., 2003) could have been available to supplement the findings presented above. Furthermore, the SCDC is a parent-report measure, and therefore reliant on a single source of information, while gold standard practice would require a multi-informant, multi-method and multi-context approach (Möricke, Buitelaar, & Rommelse, 2016). Comparison with a measure such as the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1999), which uses direct observational techniques outside of the home environment, would therefore also potentially increase the validity of the

findings, although this would be practically very challenging in a large population cohort.

A further limitation lies in the final sample being composed disproportionately of White, middle-class participants. This could potentially limit generalisability of the findings, especially in light of the evidence showing children of migrant parents for example are at increased risk of ASD (Keen, Reid, & Arnone, 2010); findings for socioeconomic status and its link to ASD however are inconsistent (Durkin et al., 2010; Larsson et al., 2005; Rai et al., 2012). Future studies should aim to replicate LCGM approaches using general population samples that are more diverse and representative of the UK population, as well as using global data. Moreover, this dataset was limited by the restriction of the age range to mid-adolescence (16 years) – it will be important for future research to examine longitudinal trajectories of ATs into adulthood to determine whether the observed trends are maintained across time.

The fact that the BIC and BLRT fit indices continued to improve until a twelve-class model was reached introduced difficulties in selecting the optimal number of classes to present. A strength of the study is that there was scope to follow the guidelines on conducting and presenting a GMM (Ram & Grimm, 2009) so that this process was transparent and the decision-making process clear. Furthermore, a range of approaches to handling the skewed data was employed, which replicated the core findings. However, as Bauer and Curran (2003) argue in their critique of the application of GMM to psychological research, there is a risk of over-extraction of spurious classes, given that GMM can serve both to represent true sub-populations in heterogeneous data *and* to simply characterise a non-normal population in terms of (ecologically meaningless) normally-distributed sub-groups of data. To this end, the careful analysis of the trajectories of the sub-groups, their concordance with

substantive theory and their differential relation to IQ and SDQ scores supported the conclusion that these were meaningful groups of individuals rather than pseudo-random packages of data. Nonetheless, two objections against the selection of the six-class model could be raised – that the twelve-class model should have been selected, or that a lower number of classes could have been selected. The final decision reflected a pragmatic compromise and attempted to select the most conservative, parsimonious model that contained both the largest number of manageable classes (to be true to the improving fit indices), and that did not include additional classes representing relatively minor improvements and lacking clinical significance or conceptual clarity. The emergence of a class that contained less than one percent of the data and was largely drawn from an existing class was therefore taken as an appropriate cut-point.

It has been established that girls are at increased risk of services failing to give a needed diagnosis (Lai & Baron-Cohen, 2015; Loomes et al., 2017), are diagnosed later than males with equivalent levels of ATs (Giarelli et al., 2010), and teachers are less likely to notice ATs in females (Posserud et al., 2006). This study highlights the fact that a distinct group of majority-female participants experienced increasing symptoms over time, but was not able to determine whether additionally this group displays different phenotypic characteristics compared to equivalent males, or whether the SCDC is as sensitive to a female phenotype of ASD that may be more characterised by internalising problems (Mandy et al., 2012; Huke, Turk, Saeidi, Kent, & Morgan, 2013). The possibility remains that the SCDC is capturing a non-autistic syndrome in these later-presenting participants, although the SCDC has been shown to measure social communication difficulties independently of other conditions (Robinson et al., 2011b; Mandy et al., 2013; Pickard et al., 2017). It is also important to note that a large number of males were also present in this later-presenting category, and it could be that rather than a different phenotype being

exclusively female, sex and phenotype may be related probabilistically rather than categorically (Baron-Cohen, 2010). Future studies are needed to establish whether the relationship between female sex and an increasing AT phenotype is replicated.

Future research could examine genetic risk in relation to AT chronogeneity. Evidence has linked genetic influences on ASD risk to the entire range of AT variability (Robinson et al., 2016; St Pourcain et al., 2010). Understanding of the aetiology of Autism could be greatly increased by accounting for heterogeneity, as previous attempts to reduce phenotypic heterogeneity in order to increase genetic homogeneity have failed (Chaste et al., 2015). Currently there is little evidence for the large effects of the aggregates of multiple individual common genetic variants for ASD (Anney et al., 2012). This can be researched using Polygenic Risk Scores (PRS; Plomin, Haworth, & Davis, 2009). These are scores that summarise risk from a range of small genetic risk loci, and are especially useful for conceptualising disorders as quantitative traits (Plomin et al., 2009). PRS could be used in conjunction with latent trajectory classes of ATs, to determine whether different genetic risk profiles are associated with these phenotypic presentations.

Finally, the relationship between AT trajectories and general psychopathology needs further clarification. While this study demonstrated similar trajectories in SCDC and SDQ scores over time within groups, the causal relationship between AT trajectories and general difficulties is still unclear. Future studies could investigate relationships between ATs and other measures of functioning such as daily living skills, or general difficulties and psychopathology, using Sequential Process GMM or Latent Transition Analysis (Muthén & Muthén, 2017b), which investigate latent classes within parallel processes. Alternatively, further work could be done to examine the relationship of latent AT classes to distal (future) outcomes of a similar nature using GMM (Muthén & Muthén, 2017b).

4.4 Conclusion: Scientific and Clinical Implications

This study was the first to apply LCGM approaches to study ATs in the general population. The finding that the chronogeneity of ATs can be characterised by distinct sub-groups including substantially increasing and decreasing trajectories has important implications for charting developmental profiles, improving prognostic prediction (Baker, Seltzer, & Greenberg, 2011; Lord, Leventhal, & Cook, 2001; Lord et al., 2012; Soke et al., 2011), and improving our understanding of ASD aetiology and nosology (Kasari et al., 2014; Magiati et al., 2014). As Szatmari et al argued, "We have known for some time that ASD is not a homogeneous syndrome. The question is not 'whether' to account for heterogeneity, but 'how'" (2009, p.1465). This study suggests that LCGM approaches such as GMM may be one useful methodology to answer this question.

The timely identification of ASD can improve quality of life, signpost access to appropriate services, promote positive identity and reduce stigma and blame (Portway & Johnson, 2005; Ruiz Calzada, Pistrang, & Mandy, 2012; Russell & Norwich, 2012: Wong et al., 2015). A key clinical implication of this study is that it will therefore be important for teachers and clinicians to be responsive to the fact that some girls with ASD may first present with noticeable symptoms in adolescence, later than the usual age at which diagnoses are given, and without a childhood history of displaying ATs. Furthermore, it may also be the case that a number of male participants may also present with these symptoms later, highlighting the need for further research to characterise these longitudinal phenotypes in probabilistic fashion regarding sex.

Georgiades et al (2017) define chronogeneity as operating at three levels of understanding: individual, group, and individual deviation from group norms. This

study has aimed to contribute to this important understanding of ASD chronogeneity, by characterising inter-individual differences in intra-individual change. Replicating these findings in new data sets and establishing the validity and generalisability of these developmental phenotypes will lead to the chance to further characterise these groups at genetic, environmental and behavioural levels, further feeding into clinical benefits including improved detection and targeted, adaptive intervention strategies (Almirall & Chronos-Tuscano, 2016).

References

Agresti, A. (1996). *An introduction to categorical data analysis* (Vol. 135). New York: Wiley.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*(3), 317-332.

Almirall, D., & Chronis-Tuscano, A. (2016). Adaptive interventions in child and adolescent mental health. *Journal of Clinical Child & Adolescent Psychology*, *45*(4), 383-395.

Amaral, D. G., Schumann, C. M., & Nordahl, C. W. (2008). Neuroanatomy of autism. *Trends in neurosciences*, *31*(3), 137-145.

Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., ... & Brennan, S. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorders. *Human molecular genetics*, *21*(21), 4781-4792.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397-438.

Baker, J. K., Seltzer, M. M., & Greenberg, J. S. (2011). Longitudinal effects of adaptability on behavior problems and maternal depression in families of adolescents with autism. *Journal of Family Psychology*, *25*(4), 601.

Banaschewski, T., Hollis, C., Oosterlaan, J., Roeyers, H., Rubia, K., Willcutt, E., & Taylor, E. (2005). Towards an understanding of unique and shared pathways in the psychopathophysiology of ADHD. *Developmental science*, *8*(2), 132-140.

Baron-Cohen, S. (2010). Empathizing, systemizing, and the extreme male brain theory of autism. *Progress in brain research*, *186*, 167-175.

Baron-Cohen, S., Scott, F. J., Allison, C., Williams, J., Bolton, P., Matthews, F. E., & Brayne, C. (2009). Prevalence of autism-spectrum conditions: UK school-based population study. *The British Journal of Psychiatry*, *194*(6), 500-509.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological methods*, *8*(3), 338.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin*, *107*(2), 238.

Berlin, K. S., Parra, G. R., & Williams, N. A. (2013). An introduction to latent variable mixture modeling (part 2): Longitudinal latent class growth analysis and growth mixture models. *Journal of Pediatric Psychology*, *39*(2), 188-203.

Bargiela, S., Steward, R., & Mandy, W. (2016). The experiences of late-diagnosed women with autism spectrum conditions: an investigation of the female autism phenotype. *Journal of autism and developmental disorders*, *46*(10), 3281.

Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., ... & Davey Smith, G. (2012). Cohort profile: the 'children of the 90s'—the index offspring of the

Avon Longitudinal Study of Parents and Children. *International journal of epidemiology*, *42*(1), 111-127.

Byrne, B. M. (2013). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press.

Chen, Q., Luo, W., Palardy, G. J., Glaman, R., & McEnturff, A. (2017). The Efficacy of Common Fit Indices for Enumerating Classes in Growth Mixture Models When Nested Data Structure Is Ignored: A Monte Carlo Study. *SAGE Open*, *7*(1), 1-19.

Constantino, J. N. (2009). How continua converge in nature: cognition, social competence, and autistic syndromes. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*(2), 97-98.

Constantino, J. N., & Todd, R. D. (2000). Genetic structure of reciprocal social behavior. *American Journal of Psychiatry*, *157*(12), 2043-2045.

Constantino, J. N., Przybeck, T., Friesen, D., & Todd, R. D. (2000). Reciprocal social behavior in children with and without pervasive developmental disorders. *Journal of Developmental and Behavioral Pediatrics*.

DiCicco-Bloom, E., Lord, C., Zwaigenbaum, L., Courchesne, E., Dager, S. R., Schmitz, C., ... & Young, L. J. (2006). The developmental neurobiology of autism spectrum disorder. *Journal of Neuroscience*, *26*(26), 6897-6906.

Durkin, M. S., Maenner, M. J., Meaney, F. J., Levy, S. E., DiGuiseppi, C., Nicholas, J. S., ... & Schieve, L. A. (2010). Socioeconomic inequality in the prevalence of autism

spectrum disorder: evidence from a US cross-sectional study. *PLoS One*, *5*(7), e11551.

Dworzynski, K., Ronald, A., Bolton, P., & Happé, F. (2012). How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders?. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(8), 788-797.

Fountain, C., Winter, A. S., & Bearman, P. S. (2012). Six developmental trajectories characterize children with autism. *Pediatrics*, *129*(5), e1112-e1120.

Georgiades, S., Bishop, S. L., & Frazier, T. (2017). Editorial Perspective: Longitudinal research in autism–introducing the concept of 'chronogeneity'. *Journal of Child Psychology and Psychiatry*, *58*(5), 634-636.

Giarelli, E., Wiggins, L. D., Rice, C. E., Levy, S. E., Kirby, R. S., Pinto-Martin, J., & Mandell, D. (2010). Sex differences in the evaluation and diagnosis of autism spectrum disorders among children. *Disability and health journal*, *3*(2), 107-116.

Gilmour, J., Hill, B., Place, M., & Skuse, D. H. (2004). Social communication deficits in conduct disorder: a clinical and community survey. *Journal of Child Psychology and Psychiatry*, *45*(5), 967-978.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), 1337-1345.

Goodman, R., Ford, T., Richards, H., Gatward, R., & Meltzer, H. (2000). The Development and Well-Being Assessment: description and initial validation of an integrated

assessment of child and adolescent psychopathology. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(5), 645-655.

Gotham, K., Pickles, A., & Lord, C. (2012). Trajectories of autism severity in children using standardized ADOS scores. *Pediatrics*, *130*(5), e1278-e1284.

Greenbaum, P. E., Del Boca, F. K., Darkes, J., Wang, C. P., & Goldman, M. S. (2005). Variation in the drinking trajectories of freshmen college students. *Journal of consulting and clinical psychology*, *73*(2), 229.

Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(2), 202-226.

Herbert, M. R., & Anderson, M. P. (2008). An Expanding Spectrum of Autism Models. In *Autism* (pp. 429-463). Humana Press.

Hsiao, M. N., Tseng, W. L., Huang, H. Y., & Gau, S. S. F. (2013). Effects of autistic traits on social and school adjustment in children and adolescents: The moderating roles of age and gender. *Research in Developmental Disabilities*, *34*(1), 254-265.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

Huke, V., Turk, J., Saeidi, S., Kent, A., & Morgan, J. (2013). Autism spectrum disorders in eating disorder populations: a systematic review. *European Eating Disorders Review*, *21*(5), 345-351.

Jedidi, K., Ramaswamy, V., & DeSarbo, W. S. (1993). A maximum likelihood method for latent class regression involving a censored dependent variable. *Psychometrika*, *58*(3), 375-394.

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and personality psychology compass*, *2*(1), 302-317.

Kasari, C., Kaiser, A., Goods, K., Nietfeld, J., Mathy, P., Landa, R., ... & Almirall, D. (2014). Communication interventions for minimally verbal children with autism: A sequential multiple assignment randomized trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(6), 635-646.

Kaufman, A. S., & Lichtenberger, E. O. (2000). *Essentials of WISC-III and WPPSI-R assessment*. John Wiley & Sons Inc.

Keen, D. V., Reid, F. D., & Arnone, D. (2010). Autism, ethnicity and maternal immigration. *The British Journal of Psychiatry*, *196*(4), 274-281.

Knapp, M., Romeo, R., & Beecham, J. (2009). Economic cost of autism in the UK. *Autism*, *13*(3), 317-336.

Kopp, S., & Gillberg, C. (2011). The Autism Spectrum Screening Questionnaire (ASSQ)-Revised Extended Version (ASSQ-REV): an instrument for better capturing the autism phenotype in girls? A preliminary study involving 191 clinical cases and community controls. *Research in developmental disabilities*, *32*(6), 2875-2888.

Lai, M. C., & Baron-Cohen, S. (2015). Identifying the lost generation of adults with autism spectrum conditions. *The Lancet Psychiatry*, *2*(11), 1013-1027.

Lai, M. C., Lombardo, M. V., Auyeung, B., Chakrabarti, B., & Baron-Cohen, S. (2015). Sex/gender differences and autism: setting the scene for future research. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(1), 11-24.

Lam, K. S., & Aman, M. G. (2007). The Repetitive Behavior Scale-Revised: independent validation in individuals with autism spectrum disorders. *Journal of autism and developmental disorders*, *37*(5), 855-866.

Larsson, H. J., Eaton, W. W., Madsen, K. M., Vestergaard, M., Olesen, A. V., Agerbo, E., ... & Mortensen, P. B. (2005). Risk factors for autism: perinatal factors, parental psychiatric history, and socioeconomic status. *American journal of epidemiology*, *161*(10), 916-925.

Le Couteur, A. N. N. E., Lord, C., & Rutter, M. I. C. H. A. E. L. (2003). The autism diagnostic interview-revised (ADI-R). *Los Angeles, CA: Western Psychological Services*.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198-1202.

Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data*. John Wiley & Sons.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*(3), 767-778.

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*.

Lord, C., & Luyster, R. (2006). Early diagnosis of children with autism spectrum disorders. *Clinical Neuroscience Research*, *6*(3), 189-194.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999). Autism diagnostic observation schedule (ADOS) manual. *Los Angeles, CA: Western Psychological Services*.

Lord, C., Leventhal, B. L., & Cook, E. H. (2001). Quantifying the phenotype in autism spectrum disorders. *American Journal of Medical Genetics Part A*, *105*(1), 36-38.

Lord, C., Luyster, R., Guthrie, W., & Pickles, A. (2012). Patterns of developmental trajectories in toddlers with autism spectrum disorder. *Journal of consulting and clinical psychology*, *80*(3), 477.

Lundström, S., Chang, Z., Kerekes, N., Gumpert, C. H., Råstam, M., Gillberg, C., ... & Anckarsäter, H. (2011). Autistic-like traits and their association with mental health problems in two nationwide twin cohorts of children and adults. *Psychological medicine*, *41*(11), 2423-2433.

Lundström, S., Chang, Z., Råstam, M., Gillberg, C., Larsson, H., Anckarsäter, H., & Lichtenstein, P. (2012). Autism spectrum disorders and autisticlike traits: similar etiology in the extreme end and the normal variation. *Archives of general psychiatry*, *69*(1), 46-52.

Magiati, I., Tay, X. W., & Howlin, P. (2014). Cognitive, language, social and behavioural outcomes in adults with autism spectrum disorders: a systematic review of longitudinal follow-up studies in adulthood. *Clinical Psychology Review*, *34*(1), 73-86.

Mandy, W., Chilvers, R., Chowdhury, U., Salter, G., Seigal, A., & Skuse, D. (2012). Sex differences in autism spectrum disorder: evidence from a large sample of children and adolescents. *Journal of autism and developmental disorders*, *42*(7), 1304-1313.

Mandy, W., Skuse, D., Steer, C., St Pourcain, B., & Oliver, B. R. (2013). Oppositionality and socioemotional competence: interacting risk factors in the development of childhood conduct disorder symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*(7), 718-727.

Mayes, S. D., & Calhoun, S. L. (2011). Impact of IQ, age, SES, gender, and race on autistic symptoms. *Research in Autism Spectrum Disorders*, *5*(2), 749-757.

McGovern, C. W., & Sigman, M. (2005). Continuity and change from early childhood to adolescence in autism. *Journal of Child Psychology and Psychiatry*, *46*(4), 401-408

McLachlan, G. J., & Peel, D. (2000). Robust mixture modelling using the t distribution. *Statistics and Computing*, *10*(4), 335-344.

Mieloo, C., Raat, H., van Oort, F., Bevaart, F., Vogel, I., Donker, M., & Jansen, W. (2012). Validity and reliability of the Strengths and Difficulties Questionnaire in 5–6 year olds: differences by gender or by parental education?. *PloS one*, *7*(5), e36805.

Möricke, E., Buitelaar, J. K., & Rommelse, N. N. (2016). Do we need multiple informants when assessing autistic traits? The degree of report bias on offspring, self, and spouse ratings. *Journal of autism and developmental disorders*, *46*, 164.

Muthen, B. (2001). Latent variable mixture modeling. *New developments and techniques in structural equation modeling*, 1-33.

Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: comment on Bauer and Curran (2003).

Muthén, B. (2004). Latent variable analysis. *The Sage handbook of quantitative methodology for the social sciences*, 345-368.

Muthén, B., & Asparouhov, T. (2015). Growth mixture modeling with non-normal distributions. *Statistics in medicine*, *34*(6), 1041-1058.

Muthén, B. O, & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and experimental research*, *24*(6), 882-891.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural equation modeling*, *9*(4), 599-620.

Muthén, L. K., & Muthén, B. O. (2017a). Mplus. *Statistical analysis with latent variables. Version 8.*

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

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*(2), 463-469.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling*, *14*(4), 535-569.

Oliver, C., Berg, K., Moss, J., Arron, K., & Burbidge, C. (2011). Delineation of behavioral phenotypes in genetic syndromes: characteristics of autism spectrum disorder, affect and hyperactivity. *Journal of Autism and Developmental Disorders*, *41*(8), 1019-1032.

Olsen, M. K., & Schafer, J. L. (2001). A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association*, *96*(454), 730-745.

Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T., ... & Zwaigenbaum, L. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: a baby siblings research consortium study. *Journal of Child Psychology and Psychiatry*, *56*(9), 988-998.

Peugh, J., & Fan, X. (2012). How well does growth mixture modeling identify heterogeneous growth trajectories? A simulation study examining gmm's performance characteristics. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(2), 204-226.

Pickard, H., Rijsdijk, F., Happé, F., & Mandy, W. (2017). Are Social and Communication Difficulties a Risk Factor for the Development of Social Anxiety?. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(4), 344-351.

Plomin, R., Haworth, C. M., & Davis, O. S. (2009). Common disorders are quantitative traits. *Nature Reviews Genetics*, *10*(12), 872-878.

Portway, S. M., & Johnson, B. (2005). Do you know I have Asperger's syndrome? Risks of a non-obvious disability. *Health, Risk & Society*, *7*(1), 73-83.

Posserud, M. B., Lundervold, A. J., & Gillberg, C. (2006). Autistic features in a total population of 7–9‐year‐old children assessed by the ASSQ (Autism Spectrum Screening Questionnaire). *Journal of Child Psychology and Psychiatry*, *47*(2), 167-175.

Rai, D., Lewis, G., Lundberg, M., Araya, R., Svensson, A., Dalman, C., ... & Magnusson, C. (2012). Parental socioeconomic status and risk of offspring autism spectrum disorders in a Swedish population-based study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(5), 467-476.

Ram, N., & Grimm, K. J. (2009). Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International journal of behavioral development*, *33*(6), 565-576.

Robinson, E. B., Munir, K., Munafò, M. R., Hughes, M., McCormick, M. C., & Koenen, K. C. (2011a). Stability of autistic traits in the general population: further evidence for a

continuum of impairment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(4), 376-384.

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.

Robinson, E. B., St Pourcain, B., Anttila, V., Kosmicki, J. A., Bulik-Sullivan, B., Grove, J., ... & Martin, J. (2016). Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nature genetics*, *48*(5), 552-555.

Ruiz Calzada, L., Pistrang, N., & Mandy, W. (2012). High-functioning autism and Asperger's disorder: utility and meaning for families. *Journal of autism and developmental disorders*, *42*(2), 230.

Russell, G., & Norwich, B. (2012). Dilemmas, diagnosis and de-stigmatization: Parental perspectives on the diagnosis of autism spectrum disorders. *Clinical Child Psychology and Psychiatry*, *17*(2), 229-245.

Russell, G., Steer, C., & Golding, J. (2011). Social and demographic factors that influence the diagnosis of autistic spectrum disorders. *Social psychiatry and psychiatric epidemiology*, *46*(12), 1283-1293.

SAS: University Edition [Computer Software]. (2017). Retrieved from https://www.sas.com/en_gb/software/university-edition.html

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, *7*(2), 147.

Seltzer, M. M., Shattuck, P., Abbeduto, L., & Greenberg, J. S. (2004). Trajectory of development in adolescents and adults with autism. *Developmental Disabilities Research Reviews*, *10*(4), 234-247.

Skuse, D. H., Mandy, W. P., & Scourfield, J. (2005). Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. *The British Journal of Psychiatry*, *187*(6), 568-572.

Skuse, D. H., Mandy, W., Steer, C., Miller, L. L., Goodman, R., Lawrence, K., ... & Golding, J. (2009). Social communication competence and functional adaptation in a general population of children: preliminary evidence for sex-by-verbal IQ differential risk. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(2), 128-137.

Soke, G. N., Philofsky, A., Diguiseppi, C., Lezotte, D., Rogers, S., & Hepburn, S. (2011). Longitudinal changes in scores on the Autism Diagnostic Interview—Revised (ADI-R) in pre-school children with autism: Implications for diagnostic classification and symptom stability. *Autism*, *15*(5), 545-562.

Spiker, D., Lotspeich, L. J., Dimiceli, S., Myers, R. M., & Risch, N. (2002). Behavioral phenotypic variation in autism multiplex families: evidence for a continuous severity gradient. *American Journal of Medical Genetics Part A*, *114*(2), 129-136.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate behavioral research*, *25*(2), 173-180.

St. Pourcain, B., Wang, K., Glessner, J. T., Golding, J., Steer, C., Ring, S. M., ... & Davey Smith, G. (2010). Association between a high-risk autism locus on 5p14 and social communication spectrum phenotypes in the general population. *American Journal of Psychiatry*, *167*(11), 1364-1372.

St Pourcain, B., Skuse, D. H., Mandy, W. P., Wang, K., Hakonarson, H., Timpson, N. J., ... & Golding, J. (2014). Variability in the common genetic architecture of social-communication spectrum phenotypes during childhood and adolescence. *Molecular autism*, *5*(1), 18.

Szatmari, P., Bryson, S., Duku, E., Vaccarella, L., Zwaigenbaum, L., Bennett, T., & Boyle, M. H. (2009). Similar developmental trajectories in autism and Asperger syndrome: from early childhood to adolescence. *Journal of Child Psychology and Psychiatry*, *50*(12), 1459-1467.

Szatmari, P., Georgiades, S., Duku, E., Bennett, T. A., Bryson, S., Fombonne, E., ... & Volden, J. (2015). Developmental trajectories of symptom severity and adaptive functioning in an inception cohort of preschool children with autism spectrum disorder. *JAMA psychiatry*, *72*(3), 276-283.

Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. *Advances in latent variable mixture models*, (Information Age Publishing, Inc), 317-341.

Venker, C. E., Ray-Subramanian, C. E., Bolt, D. M., & Weismer, S. E. (2014). Trajectories of autism severity in early childhood. *Journal of autism and developmental disorders*, *44*(3), 546.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. *Applied latent class analysis*, *11*, 89-106.

Volkmar, F. R., & Klin, A. (2005). Issues in the classification of autism and related conditions. *Handbook of Autism and Pervasive Developmental Disorders, Volume 1, Third Edition*, 5-41.

Wechsler, D., Golombok, S., & Rust, J. (1992). WISC-III UK. *Sidcup, Kent: The Psychological Corporation*.

Wickrama, K. K., Lee, T. K., O'Neal, C. W., & Lorenz, F. O. (2016). *Higher-order growth curves and mixture modeling with Mplus: A practical guide*. Routledge.

Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., ... & Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of autism and developmental disorders*, *45*(7), 1951.

Zimmerman, A. W. (Ed.). (2008). *Autism: Current theories and evidence*. Springer Science & Business Media.

Part 3. Critical Appraisal

1. Categories or dimensions?

This study was the first to use Latent Class Growth Models (LCGMs) to capture the different trajectories of change in Autistic traits (ATs) in a general population sample. It revealed that six latent trajectories characterised the "chronogeneity" (heterogeneity over time; Georgiades, Bishop, & Frazier, 2017) of general population participants. Of particular interest was the discovery that a later-increasing trajectory (10 to 16 years of age) was composed of a majority of female participants. This tallied with findings that girls receive later diagnoses (Giarelli et al., 2010), and lent support to the "adolescent emergence hypothesis" (Bargiela, Steward, & Mandy, 2016; Kopp & Gillberg, 2011), which suggests females with Autism may experience symptoms at a later stage than males. This could be because the symptoms emerge at this point, or because traits were present from childhood but were sufficiently "masked" by strategies of camouflaging and imitation (Dean, Harwood, & Kasari, 2016), so that it is only when social ecology becomes more complex during adolescence that social and communication difficulties become more apparent.

The established evidence that ATs are continuously distributed invites debate about categorical versus dimensional approaches to conceptualising mental health and neurodiversity (Silberman, 2015), and also reflection on the sometimes uncritical way in which sex and gender can themselves be "categorised". I intend to offer brief reflections on each of these diversities in turn. As Richard Bentall (2004) has argued:

> It seems reasonable to assume, as a general principle, that abnormal behaviours and experiences exist on continua with normal behaviours and experiences. This principle of continuity might be formally stated

> as follows: Abnormal behaviours and experiences are related to normal behaviours and experiences by continua of frequency… severity… and phenomenology (p.115).

Debate between categorical and dimensional approaches to conceptualising mental health and neurological difficulties has been protracted (Donner and Eliasziw, 1994), and evidence supporting dimensional approaches has been established for psychosis (van Os et al., 1999), Personality Disorder (Trull & Durrett, 2005), Social Anxiety Disorder (Ruscio, 2010) and depression (Haslam, 2003). In Autism Spectrum Disorders (ASD), a similar weight of evidence has now accumulated highlighting the value of dimensional approaches (Constantino, 2009). This importantly underscores the fact that ATs are a continuum, with no natural "cut-point" defining where "autistic" and "non-autistic" begin and end. However, as Kraemer, Noda and O'Hara have argued (2004), dimensional and categorical approaches can quite conceivably interdigitate in such a way that severity is assessed and carefully compared to need for clinical services, as well as (arbitrary) conventional cut-points. The value of thinking in both dimensional and categorical terms simultaneously is that diagnostic labels can facilitate service allocation to those in need, and self-advocacy and disability rights movements can help to increase public understanding and positive self-identity (Silberman, 2015). At the same time, clinical benefits of dimensional approaches include a sensitivity to those who may be diagnostically sub-clinical and yet still experience significant difficulty, as well as to those whose symptom profiles increase or decrease over time who may be missed in the current categorically-driven diagnostic climate. Moreover, the dimensional approach helps reduce bias against female participants who have been underrepresented in prior research (Loomes et al., 2017), which will help improve clinical recognition and service provision for those who may have been missed (Gould & Ashton-Smith, 2011). Further research into the relationship between AT

trajectories and such related areas as daily living skills, mood, behaviours and cognitive difficulties could also help further clarify where need for clinical services is most unmet.

The research process, which involved shifting perspective from categorical to dimensional ways of thinking, invites critical reflection on another category that has played a central role in this study – that of sex or gender. It is difficult, perhaps especially so in quantitative research, to avoid an uncomplicated approach in which biological sex and socioculturally defined gender are fused without reflection, and gender treated as a binary categorical variable (Butler, 2004). As Lorber has argued (1996):

> Research variables – 'sex' polarized as 'females' and 'males', 'sexuality' polarized as 'homosexuals' and 'heterosexuals', and 'gender' polarized as 'women' and 'men' – reflect unnuanced series that conventionalize bodies, sexuality and social location. Such research designs cannot include the experiences of hermaphrodites, pseudo-hermaphrodites, transsexuals, transvestites, bisexuals, third genders and gender rebels… Even if the research sample is restricted to putative 'normals', the use of unexamined categories of sex, sexuality and gender will miss complex combinations of status and identity, as well as differently gendered sexual continuities and discontinuities (p.144).

Lai, Lombardo, Auyeung, Chakrabarti and Baron-Cohen (2015), in their review of sex and gender differences in Autism, acknowledge this complexity and use the term "sex/gender" to designate an interaction of gendered socialisation and biological sex occurring across development. However, anecdotal elaborations of a

"female phenotype" help to build a broad, categorical stereotype of girls channelling their special interests into such subjects as animals, celebrities, pop music and fashion (Lai et al., 2015). To endorse such a binary "male phenotype" of a systematising brain, with interests in computers and mechanics (Baron-Cohen, 2010) and a "female phenotype" of a more socially turned brain with interests in fashion and celebrities, could help to reinforce socially-sanctioned gender roles uncritically. Highlighting alternative phenotypes of Autism is useful, but further research needs to complicate these simplistic gender binaries, and to consider using gender-neutral labels, bearing social norms in mind when interpreting increased propensity for children socialised into a particular gender to conform to any phenotypic phenomenology. Furthermore, increasingly evidence is suggesting that high levels of ATs is associated with a greater likelihood of not identifying with one's birth-assigned gender, or rejecting binary gender identities (Kristensen & Broome, 2015; Strang et al., 2014), and women with ASD diagnoses have reported conflict between their sense of gender identity and their diagnosis (Bargiela et al., 2016). This further underlines the importance of moving beyond simplistic models of binary sex/gender in future studies.

2. Dealing with dilemmas and methodological choices

An especially difficult juncture in the research project arose when fit indices, including the Bayesian Information Criterion (BIC) and Bootstrapped Lo-Mendell Rubin Test (BLRT), continued to improve until a large number of classes (eleven or twelve) were reached. I consulted two statisticians with expertise in Growth Mixture Modelling (GMM) regarding the continually improving fit indices. One expert noted that this is a frequent occurrence in GMM, especially on large data sets. Another suggested that the skewness and the kurtosis in the data may be responsible for this result, and hence I undertook two-part GMM and zero-inflated Poisson GMM

approaches to try to see if this would produce a more definitive preference for a model with a lower number of classes. These approaches in actuality replicated the pattern found in the original GMM. I could find no published material on handling the phenomenon of improving fit indices, and little published material recommending a basis for selecting optimal models aside from using fit indices. Further papers dedicated to selection of optimal GMM models, especially with a focus on the phenomenon of continually improving fit indices, would be extremely helpful to non-experts.

I also attempted to explore other avenues for modelling the data. Firstly, I designed a model using sex as a known class, which treats this exogenous variable as an additional latent variable, and performs a binary logistic regression of the categorical latent class on the unknown class. Additionally, I investigated separate GMMs between sexes, and modelled Social and Communication Disorders Checklist (SCDC; Skuse et al., 2005) scores as count data by summing the total number of items endorsed. None of these solved the problem of the increasing classes. I spent considerable amounts of time researching alternative approaches to modelling GMM data, including a two-part GMM (Olsen & Schafer, 2001; Asparouhov & Muthén, 2015). In this approach, as in the zero-inflated Poisson, floor effects are accommodated by modelling the data in a way that uses a categorical GMM to represent the probability of having zero or non-zero scores, and a continuous GMM to represent the continuous (non-zero) part of the data. It was surprising and somewhat disappointing that none of these more exotic approaches to accommodating the data appeared to address the underlying problem, if there was one. I briefly considered using entirely binary categorical approaches to modelling the data, for example using the SCDC cut-off of 8 (Robinson et al, 2011) to create a model representing probability of transitioning between “case” and “non-case” over time. However, I decided this this approach would result in too great a loss of

important information, compromising the richness of the data that was available. Furthermore, in the context of the SCDC having been instrumental in establishing that ATs are continuously-distributed across the population (Constantino, 2009) it seemed entirely inappropriate and at-odds with the underlying philosophy of the research, as well as the evidence, to revert to a categorical conceptualisation purely for convenience. In this sense, I felt I wanted to build on the substantive theory and evidence that I had reviewed prior to my analyses, even if this meant facing the statistical conundrum raised by the fit indices.

The dilemma of the improving fit indices highlighted a trade-off between statistical precision and interpretability. It appears to be the case, for example, that according to all conventional fit indices, a twelve-class GMM genuinely is the optimal representation of the heterogeneity present in the data. Part of the reason for twelve classes seeming to be an "error" is that it is not conventional in GMM to extract such a large number of classes and present them to a readership. Firstly it would be lengthy and potentially tedious to chart each and every one of these trajectories and their differences. The reason for the potential tedium is that classes become increasingly lacking in clinically-significant distinctiveness as twelve classes are approached; even if fit indices suggest an improvement, the relevance of a new trajectory that is parallel and one-point lower than an existing trajectory is hard to justify conceptually, especially when relation to external covariates is also identical.
The core dilemma therefore was: in the face of the continually improving fit indices, which model is the best to select? I opted for what I considered to be the most conservative approach. This was to select the highest number of classes that could be defended conceptually – it was at the seven-class model that a class first appeared that included less than one percent of the total sample, and it was at the six-class model that a clear "elbow" was reached in the 2-LogLikelihood statistics, which suggested that although the addition of subsequent classes remained

statistically-significant according to BLRT, they were also relatively small improvements.

However, I also anticipated a potential criticism of my approach, arguing that in fact choosing the *lowest* number of conceptually meaningful classes would be the conservative choice. As Bauer and Curran (2003) have argued in their trenchant critique of the application of GMM to psychological research, there is a risk of over-extracting spurious latent trajectory classes, due to the fact that GMM can be used both to locate "real" subgroups in data *and* to represent a non-normal distribution as normally-distributed subcomponents, but which lack real-life validity. In the face of this critique, it is important for researchers to be able to interpret the meaning of classes using substantive theory, to examine the classes carefully to ensure that the described trajectory is coherent and meaningful, and that differential relationships to external criteria can also be discerned. I therefore attempted to ensure that these tasks were achieved, by interpreting the classes in light of theory, and demonstrating relationships to IQ and Strengths and Difficulties Questionnaire (Goodman, 2001) scores. I also took the view that choosing a smaller number of classes for my final model would aid conceptual ease a little, but would require selecting an entirely arbitrary "cut-point", and losing faithfulness to the indication that a larger number of classes did offer significant statistical improvement in model fit.

3. Alternative approaches

If there had been the time available, it would have been interesting to try other growth modelling approaches to the data, such as Generalized Linear Latent And Mixed Models (GLLAMM; Rabe-Hesketh, Skrondal, & Pickles, 2004) using STATA (Statacorp, 2017), or non-parametric latent growth models (Nagin, 2005) to confirm that the same classes were extracted. Additionally, a range of alternative fit statistics

that were not available in MPlus such as the Normalized Entropy Criterion (NEC; Biernacki & Govaert, 1999; Celeux & Soromenho, 1996) and the Classification Likelihood Criterion (CLC; Biernaci & Govaert, 1997) produce outcomes in a different fashion from the BIC or BLRT prioritised in the study – instead, they prioritise models with well-separated groups, meaning class posterior probabilities approaching one or zero (McLachlan & Peel, 2000). This would have been interesting to examine, in case further evidence emerged regarding an optimal model with a lower number of classes. It would have been useful if possible to also have had the chance to model restricted and repetitive behaviours (RRB) and interests alongside the Social and Communication Disorders Checklist (Skuse et al., 2005), which focused only on social-communicative deficits.

4. Strengths and weaknesses

This study had many strengths, including being the first application of GMM to ATs in general population participants, its large sample size and the rigorous data-collection procedures employed by the Avon Longitudinal Study of Parents and Children (ALSPAC). It allowed participants' sex, IQ and SDQ scores to be compared across latent classes. The methodology is transparent and replicable, and guidelines for conducting and reporting latent class growth modelling were carefully followed (Ram & Grimm, 2009).

GMM approaches allowed for chronogeneity of participants to be explored, highlighting six different trajectories that could increase our understanding of improving and worsening phenotypes. In particular, evidence regarding those whose scores increased from a low score at ages 7 and 10 to a far more severe score by the age of 16 was a novel finding, suggesting that many of this late-developing symptom group may be missed in clinical settings due to their conventionally

atypical presentation. Previous studies that had restricted their recruitment to participants who already had ASD diagnoses are designed in a way that by definition has been unable to detect this group. Moreover, approaches that have used a single growth trajectory to represent change across time in all participants (Robinson et al., 2011) have also been unable to detect this cluster of individuals.

Limitations included the fact that the SCDC is restricted to assessing social and communication difficulties only, meaning that only one of the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM-5; American Psychiatric Association [APA], 2013) ASD criteria is under assessment. However, the SCDC has been well-validated (Skuse et al., 2005), demonstrating heritability (St Pourcain et al., 2014) and genetic risk that is consistent with diagnosed ASD (Robinson et al., 2016). Nonetheless, it would be useful and interesting to examine trajectories of RRB symptoms over time to investigate similarities and differences. Furthermore, measures of overall Autism symptomatology such as the Autism Diagnostic Interview-Revised (ADI-R; Le Couteur, Lord, & Rutter, 2003) could be use to model symptoms across both domains. Also, this study was restricted to examining scores between the ages of 7 and 16, and therefore it is not clear whether trajectories would continue as observed beyond mid-adolescence. Furthermore, a key limitation of this study was the fact that participants were overwhelmingly White (98%) and participants with higher socioeconomic status were more likely to participate, which potentially limits generalisability.

A further, more conceptual limitation is that modelling ATs in terms of social difficulties reinforces an exclusively deficit-based conceptualisation of Autism. The emerging field of Critical Autism Studies (Davidson & Orsini, 2013) has usefully drawn attention to the way that the predominant, deficit-focused construction of Autism can function as a less enabling narrative than one that focuses on abilities

(O'Dell, Bertilsdotter Rosqvist, Ortega, Brownlow, & Orsini, 2016). Neurodiversity narratives have also highlighted Autism as a positive difference rather than a flaw or lack, which has been successful in promoting rights and gaining political traction (Ortega, 2013). Focusing exclusively on deficits could implicitly reinforce the opposite of these approaches, and promote the idea that people with ASD are somehow "less than". However, the dimensional approach taken in this study also helps to remove any artificial barrier between people "with" or "without" ASD, promoting instead a conceptualisation of diversity across a range. If Critical Autism Studies aims to deconstruct the assumption of a "non-autistic 'norm'" (O'Dell et al., 2016, p. 168), then this study hopefully supports this agenda.

5. Clinical and scientific implications

Scientific implications of this study included the novel finding that around 30% of the general population follow increasing or decreasing trajectories of Autism traits over time. Particularly striking observations included the finding that around half of people with very severe AT scores at age 7 improved dramatically, showing vastly reduced scores by age 16, and that a group with later-increasing symptoms was observed that included a majority of girls. This lends support to the "adolescent emergence hypothesis" of Autism in girls, suggesting that either there is an onset of symptoms in adolescence, or that traits that have been present since childhood become apparent in the more complex social environment of later development.

It will be important for clinicians to note that improving and worsening trajectories for ATs have been observed, and in particular for clinical practice to accommodate the possibility of adolescent emergence of ATs, especially amongst girls. Diagnostic practice should change to recognise this and to ensure that these young people are provided with appropriate support and services.

6. Directions for future research

Future research should seek to conduct further LCGMs on general population participants, replicating this approach on other data sets and using additional measures, including those that capture RRBs and full ASD profiles. The addition of measurement that captures key strengths and abilities will also be a vital addition to future mapping of chronogeneity. It will be useful for future studies also to focus on the later-developing trajectory, to examine phenotypic characteristics of this group – this will be crucially helpful in potentially developing more ecologically-valid measures that capture different presentations of ASD if this is the case. Moreover, characterising improving trajectories in terms of relevant individual and contextual variables could help to inform development of future individualised treatment approaches (Almirall & Chronos-Tuscano, 2016; Georgiades et al., 2017). It will also be helpful for future studies to extend participation periods beyond mid-adolescence, to determine whether the identified trajectories extend into adulthood.

The observation of latent groups with distinct developmental trajectories would potentially offer a new direction for genetic research. Polygenic Risk Score (PRS) approaches aggregate the small effects of thousands of genetic variations (Plomin, Haworth, & Davis, 2009). PRS approaches are especially useful when conceptualising disorders as quantitative traits across the whole population (Plomin et al., 2009) and hence could be a very fruitful avenue for exploration in conjunction with latent trajectories of ASD traits – for example, it could be the case that there is a relationship between genetic profile and developmental trajectories. So far, genetic studies have yet to investigate the linkage between genes and phenotypes that are longitudinal in nature – that is to say, to investigate genetic variation alongside chronogeneity.

Finally, it will be important for future studies to investigate trajectories of ATs in general population datasets that are diverse and representative. A crucial next step will be to carry out such studies on global datasets, to replicate these findings and investigate whether these apply across cultures and ethnicities. Even more, this work could help highlight unmet health care needs around the world.

References

Almirall, D., & Chronis-Tuscano, A. (2016). Adaptive interventions in child and adolescent mental health. *Journal of Clinical Child & Adolescent Psychology*, *45*(4), 383-395.

Bargiela, S., Steward, R., & Mandy, W. (2016). The experiences of late-diagnosed women with autism spectrum conditions: an investigation of the female autism phenotype. *Journal of autism and developmental disorders*, *46*(10), 3281.

Baron-Cohen, S. (2010). Empathizing, systemizing, and the extreme male brain theory of autism. *Progress in brain research*, *186*, 167-175.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological methods*, *8*(3), 338.

Bentall, R. P. (2004). *Madness explained: Psychosis and human nature*. Penguin UK.

Biernacki, C., & Govaert, G. (1997). Using the classification likelihood to choose the number of clusters. *Computing Science and Statistics*, 451-457.

Biernacki, C., Celeux, G., & Govaert, G. (1999). An improvement of the NEC criterion for assessing the number of clusters in a mixture model. *Pattern Recognition Letters*, *20*(3), 267-272.

Butler, J. (2004). *Undoing gender*. Psychology Press.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., ... & Moffitt, T. E. (2014). The p factor: one general psychopathology factor in the structure of psychiatric disorders?. *Clinical Psychological Science*, *2*(2), 119-137.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of classification*, *13*(2), 195-212.

Constantino, J. N. (2009). How continua converge in nature: cognition, social competence, and autistic syndromes. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*(2), 97-98.

Davidson, J., & Orsini, M. (2013). *Worlds of autism: Across the spectrum of neurological difference*. University of Minnesota Press.

Dean, M., Harwood, R., & Kasari, C. (2016). The art of camouflage: Gender differences in the social behaviors of girls and boys with autism spectrum disorder. *Autism*, 1362361316671845.

Donner, A., & Eliasziw, M. (1994). Statistical implications of the choice between a dichotomous or continuous trait in studies of interobserver agreement. *Biometrics*, 550-555.

Georgiades, S., Bishop, S. L., & Frazier, T. (2017). Editorial Perspective: Longitudinal research in autism–introducing the concept of ‘chronogeneity’. *Journal of Child Psychology and Psychiatry*, *58*(5), 634-636.

Giarelli, E., Wiggins, L. D., Rice, C. E., Levy, S. E., Kirby, R. S., Pinto-Martin, J., & Mandell, D. (2010). Sex differences in the evaluation and diagnosis of autism spectrum disorders among children. *Disability and health journal*, *3*(2), 107-116.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), 1337-1345.

Gould, J., & Ashton-Smith, J. (2011). Missed diagnosis or misdiagnosis? Girls and women on the autism spectrum. *Good Autism Practice (GAP)*, *12*(1), 34-41.

Haslam, N. (2003). Categorical versus dimensional models of mental disorder: The taxometric evidence. *Australian and New Zealand Journal of Psychiatry*, *37*(6), 696-704.

Kopp, S., & Gillberg, C. (2011). The Autism Spectrum Screening Questionnaire (ASSQ)-Revised Extended Version (ASSQ-REV): an instrument for better capturing the autism phenotype in girls? A preliminary study involving 191 clinical cases and community controls. *Research in developmental disabilities*, *32*(6), 2875-2888.

Kraemer, H. C., Noda, A., & O'Hara, R. (2004). Categorical versus dimensional approaches to diagnosis: methodological challenges. *Journal of Psychiatric Research*, *38*(1), 17-25.

Kristensen, Z. E., & Broome, M. R. (2015). Autistic traits in an internet sample of gender variant UK adults. *International Journal of Transgenderism*, *16*(4), 234-245.

Lai, M. C., Lombardo, M. V., Auyeung, B., Chakrabarti, B., & Baron-Cohen, S. (2015). Sex/gender differences and autism: setting the scene for future research. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(1), 11-24.

Le Couteur, A., Lord, C., & Rutter, M. (2003). The autism diagnostic interview-revised (ADI-R). *Los Angeles, CA: Western Psychological Services*.

Lorber, J. (1996). Beyond the binaries: Depolarizing the categories of sex, sexuality, and gender. *Sociological Inquiry*, *66*(2), 143-160.

McLachlan, G. J., & Peel, D. (2000). Robust mixture modelling using the t distribution. *Statistics and Computing*, *10*(4), 335-344.

Muthén, B., & Asparouhov, T. (2015). Growth mixture modeling with non‐normal distributions. *Statistics in medicine*, *34*(6), 1041-1058.

Nagin, D. (2005). *Group-based modeling of development*. Harvard University Press.

O'Dell, L., Bertilsdotter Rosqvist, H., Ortega, F., Brownlow, C., & Orsini, M. (2016). Critical autism studies: exploring epistemic dialogues and intersections, challenging dominant understandings of autism. *Disability & Society*, *31*(2), 166-179.

Olsen, M. K., & Schafer, J. L. (2001). A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association*, *96*(454), 730-745.

Ortega, F. (2013). Cerebralizing autism within the neurodiversity movement. *Worlds of autism: Across the spectrum of neurological difference*, 73-96.

Plomin, R., Haworth, C. M., & Davis, O. S. (2009). Common disorders are quantitative traits. *Nature Reviews Genetics*, *10*(12), 872-878.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). GLLAMM manual.

Ram, N., & Grimm, K. J. (2009). Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International journal of behavioral development*, *33*(6), 565-576.

Robinson, E. B., Munir, K., Munafò, M. R., Hughes, M., McCormick, M. C., & Koenen, K. C. (2011). Stability of autistic traits in the general population: further evidence for a continuum of impairment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(4), 376-384.

Robinson, E. B., St Pourcain, B., Anttila, V., Kosmicki, J. A., Bulik-Sullivan, B., Grove, J., ... & Martin, J. (2016). Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nature genetics*, *48*(5), 552-555.

Ruscio, A. M. (2010). The latent structure of social anxiety disorder: Consequences of shifting to a dimensional diagnosis. *Journal of Abnormal Psychology*, *119*(4), 662.

Silberman, S. (2015). *Neurotribes: The legacy of autism and the future of neurodiversity.* London: Penguin

Skuse, D. H., Mandy, W. P., & Scourfield, J. (2005). Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. *The British Journal of Psychiatry*, *187*(6), 568-572.

Stata, S. (2017). Release 15. *Statacorp Statistical Software*.

St Pourcain, B., Skuse, D. H., Mandy, W. P., Wang, K., Hakonarson, H., Timpson, N. J., ... & Golding, J. (2014). Variability in the common genetic architecture of social-communication spectrum phenotypes during childhood and adolescence. *Molecular autism*, *5*(1), 18.

Strang, J. F., Kenworthy, L., Dominska, A., Sokoloff, J., Kenealy, L. E., Berl, M., ... & Luong-Tran, C. (2014). Increased gender variance in autism spectrum disorders and attention deficit hyperactivity disorder. *Archives of Sexual Behavior*, *43*(8), 1525.

Trull, T. J., & Durrett, C. A. (2005). Categorical and dimensional models of personality disorder. *Annu. Rev. Clin. Psychol.*, *1*, 355-380.

van Os, J., Gilvarry, C., Bale, R., Van Horn, E., Tattan, T., & White, I. (1999). A comparison of the utility of dimensional and categorical representations of psychosis. *Psychological medicine*, *29*(3), 595-606.

Appendix 1: Descriptive Statistics

Appendix 1.1.

Participant demographic information

	N	%
Enrolment Phase		
Phase I	9354	96%
Phase II	327	3.4%
Phase III	63	.6%
Sex		
Female	4784	49.1%
Male	4960	50.9%
Maternal Education		
CSE	1372	15.2%
Vocational	815	9.0%
O-Level	3193	35.4%
A-Level	2264	25.1%
Degree	1372	15.2%
Maternal Occupation		
Professional	517	6.7%
Managerial and technical	2562	33.4%
Skilled non-manual	3272	42.6%
Skilled manual	534	7.0%
Partly-skilled manual	650	8.5%
Non-skilled manual	134	1.7%
Armed forces	3	<.001%
Accommodation		
Mortgaged / owned	7293	79.5%
Rented (private)	528	5.8%
Rented (council / Housing Association)	985	10.9%
Other	257	2.8%
Ethnicity		
White	8810	98.0%
Black-Caribbean	38	.4%
Black-African	7	1%
Black-Other	24	.3%
Indian	30	.3%
Pakistani	9	.1%

Bangladeshi	1	<.001%
Chinese	18	.2%
Other	52	.6%

Appendix 1.2.

Descriptive statistics for the four study time points

Time point	Statistic	Value	Standard Error
1	Mean: overall (SD)	2.83 (3.72)	.04
	Mean: male (SD)	3.25 (4.15)	.064
	Mean: female (SD)	2.39 (3.14)	.050
	Median	2	
	Variance	13.72	
	Minimum	0	
	Maximum	24	
	Range	24	
	Interquartile range	4	
	Skewness	2.16	.027
	Kurtosis	5.91	.054
2	Mean: overall (SD)	2.40 (3.63)	.041
	Mean: male (SD)	2.78 (4.09)	.065
	Mean: female (SD)	2.01 (3.05)	.049
	Median	1	
	Variance	13.23	
	Minimum	0	
	Maximum	24	
	Range	24	
	Interquartile range	3	
	Skewness	2.46	.028
	Kurtosis	7.51	.056
3	Mean: overall (SD)	2.54 (3.64)	.043
	Mean: male (SD)	2.74 (3.90)	.065
	Mean: female (SD)	2.35 (3.34)	.056
	Median	1	
	Variance	13.28	
	Minimum	0	

	Maximum	24	
	Range	24	
	Interquartile range	4	
	Skewness	2.20	.029
	Kurtosis	6.00	.058
4	Mean: overall (SD)	2.85 (3.80)	.050
	Mean: male (SD)	2.80 (3.83)	.073
	Mean: female (SD)	2.89 (3.77)	.069
	Median	1	
	Variance	14.44	
	Minimum	0	
	Maximum	24	
	Range	24	
	Interquartile range	4	
	Skewness	1.90	.032
	Kurtosis	4.08	.065

Appendix 1.3.

Kolmogorov-Smirnov Tests, with Lilliefors Significance Correction applied

Time point	Statistic	Df	Sig.
1	.223	8115	<0.001
2	.254	7727	<0.001
3	.242	7010	<0.001
4	.226	5681	<0.001

Appendix 2: Example syntax

Syntax specifying single-class growth model

```
Variable:   Names are t1 t2 t3 t4;
            Usevar = t1-t4;
            Missing = all (999);
Model:      i s q| t1@0 t2@1 t3@2 t4@3;
```

Syntax specifying two-class LCGA

```
Variable:   Names are t1 t2 t3 t4;
            Usevar = t1-t4;
            Missing = all (999);
            Classes = c(2);
Analysis:   Type = Mixture Missing;
            Starts = 2000 200;
            Stiterations = 400;
Model:      %OVERALL%
            i s q | t1@0 t2@1 t3@2 t4@3;
            i-q@0;
            t1 t2 t3 t4 (resfix);
```

Syntax specifying two-class GMMs

Fully invariant

```
Variable:   Names are t1 t2 t3 t4;
            Usevar = t1-t4;
            Missing = all (999);
            Classes = c(2);
Analysis:   Type = Mixture missing;
            Starts = 2000 200;
            Stiterations = 400;
Model:      %OVERALL%
            i s q| t1@0 t2@1 t3@2 t4@3;

            %c#1%
            [i s q];
            i  (ivar);
            s  (svar);
            q  (qvar);
            i with s (covis);
            i with q (coviq);
            q with s (covqs);
            t1 t2 t3 t4 (resfix);

            %c#2%
            [i s q];
            i  (ivar);
            s  (svar);
            q  (qvar);
            i with s (covis);
            i with q (coviq);
            q with s (covqs);
            t1 t2 t3 t4 (resfix);
```

Fully variant

```
Variable:   Names are t1 t2 t3 t4;
            Usevar = t1-t4;
            Missing = all (999);
            Classes = c(2);
Analysis:   Type = Mixture missing;
            Starts = 2000 200;
            Stiterations = 400;
Model:      %OVERALL%
            i s q| t1@0 t2@1 t3@2 t4@3;

            %c#1%
            [i s q];
            i  (c1_ivar);
            s  (c1_svar);
            q  (c1_qvar);
            i with s (c1_covis);
            i with q (c1_coviq);
            q with s (c1_covqs);
            t1 t2 t3 t4 (c1_resfix);

            %c#2%
            [i s q];
            i  (c2_ivar);
            s  (c2_svar);
            q  (c2_qvar);
            i with s (c2_covis);
            i with q (c2_coviq);
            q with s (c2_covqs);
            t1 t2 t3 t4 (c2_resfix);
```

Mixed

```
Variable:   Names are t1 t2 t3 t4;
            Usevar = t1-t4;
            Missing = all (999);
            Classes = c(2);
Analysis:   Type = Mixture missing;
            Starts = 2000 200;
            Stiterations = 400;
Model:      %OVERALL%
            i s q| t1@0 t2@1 t3@2 t4@3;

            %c#1%
            [i s q];
            i  (c1_ivar);
            s  (c1_svar);
            q  (c1_qvar);
            i with s (c1_covis);
            i with q (c1_coviq);
            q with s (c1_covqs);
            t1 t2 t3 t4 (resfix);

            %c#2%
            [i s q];
            i  (c2_ivar);
            s  (c2_svar);
            q  (c2_qvar);
            i with s (c2_covis);
            i with q (c2_coviq);
            q with s (c2_covqs);
            t1 t2 t3 t4 (resfix);
```

Appendix 3: Measures

Social and Communication Disorders Checklist

(Skuse, Mandy and Scourfield, 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.

	Not true	Quite or sometimes true	Very or often true
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.)

Strengths and Difficulties Questionnaire

(Goodman, 2001)

Strengths and Difficulties Questionnaire

For each item, please mark the box for Not True, Somewhat True or Certainly True. It would help us if you answered all items as best you can even if you are not absolutely certain or the item seems daft! Please give your answers on the basis of the child's behaviour over the last six months or this school year.

Child's Name ..... Male/Female

Date of Birth.....

	Not True	Somewhat True	Certainly True
Considerate of other people's feelings	☐	☐	☐
Restless, overactive, cannot stay still for long	☐	☐	☐
Often complains of headaches, stomach-aches or sickness	☐	☐	☐
Shares readily with other children (treats, toys, pencils etc.)	☐	☐	☐
Often has temper tantrums or hot tempers	☐	☐	☐
Rather solitary, tends to play alone	☐	☐	☐
Generally obedient, usually does what adults request	☐	☐	☐
Many worries, often seems worried	☐	☐	☐
Helpful if someone is hurt, upset or feeling ill	☐	☐	☐
Constantly fidgeting or squirming	☐	☐	☐
Has at least one good friend	☐	☐	☐
Often fights with other children or bullies them	☐	☐	☐
Often unhappy, down-hearted or tearful	☐	☐	☐
Generally liked by other children	☐	☐	☐
Easily distracted, concentration wanders	☐	☐	☐
Nervous or clingy in new situations, easily loses confidence	☐	☐	☐
Kind to younger children	☐	☐	☐
Often lies or cheats	☐	☐	☐
Picked on or bullied by other children	☐	☐	☐
Often volunteers to help others (parents, teachers, other children)	☐	☐	☐
Thinks things out before acting	☐	☐	☐
Steals from home, school or elsewhere	☐	☐	☐
Gets on better with adults than with other children	☐	☐	☐
Many fears, easily scared	☐	☐	☐
Sees tasks through to the end, good attention span	☐	☐	☐

Signature ..... Date .....

Parent/Teacher/Other (please specify:)

Thank you very much for your help

Appendix 4: Model fit statistics

Appendix 4: Fit statistics

Appendix 4.1.

Fit statistics for core models

Models	2 Classes	3 Classes	4 Classes	5 Classes	6 Classes	7 Classes
LCGA (Quadratic)						
LL (No. of parameters)	-72370.80 (11)	-70496.06 (15)	-69838.53 (19)	-69236.46 (23)	-68716.11 (27)	-68401.35 (31)
AIC	144763.60	141022.13	139715.07	138518.92	137486.22	136864.71
BIC	144842.62	141129.90	139851.57	138684.16	137680.19	137087.43
SSABIC	144842.62	141082.23	139791.19	138611.07	137594.39	136988.91
Entropy	.94	.91	.88	.87	.86	.85
ALMR (*p*)	10553.63 (<0.0001)	3650.10 (.001)	1280.21 (.15)	1172.24 (.72)	1013.12 (.23)	612.82 (.18)
BLRT (*p*)	10840.90 (<0.0001)	3749.46 (<0.0001)	1315.06 (<0.0001)	1204.15 (<0.0001)	1040.70 (<0.0001)	629.50 (<0.0001)
Group size (%)						
C1	1021 (10.5%)	1398 (14.35%)	323 (3.32%)	7564 (77.63%)	137 (1.41%)	728 (7.47%)
C2	8723 (89.5%)	347 (3.56%)	7810 (80.15%)	376 (3.86%)	252 (2.59%)	592 (6.08%)
C3		7999 (82.09%)	784 (8.05%)	162 (1.66%)	623 (6.39%)	7280 (74.71%)

C4			827 (8.49%)	718 (7.37%)	7357 (75.55%)	220 (2.26%)
C5				924 (9.48%)	1115 (11.44%)	529 (5.43%)
C6					260 (2.67%)	135 (1.39%)
C7						260 (2.67%)

GMM-FI (quadratic free)		
LL (No. of parameters)	-70887.97 (14)	Error [1]
AIC	141803.95	
BIC	141904.53	
SSABIC	141860.04	
Entropy	.94	
ALMR (*p*)	3864.89 (.001)	
BLRT (*p*)	3970.10 (<.0001)	
Group size (%)		
C1	9053 (92.9%)	
C2	691 (9.1%)	

GMM-FV (quadratic free)		
LL (No. of parameters)	-63426.81 (21)	Error [2, 4, 5]
AIC	126895.63	
BIC	127046.50	
SSABIC	126979.76	
Entropy	.85	
ALMR (*p*)	18707.25	
	(<.0001)	
BLRT (*p*)	18892.42	
	(<.0001)	
Group size (%)		
C1	3596 (36.9%)	
C2	6148 (63.1%)	
GMM-Mixed (quadratic free)	Error [1]	
GMM-FI (quadratic constrained)		

LL (No. of parameters)						
AIC	141904.38	140293.16	138893.35	137828.65	**137161.64**	136571.64
BIC	141983.40	140400.92	139029.85	137993.89	**137355.62**	136794.36
SSABIC	141948.45	140353.25	138969.47	137920.80	**137269.82**	136695.84
Entropy	.95	.89	.88	.88	**.87**	.86
ALMR (*p*)	3847.14 (<.0001)	1576.31 (.12)	1370.50 (.48)	1044.25 (.01)	**657.122 (.63)**	584.144 (.26)
BLRT (*p*)	3951.86 (<.0001)	1619.22 (<.0001)	1407.80 (<.0001)	1072.70 (<.0001)	**675.01 (<.0001)**	600.04 (<.0001)
Group size (%)						
C1	9078 (93.2%)	642 (6.6%)	687 (7.1%)	606 (6.2%)	**168 (1.7%)**	584 (6.0%)
C2	666 (6.8%)	8623 (88.5%)	423 (4.3%)	167 (1.7%)	**475 (4.9%)**	204 (2.1%)
C3		479 (4.9%)	283 (2.9%)	352 (3.6%)	**781 (8.0%)**	590 (6.1%)
C4			8351 (85.7%)	8100 (83.1%)	**7763 (79.6%)**	92 (.9%)
C5				519 (5.3%)	**400 (4.1%)**	7551 (77.5%)
C6					**157 (1.6%)**	519 (5.3%)

C7			204 (2.0%)
GMM-FV (quadratic constrained)			
LL (No. of parameters)	-63737.62 (15)	Error [2, 3, 4]	
AIC	127505.24		
BIC	127613.00		
SSABIC	127565.34		
Entropy	.89		
ALMR (*p*)	18112.49 (<.0001)		
BLRT (*p*)	18359.00 (<.0001)		
Group size (%)			
C1	7039 (72.2%)		
C2	2705 (27.8%)		
GMM-Mixed (quadratic constrained)	Error [1]		

Key: Errors in model identification: 1– Psi-Matrix not positive definite, 2 – Ill-conditioned Fisher information matrix, 3 – Fisher information matrix not positive definite, 4 – Standard errors could not be computed, 5 - First-order derivative product matrix not positive definite

Appendix 4.2.

Fit statistics for additional models

Models	2 Classes	3 Classes	4 Classes	5 Classes	6 Classes	7 Classes
GMM (Known Class)						
LL (No. of parameters)	-77915.54 (16)	-76884.89 (22)	-76205.63 (28)	-75871.16 (34)	-75521.39 (40)	-75314.63 (46)
AIC	155863.08	153813.78	152467.25	151810.33	151122.79	150721.27
BIC	155978.03	153971.84	152668.42	152054.60	151410.17	151051.75
SSABIC	155927.18	153901.93	152579.44	151946.55	151283.05	150905.57
GMM (Females only)						
LL (No. of parameters)	-20918.31 (11)	-20675.05 (15)	-20503.21 (19)	-20339.66 (23)	-20235.53 (27)	-20147.91 (31)
AIC	41858.63	41380.11	41044.42	40725.32	40525.07	40357.82

BIC	41921.36	41465.65	41152.77	40856.48	40679.04	40534.60
SSABIC	41886.41	41417.99	41092.41	40783.41	40593.26	40436.11
GMM (Males only)						
LL (No. of parameters)	-21178.17 (11)	-20906.50 (15)	-20701.49 (19)	-20564.19 (23)	Error[1]	
AIC	42378.34	41843.00	41440.98	41174.38		
BIC	42440.83	41928.21	41548.91	41205.04		
SSABIC	42405.88	41880.55	41488.54	41231.96		
GMM (Count Data)						
LL (No. of parameters)	-31908.70 (7)	-30249.37 (11)	-29671.76 (15)	-29322.18 (19)	-29120.52 (23)	-29051.38 (29)
AIC	63831.41	60520.75	59373.53	58682.37	58287.04	58156.76
BIC	63875.56	60590.12	59468.13	58802.20	58432.10	58327.05
SSABIC	63853.32	60555.17	59420.47	58741.83	58359.02	58241.25
Entropy	.90	.85	.80	.77	.75	.73
GMM - Zero Inflated Poisson						
LL (No. of parameters)	-41909.77 (9)	-40855.970 (13)	-40468.391 (17)	-40197.638 (21)	-40037.226 (25)	-39978.632 (29)

AIC	83837.54	81737.939	80970.781	80437.276	80124.451	80015.265
BIC	83896.75	81823.460	81082.615	80575.424	80288.913	80206.041
SSABIC	83868.15	81782.150	81028.595	80508.693	80209.471	80113.888
Entropy	.70	.72	.68	.65	.65	.64
ALMR (*p*)	6303.14	2047.92	753.20	526.17	311.74	113.86
	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(0.10)
BLRT, 2-LL (*p*)	6486.84	2107.61	775.15	541.50	320.83	117.18
	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)

Key: Errors in model identification. 1 Psi-Matrix not positive definite

Appendix 5: Further statistics for final GMM model

Appendix 5.1.

Average class posterior probabilities for most likely class membership (row) by latent class (column)

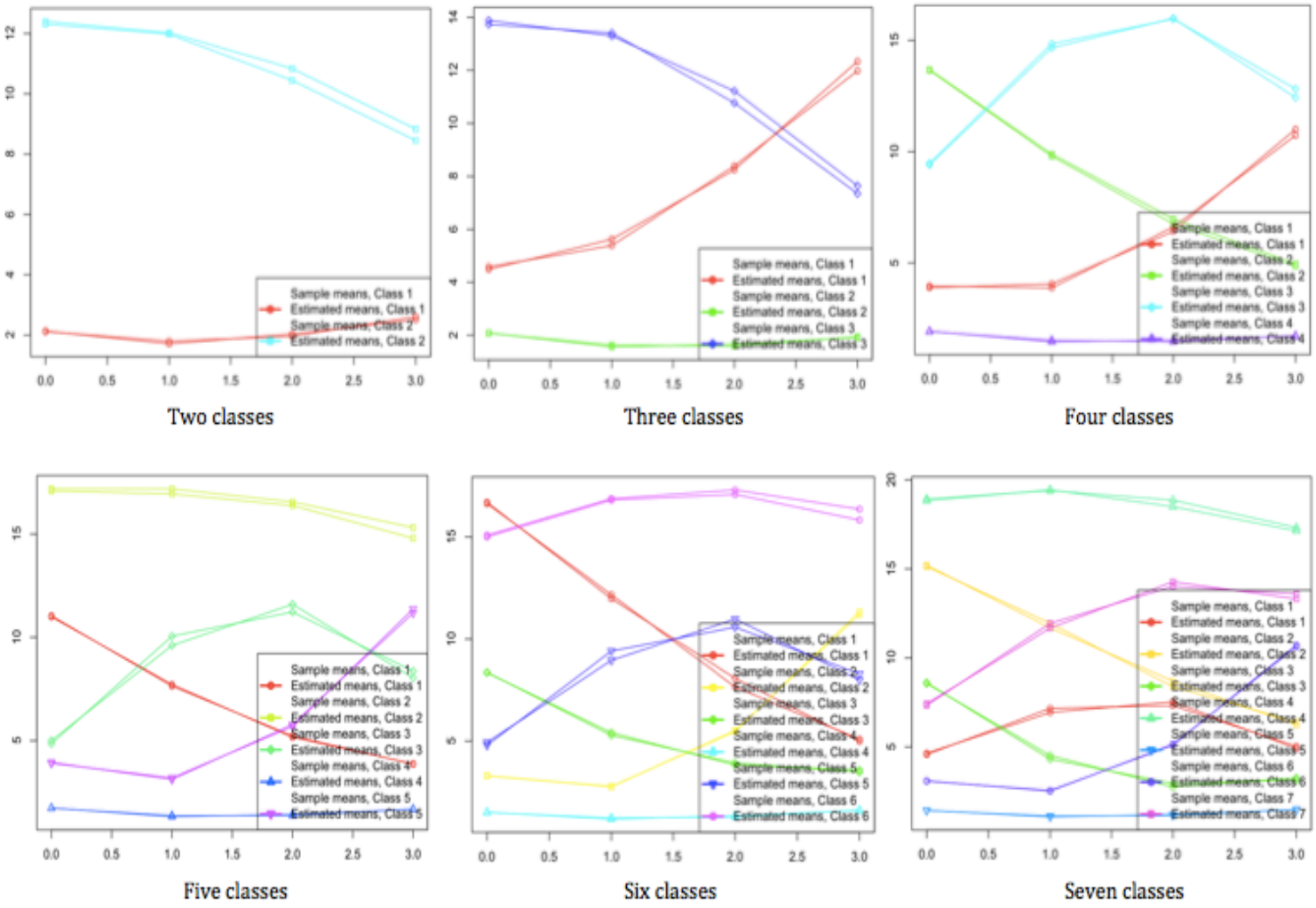
Class	1 (High Decrease)	2 (Delayed Increase)	3 (Moderate Decrease)	4 (Persistent Low)	5 (Early Increase)	6 (Persistent High)
1	**.866**	.000	.048	.003	.024	.058
2	.004	**.814**	.028	.103	.050	.001
3	.015	.032	**.791**	.109	.048	.005
4	.000	.030	.022	**.941**	.006	.000
5	.028	.058	.057	.036	**.809**	.012
6	.066	.004	.020	.000	.026	**.884**

Appendix 5.2.

Variances for the six-class GMM

Variances	Variance	Standard Error	Two-tailed significance
Intercept	0.68	.42	.10
Slope	0.04	.08	.59
Quadratic	0.00	.00	Not applicable
Residual variances	0.39	.16	.00

Appendix 6: Graphs for final GMM model



Two classes

Three classes

Four classes

Five classes

Six classes

Seven classes