Title Page

Title: The Longitudinal Heterogeneity of Autistic Traits: A Systematic Review

Author names and affiliations:

1. Dr Richard **Pender,** DClinPsy (**Corresponding author)**

Dr Richard Pender is NIHR/BRC Clinical Research Fellow at the Division of Psychiatry, University College London. He can be contacted at: Maple House, 149 Tottenham Court Road, London W1T 7BN, United Kingdom | e-mail: Richard.Pender.14@ucl.ac.uk | Telephone: +44 (0)2076791675 | Cellular: +44 (0)7730594549

2. Professor Pasco **Fearon,** PhD DClinPsy

Professor Pasco Fearon is Chair in Developmental Psychopathology in the Research Department of Clinical, Educational and Health Psychology at University College London. He can be contacted at: 1-19 Torrington Place, London WC1E 7HB, United Kingdom | e-mail: P.Fearon@ucl.ac.uk | Telephone: +44 (0)2076791675

3. Dr Jon **Heron,** DPhil

Dr Jon Heron is a Senior Research Fellow at Bristol Medical School's Department of Population Health Sciences. He can be contacted at: Bristol Medical School, University of Bristol, Population Health Sciences, Oakfield House, Clifton BS8 2BN, United Kingdom | e-mail: Jon.Heron@Bristol.ac.uk | Telephone: +44 (0)1173310093

4. Dr William **Mandy,** DClinPsy PhD

Dr William Mandy is an Associate Professor and Research Director in the Research Department of Clinical, Educational and Health Psychology at University College London. He can be contacted at: 1-19 Torrington Place, London WC1E 7HB, United Kingdom | e-mail: W.Mandy@ucl.ac.uk | Telephone: +44 (0)2076791675

Abstract and Keywords

Background: Previous reviews have characterised the mean stability of autistic traits (ATs) across samples on a single measure. However, no review has yet assessed mean change across a range of measures, or described the longitudinal heterogeneity of ATs, ie. variation in direction and degree of change.

Method: A systematic literature review was conducted using PubMed, PsycINFO and EMBASE up to May 31 2020. Forty-four studies meeting inclusion criteria were identified.

Results: Retrieved studies ranged from N=20 to N=9,744. Ages spanned one to 15 years at baseline and two to 23 years at follow-up. Female participants ranged from 0 to 51%. There is some evidence that overall ATs tend to reduce over time for autistic children, reflecting decreases in social communication difficulties but not restricted behaviours. This effect was strongest in clinical samples and using parent-report measures. However, there was good evidence that statistics of mean change obscure reliable variation. Decreasing ATs appear linked to higher verbal and non-verbal IQ and female gender in autistic participants. Four patterns of change: increasing, decreasing and stable high and low best characterised the data. Social and non-social traits showed distinct, separable profiles longitudinally.

Conclusions: Individuals experience diverse patterns of change over time. More general population studies are needed to reduce male bias. More work is needed to characterise the relationship between trajectories and well-being, functioning and quality of life outcomes. This will help to understand factors that promote resilience and reduce risk, and therefore to improve the timing and targets of intervention.

Keywords:

autism; autism phenotype; autism spectrum; chronogeneity; developmental trajectories; heterogeneity; latent

class growth model; neurodevelopment

Introduction

Autism Spectrum Disorder (hereafter "autism") is a neurodevelopmental condition defined by difficulties with social communication and interaction, and restricted and repetitive interests and behaviour (APA, 2013). As well as being a categorical diagnosis, a growing body of literature has demonstrated that autism can be conceptualised as a continuum of traits that are distributed throughout the general population (Constantino, 2009; Posserud, Lundervold, & Gillberg, 2006; Skuse et al., 2009). Measuring continuous autistic traits (ATs) naturalistically over time confers the distinct advantage that we can track fine-grained trajectories of ATs, charting their developmental course, their stability and predictors of change.

Our current knowledge of the developmental course of ATs suggests they appear to show little substantive change over time (Bieleninik et al., 2017). In a recent, high-quality systematic review, Bieleninik and colleagues found across studies there was a small reduction in autistic social traits and stability in restricted and repetitive behaviours (RRB). However, this review was limited to only one measure, the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1999). Additionally, this review examined only mean-level change – that is to say, a single statistic of change across an entire sample. Therefore, it was not designed to examine the presence of different patterns of change between participants, i.e. subgroups of individuals showing meaningful change over time, despite the mean trend for stability.

Scholars have recently called for more work to characterise the "chronogeneity" of autism – ie. its heterogeneity over time ("chrono") (Georgiades, Bishop and Frasier, 2017). This reflects the fact that some individuals show ATs more clearly later in development than others (Ozonoff et al., 2018), while others show markedly attenuated ATs over time (Orinstein et al., 2015). An additional literature is available, which did not fall within the scope of Bieleninik and colleagues' review, and which describes the chronogeneity of change in ATs. For example, Lord, Luyster, Guthrie and Pickles (2012) explored longitudinal trajectories of ATs in toddlers using latent class growth modelling (LCGM), a statistical approach that allows relatively homogeneous subgroups to emerge from heterogeneous data. This approach revealed that within the data, around 20% of participants showed each of significantly increasing or decreasing ATs, while 40% showed stable low and 20% stable high ATs over time. A single statistic of change across the sample would have obscured this variation.

Understanding the chronogeneity of ATs is important for several reasons. Firstly, there is an overdue but growing attention to resilience in autism research (Szatmari, 2018). Resilience has been defined as a good outcome, despite exposure to developmental risk (Rutter, 2012). Characterising the chronogeneity of ATs will greatly help to understand critical developmental periods and key variables that reduce risk and promote resilience (Georgiades & Kasari, 2018). Secondly, it will help develop a more fine-grained and accurate representation of behavioural phenotypes, which can in turn improve prognostic prediction and timing of intervention (Almirall & Chronis-Tuscano, 2016; Georgiades et al., 2017). Thirdly, our understanding of the genetic aetiology of autism could also be greatly improved by accounting for chronogeneity (Chaste et al., 2015).

Chronogeneity is a multi-faceted idea and for clarity can be broken down into three components, all three of which may be present in the development of ATs: *intra-individual, inter-individual, and inter-group*. *Intra-individual* (within-person) chronogeneity would be reflected by different domains of autistic characteristics changing differently over time for an individual. For example, cross-sectional evidence has shown that the social and RRB domains of autism are fractionable (Happé, Ronald, & Plomin, 2006; Mandy & Skuse, 2008; Ronald, Happé, & Plomin, 2005) and intra-individual chronogeneity would be supported by evidence that they are also fractionable longitudinally. *Inter-individual* (between-person) chronogeneity would be evidenced by substantively different developmental trajectories over time between participants. For example, some individuals may show significant increases in ATs, and some significant decreases. *Inter-group* chronogeneity would be supported by different groups showing different rates of change, such as female participants showing later increases in ATs than male participants (Mandy, Pellicano, St Pourcain, Skuse, & Heron, 2018).

It is important for studies charting AT trajectories to include general population participants as well as those with a clinical autism diagnosis, to represent the broader AT continuum. Across the continuum of AT severity, there is evidence for chronogeneity – for example, in a recent study on a large general population dataset, Mandy and colleagues (2018) found shifting levels of ATs between 7 and 16 years, with a greater tendency for within-person change for girls compared to boys. The first advantage of such approaches is that they can provide information about the female autism phenotype (a type of inter-group chronogeneity). It is established that females with high levels of ATs are less likely than equivalent males to be diagnosed with autism and may be underrepresented in clinical samples (Dworzynski, Ronald, Bolton, & Happé, 2012; Loomes, Hull, & Mandy, 2017; Russell, Steer, & Golding, 2011). ATs may also emerge later in adolescence for women (Bargiela, Steward & Mandy, 2016; Kopp & Gillberg, 2011), and purely diagnostic samples may fail to chart trajectories of change from lower to higher

ATs, representing unmet clinical need. Secondly, this can provide insights into diverse forms of clinical need. Subclinical ATs are an important risk factor for difficulties including behavioural and emotional problems (Saito et al., 2017), attention difficulties (St Pourcain et al., 2011), depression (Rai et al., 2018), psychosis (Sullivan, Rai, Golding, Zammit, & Steer, 2013) and social anxiety (Pickard et al., 2017). Therefore we need to understand the developmental course of these characteristics right across the severity continuum.

There are two extant systematic reviews of longitudinal studies of ATs (Bieleninik et al., 2017; Seltzer, Shattuck, Abbeduto, & Greenberg, 2004). Seltzer and colleagues' (2004) review provided valuable insights into the course of autism, but nevertheless was constrained by limitations of the literature available at the time of its composition. Thus, the authors had to mainly rely on studies employing retrospective designs, which are subject to potential recall bias, and cross-sectional investigations, which cannot yield information about change over time. Ten prospective studies were located, but none reported repeated assessments of ATs using a validated measure (Seltzer et al., 2004). Subsequently to that review, Bieleninik and colleagues (2017) conducted a comprehensive and rigorous review of the longitudinal literature on ATs. This included four meta-analyses of 11 to 16 papers from a total of 31 studies reporting ADOS scores at two time-points. It did not investigate chronogeneity in depth, and so was not designed to identify subgroups based on their symptom trajectories. Furthermore, it was limited to outcomes of one measure, the ADOS and looked exclusively at clinical and at-risk populations, and so did not provide information on change and continuity of ATs right across the continuum of severity.

The present review therefore seeks to build on the work of Bieleninik and colleagues (2017) by systematically reviewing prospective longitudinal studies of ATs, in order to answer the following research questions:

1. What is the quality of the evidence for mean-level change in ATs?
2. What is the quality of the evidence for chronogeneity of ATs at intra-individual, inter-individual and inter-group levels?
3. What is the nature of the change and development in ATs?

Methods

Search strategy

A systematic search was conducted in EMBASE, PsycINFO and PubMed up to and including 31st May 2020. 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.

Inclusion/exclusion criteria

The following inclusion criteria were applied:

(i) Study reports analysis of ATs, either in those with autism, in the general population, or both.

(ii) ATs assessed using a measure with proven reliability and validity

(iii) Autism measure is a continuous scale reflecting severity of core autistic characteristics

(iv) Study is a longitudinal design

(v) ATs are measured on at least two occasions

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

The following exclusion criteria were applied:

(i) Study exclusively measured diagnostic stability over time, without addressing variability in AT severity

(ii) Retrospective measurement of ATs was used

(iii) Study reports mean scores only, without statistical analysis

Study selection process

The flow of papers through screening for inclusion in the review is shown in Figure 1.

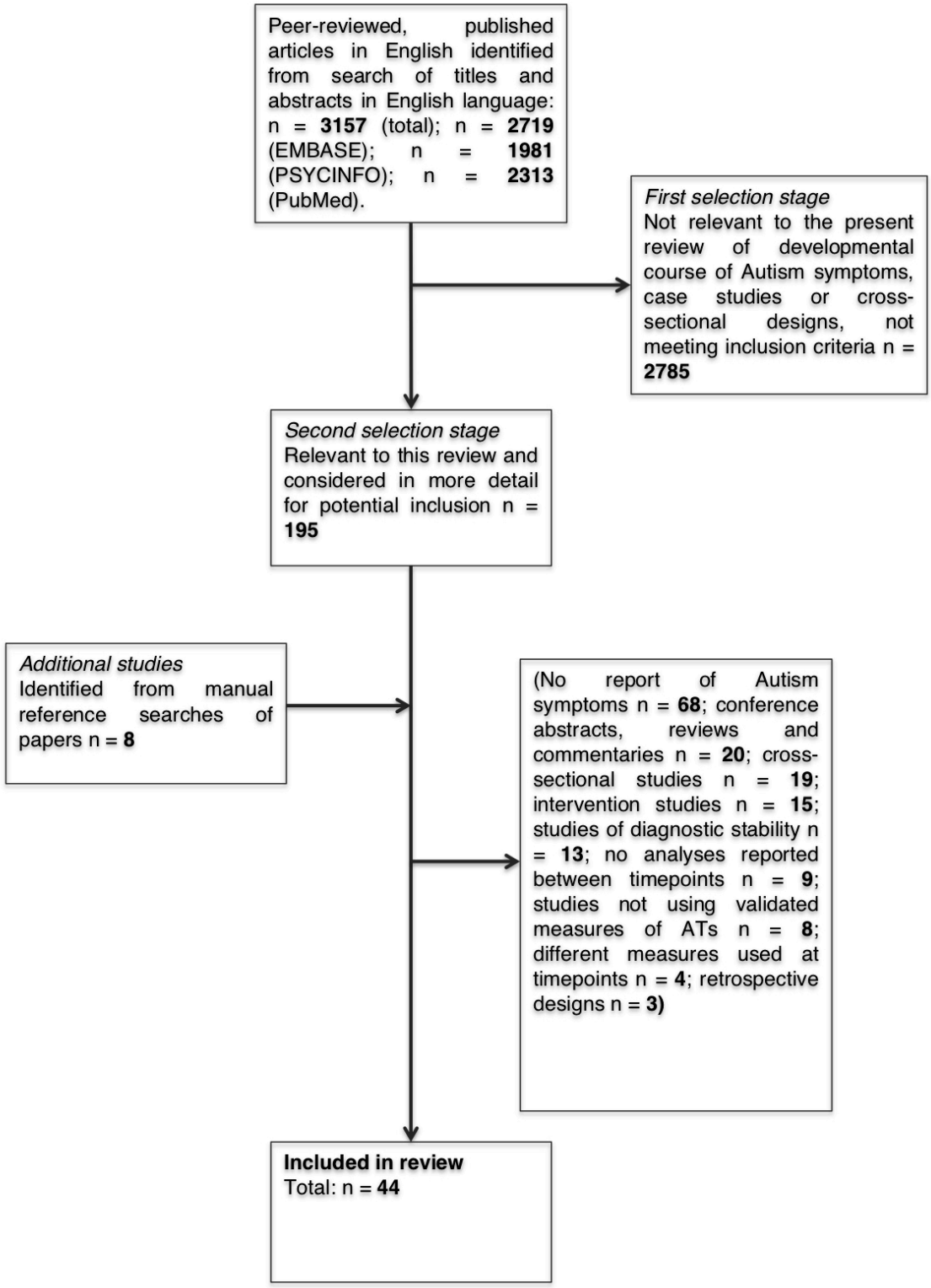


Figure 1: Flow diagram

Reporting

The 44 studies meeting inclusion criteria were divided into those that reported a single trajectory of change for the overall sample (*n* = 27; Table 2) and those that reported multiple trajectories of change between participants (*n* = 29; Table 3). This division was made for clarity of reporting. The studies reporting a single trajectory characterise mean change across the entire sample and therefore directly address our first research question (mean change across samples). The studies in Table 3 reporting multiple trajectories of change characterise the chronogeneity and therefore directly address our second and third research questions. Also, some studies presented in Table 2 investigated extraneous variables predicting variability in change of autistic traits, and so also address our second and third research questions (chronogeneity).

For clarity, studies in Table 3 reporting multiple trajectories are further categorised according to the way subgroups have been identified. A first approach is to identify *a priori groups*, meaning that the groups are specified in advance according to theory and/or diagnostic convention (for example, comparing between genders or diagnostic groupings). A second approach is to identify subgroups using *cross-sectional latent classes*. This means statistical techniques are used to allow subgroups to emerge from the data at a single time point. These groups are then followed up over time, and their trajectories compared. Thirdly, subgroups can be identified using *longitudinal latent classes*, meaning groupings emerge from the nature of the trajectories over multiple time-points. An example of a longitudinal latent class approach is LCGM.

The selection and coding procedure was conducted by the first author and overseen by supervising authors, who were consulted when dilemmas and uncertainties arose.

Nine of the thirty-one papers reviewed by Bieleninik and colleagues (2017) were included in the present review, due to our decision to exclude papers that reported scores without any statistical analysis or measures of significance. Studies were assessed for bias using a version of the NIH Quality Assessment Tool for Before-After (Pre-Post) that we modified to match the requirements of the current review.

Where effect sizes were reported in papers, they were entered into tables as definitive effects. Where information was provided such than an effect size could be estimated, effect size calculators were used to produce an effect size estimate (Supplement). See Table 2 key for further information.

Reliability and validity of measures was determined by reference to peer-reviewed validation studies. Only one study was excluded on this basis (Fountain, Winter, & Bearman, 2012), which used a Client Development Evaluation Report (CDER) register score primarily designed to measure functioning rather than to serve as a diagnostic instrument, and for which information on reliability and validity was not available.

Synthesis

As **Table 2** highlights, there is a large amount of heterogeneity in the measures used, the ages of measurement, time periods between measurement and sample compositions. Reporting of effect sizes was sparse. Therefore we concluded that a meta-analysis would not be appropriate. Bieleninik and colleagues (2017) addressed this difficulty by limiting their meta-analysis to one measure, the ADOS. This was informative, however heterogeneity remained high due to additional multiple domains of difference, and the authors urged caution in interpreting its results.

We therefore aimed to produce a narrative synthesis that could explore and navigate the multiple domains of heterogeneity, including between-study heterogeneity. Instead of aiming only for a summary of overall pattern change, we provide a synthesis of the evidence for chronogeneity of ATs across the literature. We offer a critique of the ways in which this has been characterised, and some conclusions about methodology and the state of the evidence to date. The synthesis is therefore structured by first investigating single trajectories of change, and then investigating chronogeneity. Chronogeneity is investigated at three levels, highlighted by Table 1: *intra-individual*, *inter-individual* and *inter-group.*

Table 1: *All included studies*

Study number	Authors	Country (Study)	N	Eligible N	Time points	Participants	Male (%)	Measures	Risk (/10)
1	Bacon et al (2017)	USA (1-Year Well-Baby Check-Up Approach cohort, San Diego)	273	NR	2-5 per participant: 12m – 36m ‡	ASC (39%), language delay, typically developing (42%)	NR	ADOS (Toddler, 1 and 2, raw scores) ●	9
2	Bal, Kim, Fok and Lord (2018)	USA (North Carolina and Chicago)	140	192	3: 2, 3 and 9 OR 19y	ASC	87	ADI-R SC sub-domain ●	9
3	Barbaro and Dissanayake (2017)	Australia (Social Attention and Communication Study [SACS])	77	99	2: 24m, 48m	At risk for ASC: Autism, AS, PDD-NOS, developmental delay, language disorder	76	ADOS-G Module 1 (2007 algorithm) ● ●	8
4	Charman et al (2005)	UK (Pilot RCT, London)	26	29	2: 4-5, 7 y	Autism, Atypical Autism	84.6%	ADI-R ● ●	8
5	Chawarska, Klin, Paul and Volkmar (2007)	USA (Clinic referrals, location not specified)	31	NR	2: 14-25m, 15m later ‡	ASC	71	ADOS-G Module 1 (Raw scores) ● ●	6
6	Chawarska et al (2014)	Canada and USA (Multisite: Baby Siblings Research Consortium [BSRC])	719	NR	2: 18m, 36m	High-risk siblings (53.4% typically developing outcome)	57.4	ADOS (CSS) ●	7
7	Clark, Barbaro and Dissanayake (2017)	Australia (SACS)	48	79	3: 24m, 48m, 7-9 y	ASC	75	ADOS-G (CSS) ●	7
8	Constantino et al (2009)	USA (Clinic referrals, Washington University and greater St Louis)	95	95	2: 8-15y, 5-6 y later †‡	PDD	100	Social Responsiveness Scale (SRS; Constaninto et al., 2003) ●	9
9	Darrou et al (2010)	France, Luxembourg and Switzerland (Clinic referrals)	208	280	2: Median 5y, 3 y later ‡	PDD	80	Childhood Autism Rating Scale (CARS; Schopler, Reichler, & Renner, 1988) ●	8
10	Eaves and Ho (2004)	Canada (Clinic referrals, Sunny Hill Health Centre, British Columbia)	49	NR	2: Mean 2y 9m, mean 4y 11m ‡	"Possible autism" at screening, 18% not autistic	80	CARS ●	5
11	Georgiades et al (2014)	Canada (Multisite: ASD Pathways)	280	391	2: 2-4y, 6y ‡	ASC	86	ADI-R	7

12	Giserman-Kiss and Carter (2019)	USA (University-based Early Intervention clinic)	60	61	2: 19-34m, 42-70m ‡	ASC	86.7	SRS ●	9
13	Gotham, Pickles and Lord (2012)	USA ("Early Diagnosis of ASD" cohort; Universities of North Carolina and Chicago)	345	NR	2-8 per participant: Various ‡	ASC	81.7	ADOS (CSS) ●	8
14	Haraguchi, Stickley, Saito, Takahashi, & Kamio (2018)	Japan (Tama Children's Survey; TCS)	168	2953	2: 5, 8 y	General population	53	SRS (Total, SCI, RRB) ● ● ●	9
15	Honey, McConachie, Randle, Shearer and Le Couteur (2006)	UK (Clinic referrals, North East England)	89	104	2: 24-48m, 13m later †‡	Suspected ASC, 24% not autistic	80	ADI-R ● ●	8
16	Joseph, Thurm, Farmer and Shumway (2013)	USA (PDN NIMH cohorts*)	39	128	3: Mean 4, 5.4, 6.2 y ‡	Autism	92.3	Repetitive Behavior Scale-Revised (RBS-R; Lam & Aman, 2007) ●	7
17	Kim et al (2018)	USA (University of Michigan clinic and projects)	149	NR	Multiple (mean 6 [2-21]): 14 – 36m ‡	ASC	80	ADOS (CSS) ●	7
18	Kim, Macari, Koller and Chawarska (2016)	USA (Clinic referrals, Yale Child Study Centre)	100	100	2: 14-27m, 1-2y later †‡	ASC	84	ADOS-G (CSS) ●	8
19	Lord, Bishop and Anderson (2015)	USA (Clinical referrals, North Carolina clinics, University of Chicago)	85	213	5: 2, 3, 5, 9, 19 y	ASC	92	ADI-R ● ●	8
20	Lord and Luyster (2006)	USA (North Carolina and Chicago)	297	NR	4: 2, 3, 5, 9 y	ASC	NR	PL-ADOS (raw scores) ●	8
21	Lord, Luyster, Guthrie and Pickles (2012)	USA (University of Michigan clinic and projects)	78	NR	2-10+ per participant: T1 12-19m; mean 6 times over 20m ‡	Possible ASC referrals, 38.5% not autistic	76.9	ADOS (2007 algorithm) ●	8
22	Lord et al (2006)	USA (Clinical referrals, North Carolina clinics, University of Chicago)	172	192	2: 2y, 9y	Referrals for possible ASC: Autism, PDD-NOS, non-spectrum (22%)	80	ADI-R, ADOS (raw scores) ●	8
23	Louwerse et al (2015)	Netherlands (Clinic referrals, Erasmus MC Hospital)	72	97	2: Mean 9.2y, mean 16.1y ‡	PDD-NOS	88	ADOS (CSS) ●	7
24	Mandy, Pellicano, St Pourcain, Skuse and Heron (2018)	UK (Avon Longitudinal Study of Parents and Children; ALSPAC)	9744	9744	1-4 per participant: 7, 10, 13, 16 y	General population	49	Social and Communication Disorders Checklist (SCDC; Skuse, Mandy, & Scourfield, 2005). ●	10
25	McGovern and Sigman (2005)	USA (Clinic referrals, UCLA)	45	70	2: Mean 12y8m, mean 19y ‡	Autism	88	ADI-R ● ●	6

26	Messinger et al (2015)	Canada and USA (BSRC)	1241	NR	2: 24m, 36m	ASC (14%); high-risk (54%) and low-risk (32%) non-ASC	77	ADOS (CSS) ● ●	7
27	Moore and Goodson (2003)	UK (Clinic referrals, Southampton)	20	NR	2: Mean 2y 10m, mean 4y 5m ‡	Suspected ASC: Autism, Atypical Autism, language disorder	80	ADI-R ● ●	6
28	Moss, Magiati, Charman and Howlin (2008)	UK (Clinic referrals, country-wide)	35	75	2: Mean 3.5y, mean 10.5y ‡	ASC	91	ADI-R ● ● ●	6
29	Ozonoff et al (2015)	Canada and USA (BSRC)	418	NR	3: 18, 24, 36 m	Sibling with ASC	59	ADOS (raw scores) ●	7
30	Pellicano (2012)	Australia (Clinic referrals, location not specified)	37	45	2: Mean 5y 8m, mean 8y 4m ‡	ASC	89	Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003) ● ● ●	7
31	Postorino et al (2015)	Italy (Clinic referrals, Bambino Gesu Hospital)	60	NR	2: NR	ASC	50	ADOS-G (CSS) ●	6
32	Richler, Huerta, Bishop and Lord (2010)	USA (Early Diagnosis Cohort; North Carolina and Chicago)	192	214	4: 2, 3, 5, 9 y	ASC	80	ADI-R ●	9
33	Robinson et al (2011)	UK (ALSPAC)	6539	7173	3: 7, 10, 13 y	General population	50.1	SCDC ●	10
34	Shumway et al (2012)	USA (PDN NIMH cohorts*)	89	157	2: 2-12y, 12-24m later †‡	Autism	NR	ADOS (raw and CSS) ●	7
35	Simonoff et al (2019)	UK (Special Needs and Autism Project; SNAP)	126	158	3: 10-12, 15-16, 23 y ‡	ASC	87.3	SRS ●	10
36	Soke et al (2011)	USA (Early Start Denver cohort, Colorado)	28	36	2: Mean 2y 9m, mean 4y 10m ‡	Autism	79	ADI-R ● ● ●	7
37	Starr, Szatmari, Bryson and Zwaigenbaum (2003)	Canada (Southern Ontario longitudinal cohort)	58	68	2: 6-8y, 2 y later †‡	Autism, AS	88	ADI ● ●	7
38	Szatmari et al (2009)	Canada (Southern Ontario longitudinal cohort)	57	57	4: 6-8, 10-14, 14-17, 17-19 y ‡	Autism	87.7	Autism Behavior Checklist (ABC; Krug, Arick, & Almond, 1988) ●	10
39	Szatmari et al (2015)	Canada (Multisite: ASD Pathways)	421	723	3: 2-5 yo, 6m later, 6 y †‡	ASC	84.3	ADOS (CSS) ●	7
40	Venker, Ray-Subramanian, Bolt and Weismer (2014)	USA (Not specified)	129	NR	4: 2.5, 3.5, 4.5, 5.5 y	ASC	87	ADOS (CSS) ●	8
41	Visser et al (2017)	Netherlands (Diagnosis and Intervention of Autism in the Netherlands [DIANE] study)	203	252	3: mean 2.68, 4.07, 5.62 y ‡	Referrals for possible ASC – diagnosis rate NR	80.3	ADOS (2007 algorithms) ●	8

42	Wagner et al (2018)	USA (Clinic referrals, Washington University and greater St Louis)	602	1026	Multiple per participant ‡	ASC, siblings (46.2% non autistic)	82.4	SRS •	8
43	Waizbard-Bartov et al (2020)	USA (UC David Autism Phenome Project; APP)	125	NR	2: 2-3.5, 5 y	ASC	71.2	ADOS (CSS) •	9
44	Yirmiya, Seidman, Koren-Karie, Oppenheim and Dolev (2015)	Israel (General recruitment)	39	61	2: Mean 8y, mean 11y †‡	ASC	56	ADOS (CSS) •	7

Key:

• = Global autism scores reported • = Social domain scores reported • = RRB domain scores reported

† = overlapping ages at timepoints ‡ = variable ages

* = Paediatrics and Developmental Neuroscience Branch, National Institute for Mental Health

Table 2: *Single trajectories*

Study number	Authors	Analysis	Statistically Significant?	Direction	Clinically Significant?	Effect size	Covariates	Intervention	Functioning	Communication
Global ATs										
1	Bacon et al (2017)	Growth modelling	No p = .864	Increasing	Unknown	Small Partial η^2 = .000107	NR	NR	Improved	Improved
8	Constantino et al (2009)	*t*-test	Yes p <.001	Decreasing	No	Small d = .357	NR	No effect	NR	NR
9	Darrou et al (2010)	*t*-test	Yes p value NR	Decreasing	Unknown	Unknown	NR	No effect	NR	NR
10	Eaves and Ho (2004)	NR	No p >.05	Unknown	No	Unknown	PIQ predicted AT decrease. No association with VIQ.	No effect	Worsened	NR
12	Giserman-Kiss and Carter (2019)	*t*-test	Yes p<.001	Decreasing	Unknown	Large d = .89	NR	NR	Improved	Improved
14	Haraguchi et al (2018)	ANOVA	No p = .291	Decreasing	No	Small Cohen's d = .081	No association with gender	NR	NR	NR
22	Lord et al (2006)	*t*-test	Yes p <.001 (ADOS and ADI-R)	Decreasing	Unknown	Unknown	No association with gender, maternal education, VIQ, NVIQ or adaptative behaviour	Negative effect	Worsened	NR
23	Louwerse et al (2015)	Reliable Change Index	No p value NR	Increasing	No	Unknown	NR	NR	NR	NR
26	Messinger et al (2015)	Hierarchical generalised linear model	No p = .41	Unknown	No	Unknown	NR	NR	NR	Improved
28	Moss et al (2008)	*t*-test	Yes p = .002	Decreasing	No	Large Partial η^2 =.2509	NR	NR	NR	NR
30	Pellicano (2012)	ANOVA, *t*-test	Yes p <.001	Decreasing	No	Large Partial η^2 =.584	NR	Positive effect	NR	NR
31	Postorino et al (2015)	ANOVA	No p >.05	Decreasing	No	Unknown	NR	NR	Improved	Improved
34	Shumway et al (2012)	ANOVA	No p = .22, .09	Decreasing	No	Small Cohen's d (CSS) = 0.27	Mullen VDQ and NVDQ predicted change in ADOS	NR	NR	NR

						d (Raw) = 0.36	raw scores, not CSS.			
35	Simonoff et al (2019)	Latent Growth Curve	No $p = .203$	Increasing	No	NA	IQ predicted decreasing ATs; specialist school attendance predicted increase	NR	NR	NR
36	Soke et al (2011)	*t*-test	Yes $p = .01$	Decreasing	No	Unknown	NR	NR	NR	NR
38	Szatmari et al (2009)	Hierarchical linear models, *t*-tests	No $p > .05$	Decreasing	Unknown	Unknown	NR	NR	Improved	Improved
42	Wagner et al (2018)	Growth models	No $p = .09$	Decreasing	No	Small Partial η^2 =.00717	NR	NR	NR	NR
44	Yirmiya et al (2015)	ANOVA	Yes $p = .04$	Decreasing	Unknown	Medium Partial η^2 = .06)	NR	NR	Worsened	NR
Social ATs										
3	Barbaro and Dissanayake (2017)	ANOVA	Yes $p < .001$	Decreasing	Unknown	Large Partial η^2 = .51	VDQ at 24m associated with AT change	NR	NR	NR
4	Charman et al (2005)	*t*-test	Yes $p < .001$	Decreasing	No	Large $d = 1.10$	NVIQ, not functioning, associated with AT change	NR	NR	Clinically significant improvement
5	Chawarska, Klin, Paul and Volkmar (2007)	*t*-test	Yes $p < .001$	Decreasing	Unknown	Unknown	VIQ and NVIQ associated with AT change	NR	NR	NR
14	Haraguchi et al (2018)	ANOVA	No $p = .408$	Decreasing	No	Small $d = .063$	No association with gender	NR	NR	NR
15 a)	Honey et al (2006): Autism sample (n=51)	*t*-test	No $p > .05$	Increasing	No	Small $d = -.178$	NR	NR	NR	NR
15 b)	Honey et al (2006) ASD sample (n=28)	*t*-test	Yes $p < .05$	Increasing	Yes	Medium $d = -.585$	NR	NR	NR	NR
15 c)	Honey et al (2006) non-autistic sample (n=25)	*t*-test	No $p > .05$	Decreasing	No	Small $d = .387$	NR	NR	NR	NR
25	McGovern and Sigman (2005)	ANOVA	Yes $p < .001$	Decreasing	Yes	Large	NR	NR	Improved	No change

						Partial η^2 =.5156				
26a)	Messinger et al (2015): ASD sample (*n*=252)	ANOVA, *t*-test	Yes p <.001	Increasing	Unknown	Medium d = .60	NR	NR	NR	Improved
26 b)	Messinger et al (2015): Sibling sample (*n*=989)	ANOVA, *t*-test	No p = .183	Decreasing	Unknown	Small d = .13	NR	NR	NR	Improved
26 c)	Messinger et al (2015): General population sample (*n*=583)	ANOVA, *t*-test	Yes p = .017	Increasing	Unknown	Medium d = .78	NR	NR	NR	Improved
27	Moore and Goodson (2003)	Wilcoxon signed rank test	No p value NR	Increasing	No	Unknown	No association between VIQ, NVIQ and AT change	NR	NR	No change
28	Moss et al (2008)	*t*-test	Yes p = .01	Decreasing	No	Medium d = .497	NR	NR	NR	Improved
30	Pellicano (2012)	ANOVA, *t*-test	Yes p <.001	Decreasing	No	Large d = 1.557	NR	Positive effect	NR	Improved
33	Robinson et al (2011)	Growth modelling	Yes p < .001	Decreasing	No	Unknown	NR	NR	NR	NR
36	Soke et al (2011)	*t*-test	Yes p = .005	Decreasing	No	Unknown	NR	NR	NR	NR
37	Starr et al (2003)	ANOVA	Yes p <.001	Increasing	No	Large Partial η^2 =.234	NR	NR	NR	Improved
RRB ATs										
3	Barbaro and Dissanayake (2017)	ANOVA	No p > .025	Unknown	Unknown	Unknown	MSEL VDQ at 24m not associated with AT change	NR	NR	NR
4	Charman et al (2005)	*t*-test	Yes p < .001	Decreasing	No	Large d = .998	NVIQ and functioning not associated with AT change	NR	NR	Improved
5	Chawarska, Klin, Paul and Volkmar (2007)	*t*-test	No p > .05	Decreasing	Unknown	Unknown	VIQ and NVIQ associated with AT change	NR	NR	NR
14	Haraguchi et al (2018)	ANOVA	No p = .113	Decreasing	No	Small d = .121	No association with gender	NR	NR	NR

15 a)	Honey et al (2006): Autism sample (*n*=51):	*t*-test	No *p* > .05	Increasing	No	Small: *d* = -.33	NR	NR	NR	NR
15 b)	Honey et al (2006): ASD sample (n=28)	*t*-test	Yes *p* < .001	Increasing	No	Large *d* = -1.34	NR	NR	NR	NR
15 c)	Honey et al (2006): Non-autistic sample (n=25):	*t*-test	No *p* > .05	Increasing	No	Small *d* =-.229	NR	NR	NR	NR
16	Joseph et al (2013)	Linear mixed models	No *p* value NR	Decreasing	No	Small Partial η^2 RSM = .015544 IS <.0001	NR	NR	NR	NR
25	McGovern and Sigman (2005)	ANOVA	Yes *p* <.001	Decreasing	No	Large Partial η^2 =.2509	NR	NR	Improved	No change
26 a)	Messinger et al (2015): ASD sample (*n*=252)	ANOVA, *t*-test	No *p* = .240	Increasing	No	Small *d* =.21	No association with gender	NR	NR	Improved
26 b)	Messinger et al (2015): Sibling sample (*n*=989)	ANOVA, *t*-test	No *p* = .381	Decreasing	No	Small *d* =.11	No association with gender	NR	NR	Improved
26 c)	Messinger et al (2015): General population sample (*n*=583)	ANOVA, *t*-test	No *p* = .076	Decreasing	No	Small *d* =.40	No association with gender	NR	NR	Improved
27	Moore and Goodson (2003)	Wilcoxon signed rank test	Yes *p* <.001	Increasing	Unknown	Large *r* = -.722	NR	NR	NR	NR
28	Moss et al (2008)	*t*-test	No *p* = .58	Decreasing	No	Small *d* = .111	NR	NR	NR	NR
30	Pellicano (2012)	ANOVA, *t*-test	Yes *p* <.001	Decreasing	No	Large *d* = .866	NR	Positive effect	NR	NR
36	Soke et al (2011)	*t*-test	No *p* = .99	Increasing	No	Unknown	NR	NR	NR	NR
37	Starr et al (2003)	ANOVA	No *p* = 0.62	Decreasing	No	Medium Partial η^2 =.138	NR	NR	NR	NR

Table 3: *Multiple trajectories*

Study number	Authors	Analysis	No. of groups	Groups compared	Results	Reveals heterogeneity?	Heterogeneity defined
A priori groups							
Diagnostic groupings							
1	Bacon et al (2017)	Growth models	4	1. Early-diagnosed ASD 2. Late-diagnosed ASD 3. Language delay 4. Typically-developing.	Significant effect of both group and time on ATs. Early-diagnosed, language delayed and typically-developing groups showed decreasing scores, while late-diagnosed showed sharp increase.	Yes	Differences in direction of change, associated with time of diagnosis
3	Barbaro and Dissanayake (2017)	ANOVA	2	1. Stable diagnosis 2. Cross-over diagnosis	Significant decreases in social ATs, and small, non-significant reductions in RRB ATs in both groups.	No	-
6	Chawarska et al (2007)	*t*-test	2	1. Autism 2. PDD-NOS	Significant decrease in social, communication and play ATs in both groups, and no significant change in RRB ATs in either group.	No	-
7	Clark et al (2017)	ANOVA	2	1. Stable diagnosis 2. Non-stable diagnosis	Significant main effects for group, and group-age interaction. ATs in the stable group decreased significantly at pre-school age, but increased again at school age such that there was no difference from baseline. The ASD non-stable group significantly decreased in severity between toddlerhood and school age.	Yes	Diagnostic stability is a marker for AT stability over longer time-periods
15	Honey et al (2006)	*t*-tests	3	1. Autism 2. ASD 3. Other (non-autistic)	Significant increases for ASD group in social and RRB ATs. No significant changes for Autism and Other groups. No changes in Communication in any group.	Yes	Diagnostic category associated with change.
26	Messinger et al (2015)	ANOVA, *t*-tests	3	1. High-Risk ASD 2. High-Risk Non-ASD 3. Low Risk Non-ASD	Social ATs increased significantly for both ASD and Low Risk-non ASD groups, but not High Risk-non ASD. No change in RRB for any group.	Yes	Diagnostic category and sibling status associated with change
29	Ozonoff et al (2015)	Mixed effects linear models	4	1. True Positive (TP) 2. True Negative (TN) 3. False Positive (FP) 4. False Negative (FN)	TP showed a stable, high AT trajectory, and the TN group a stable and low trajectory. The FN group demonstrated an increasing trajectory, and by 36 months still showed significantly higher scores than the TN group, indicating they were sub-threshold for diagnoses but still atypical. The FP showed	Yes	Increasing and decreasing trajectory subgroups linked to diagnostic status

					an initially increasing trajectory, followed by a rapid decrease.		
37	Starr et al (2003)	ANOVA	2	1. Autism 2. Asperger Syndrome	No significant change in RRB ATs for either group; significant increase in social ATs for both groups. Significant reduction in communication difficulties for Autism group, no difference for Asperger group.	Yes	Diagnostic category associated with communication change.
38	Szatmari et al (2009)	Hierarchical linear models, *t*-tests	2	1. Autism 2. Asperger Syndrome	ATs decreased over time in both groups at a similar rate, with AS significantly lower at all time points (p <.05).	Yes	Difference in level of ATs, but no difference in change
IQ groupings							
2	Bal et al (2018)	Growth models	3	1. Verbal-verbal (V-V) 2. Delayed-verbal (D-V) 3. Delayed-minimally verbal (D-MV)	All groups showed significant decreases in ATs. The D-MV group had the slowest curve, the V-V group the steepest. In the socioemotional reciprocity and non-verbal communication subdomains, the same pattern was repeated.	Yes	Differences in speed of change, associated with verbal development
19	Lord et al (2015)	Growth models	3	1. VIQ≥70 2. VIQ<70 3. "Very Positive Outcome" (VPO).	Social ATs for the VIQ<70 and VPO groups decreased following a quadratic pattern; scores for the VIQ≥70 group followed a linear pattern. RSM RRB ATs declined gradually (linear) for all groups. RRB IS ATs increased for the VIQ<70 group, while no significant change for other groups.	Yes	Differences in speed of change in social ATs, and presence of change in IS ATs, reflect VIQ.
25	McGovern and Sigman (2005)	ANOVA	2	1. IQ ≥ 70 2. IQ < 70	The IQ ≥ 70 group showed significantly greater decreases in social and RRB ATs.	Yes	Extent of change in ATs reflects VIQ.
Gender groupings							
24	Mandy et al (2018)	Growth models	2	1. Male 2. Female	Males' scores significantly decreased between 7 and 16, while females' scores significantly increased. At 7, males had significantly higher scores and were more likely to score in the clinical range.	Yes	Gender difference in direction of change.
33	Robinson et al (2011a)	Growth modelling	10	Male/female for each of: 1. Complete sample 2. ASD diagnosis 3. 90th IQ percentile 4. 95th IQ percentile 5. Exceeding measure cut-off	Boys were slightly lower at final follow-up than at baseline. There were no significant differences from baseline to follow-up in any other grouping.	Yes	Gender associated with difference in change.
42	Wagner et al (2018)	Growth models	7	1. Male ASD simplex 2. Male ASD multiplex	Change was most marked for a female ASD group, whose scores significantly decreased over time.	Yes	Rate of change was associated with gender,

				3. Female ASD simplex/multiplex 4. Male non-ASD siblings 5. Female non-ASD siblings 6. Male psychological condition 7. Male non-ASD, psychological condition in family			diagnosis and family status.
44	Yirmiya et al (2015)	ANOVA	2	1. Male 2. Female	Females showed increasing scores and males decreasing scores	Yes	Gender associated with direction of change
Cross-sectional groups							
6	Chawarska et al (2014)	Classification and regression tree (CART) analysis	8	Groups I-VIII defined according to mixed subgroup profiles	Analysis at 18 months found 8 groups. Three groups held participants with ASD diagnoses – two were defined by lack of eye contact and reciprocity and had stable scores; one was defined by greater eye contact and severe RRBs and had increasing scores.		
9	Darrou et al (2010)	Latent class analysis, *t*-tests	4	1. Stable high-level (HL; lower severity ATs and language scores) 2. Stable low-level (LL; higher severity) 3. HL to LL 4. LL to HL	Stable outcome for 72%, 27% moved from LL-HL, and 1% from HL-LL. The HL-HL and LL-HL groups' CARS scores significantly decreased. No significant change in LL-LL group's scores.	Yes	Language and AT severity defines subgroups with decreasing ATs.
11	Georgiades et al (2014)	Factor Mixture Modelling	6	1. Stable high AT levels 2. Stable low ATs 3. High-low ATs 4. Moderate-high ATs 5. Moderate-low ATs 6. Low-high ATs	Of the low group at time 1, 65% moved to the high group by time 2, and of the time 1 high group 18% moved to the low group. Nine percent of the entire sample showed notable reduction in ATs.	Yes	Increasing and decreasing subgroups highlighted
18	Kim et al (2016)	Hierarchical clustering analysis, generalised linear mixed models	4	1. Stable cluster I 2. Stable cluster II 3. Stable cluster III 4. Increasing cluster	Stability shown in three clusters, but significant increase over time in one for global, social and RRB ATs.	Yes	Increasing AT subgroup highlighted
Longitudinal groups							
13	Gotham et al (2012)	Latent class growth models;	4	1. Persistent High (46%)	A trend for decrease in social ATs and increase in RRB ATs for all but the	Yes	Increasing and decreasing groups highlighted,

		multinomial logistic regression		2. Persistent Moderate (38%) 3. Increasing (9%) 4. Decreasing (7%)	increasing group, which rose for both. Gender, race and NVIQ did not predict class membership; lower VIQ predicted membership of Persistent High. VIQ increased for all classes, but steeper in Decreasing. Decreasing had higher and improving daily living skills; other groups significantly declined.		associated with VIQ, NVIQ, daily living skills.
17	Kim et al (2018)	Latent class growth models	4	1. Persistent high (23%) 2. Persistent low (25%) 3. Increasing (27%) 4. Decreasing (25%)	Increasing showed a significant rise in ATs, and Decreasing a non-significant trend. NVIQ was stable for all groups; VIQ improved significantly for Low, Increasing and Decreasing. Children with IQ below 85 were significantly more likely to belong to High or Decreasing. Groups did not differ by gender, siblings with autism or maternal education. Increasing was more cognitively able, and showed improvements in language.	Yes	Increasing and decreasing groups highlighted, associated with VIQ, IQ
20	Lord and Luyster (2006)	Latent class growth models	4	1. Persistent high 2. Persistent low 3. Increasing 4. Decreasing	Model reported	Yes	Increasing and decreasing groups highlighted
21	Lord et al (2012)	Latent class growth models; marginal regression models	4	1. Persistent high (21%) 2. Persistent low (40%) 3. Increasing (21%) 4. Decreasing (19%)	No significant change over time in social or RRB ATs in High; both significantly rose for Increasing. Only social ATs significantly declined in Decreasing. Marginal decrease in social ATs for Low. Groups did not significantly differ by gender or maternal education. Treatment only distinguished High from Low. Decreasing showed significantly faster increase in NVIQ than High, and Improving and Low classes faster increase in VIQ than High.	Yes	Increasing and decreasing groups highlighted, associated with NVIQ and VIQ
32	Richler et al (2010)	Latent class growth models	6	1. RSM Persistent High (25%) 2. RSM Persistent Low (25%) 3. RSM Decreasing (50%) 4. IS Persistent Moderate (16%) 5. IS Persistent Low (13%)	In RSM Decreasing, 88% continued to show high numbers of RSM behaviours. Higher cognitive ability at age 2 predicted fewer and decreasing RSM. RSM was not associated with social difficulties. IS was not associated with cognitive ability, but was associated with social difficulties – milder IS was associated with greater social impairment.	Yes	Increasing and decreasing groups highlighted in both RSM and IS domains, associated with IQ and social difficulties. RSM and IS shown to be separable.

				6. IS Increasing (71%)			
39	Szatmari et al (2015)	Latent class growth models	2	1. Persistent High (88.6%) 2. Decreasing (11.4%)	Decreasing significantly decreased. Girls were more significantly more likely to be in Decreasing. Decreasing showed significantly higher developmental and language scores.	Yes	Decreasing group highlighted, associated with gender, development and language
40	Venker et al (2014)	Latent class growth models; multinomial logistic regression; multi-level models	4	1. Persistent High (36.4%) 2. Persistent Moderate (41.8%) 3. Increasing (7.8%) 4. Decreasing (14%)	Class membership was not significantly associated with gender, ethnicity, maternal education or language loss. Intervention significantly predicted membership of High. NVIQ was significantly lower in High than Decreasing. Daily living skills were significantly lower in High than all others. Receptive language growth was significantly higher for Increasing and Decreasing than High. High had significantly slower growth in expressive language than others.	Yes	Increasing and decreasing groups highlighted, associated with NVIQ, daily living skills and receptive language.
41	Visser et al (2017)	Latent class growth models	5	1. Persistent High (19.5%) 2. Persistent Moderate (21.7%) 3. Persistent Low (48%) 4. High Decreasing (5.4%) 5. Moderate Decreasing (5.4%)	High had stable low NVIQ, and was the only group without improved language skill. ADHD scores were significantly lower than other groups. All participants had ASD diagnoses. Moderate had below average and increasing NVIQ. ADHD scores significantly increased. Low group ASD diagnoses increased from 35 to 58%. High Decreasing showed increasing NVIQ and low ADHD scores. Moderate Decreasing had improving NVIQ and moderate declining ADHD scores. Three of eight children in this group lost their ASD diagnosis.	Yes	Increasing and decreasing groups highlighted, associated with ADHD, NVIQ and diagnostic change.
43	Waizbard-Bartov et al (2020)	Reliable Change Index (RCI)	3	1. Decreased Severity Group (28.8%) 2. Stable Severity Group (54.4%) 3. Increased Severity Group (16.8%)	Almost half the sample showed reliable change in ADOS CSS scores over time. Increased group had lower initial severity scores. No relationship with intervention. Girls were overrepresented in Decreased and underrepresented in Increased groups. Decreased group showed higher IQ scores and increases in Communication and Daily Living, without loss of Motor Skills shown in Increased group.	Yes	Increased and Decreased groups' presence detected based on reliable change. Group membership linked to differences in gender, IQ, functioning and communication.

Results

Quality and bias

Risk of bias was found to generally be moderate to low. Studies mostly scored highly for clarity of objectives, clear eligibility, selection and diagnostic processes, representative clinical samples and validated, appropriate outcome measures. Frequent limitations included small sample sizes, poor reporting of loss to follow-up (increasing risk of attrition bias) and low levels of enrolment from eligible participants (increasing risk of selection bias). Only 4 of 44 studies scored 10 (lowest possible risk of bias) (Mandy, Pellicano, St Pourcain, Skuske, & Heron, 2018; Robinson et al., 2011; Simonoff et al., 2019; Szatmari et al., 2009), and 7 scored 9 (Bacon et al., 2017; Bal, Kim, Fok, & Lord, 2018; Constantino et al., 2009; Giserman-Kiss & Carter, 2019; Haraguchi, Stickley, Saito, Takahashi, & Kamio, 2018; Richler, Huerta, Bishop, & Lord, 2010; Waizbard-Bartov et al., 2020). These studies notably had large sample sizes, measured ATs at multiple time points, and offered detailed statistical output. Studies with higher risk of bias generally offered briefer methods reports and omitted key aspects of statistical output. The quality of statistical output was more variable, with many studies scoring lower due to overlapping and variable time-points, unclear sample composition and failure to report either effect sizes or sufficient information to allow them to be estimated.

Mean trajectories of ATs

Overall, there was some evidence to suggest a general trend for mean decreases in ATs over time, linked to social but not RRB ATs. This appears clearest for those with autism diagnoses, rather than subclinical ATs. Decreasing scores were associated with higher verbal and non-verbal IQ, but not gender, functioning, language or intervention. However, this broad finding is limited by methodological issues including quality of reporting, and heterogeneity between samples and measures. Furthermore, the presence of chronogeneity within samples highlights a key conceptual limitation, which is that mean statistics across entire populations obscure significant variability.

Global ATs

Outcomes for change in global ATs across entire samples (single trajectories) were divided between those finding no change (Bacon et al., 2017; Eaves & Ho, 2004; Louwerse et al., 2015; Haraguchi et al., 2018; Messinger et al., 2015; Postorino et al., 2015; Shumway et al., 2012; Simonoff et al., 2019; Szatmari et al., 2009; Wagner et al., 2018) and those finding evidence for a general trend for decrease over time (Constantino et al., 2009; Darrou et al., 2010; Giserman-Kiss & Carter, 2019; Lord et al., 2006; Moss, Magiati, Charman, & Howlin, 2008; Pellicano, 2012; Soke et al., 2011; Yirmiya et al., 2015). No studies reported a significant increase over any given time period. Of the studies reporting a significant decrease, three (Giserman-Kiss & Carter, 2019; Moss et al., 2008; Pellicano, 2012) showed large effect sizes, across small sample sizes and relatively long (3-5 year) time periods. There was a notable difference in measurement: 4 of 10 studies that found no significant change used ADOS Calibrated Severity Scores, which are standardised for age and language ability (CSS; Gotham, Pickles, & Lord, 2009). Three of eight studies that observed a significant change used the Autism Diagnostic Interview-Revised (ADI-R; Le Couteur, Lord, & Rutter, 2003), a reported, rather than observational measure. This is in keeping with Bieleninik and colleagues' (2017) finding that ADOS CSS scores are stable over time, and this is likely to be explained by the fact that they are adjusted for development (Gotham et al., 2009).

Social ATs

A majority of studies (8 of 13) investigating change in social ATs using a single-trajectory approach reported a significant decrease in scores over time (Barbaro & Dissanayake, 2017; Charman et al., 2005; Chawarska, Klin, Paul, & Volkmar, 2007; McGovern & Sigman, 2005; Moss et al., 2008; Pellicano, 2012; Robinson et al., 2011; Soke et al., 2011). However, some studies showed no change, or even an increase in autistic social difficulties over time. Two studies showed no significant change (Haraguchi et al., 2018; Moore & Goodson, 2003) and one a significant increase (Starr, Szatmari, Bryson, & Zwaigenbaum, 2003). Two studies (Honey, McConachie, Randle, Shearer & LeCouteur, 2006; Messinger et al., 2015) showed both increasing and stable social ATs across multiple samples. Where ADOS CSS were reported, they found either no change or an increase, replicating the finding that ADOS CSS scores are relatively less likely to show a decreasing trend found across other measures. Where effect sizes were estimated, they were found to be large for ADI-R and ADOS raw scores, while

the Social Communication Disoders Checklist (SCDC; Skuse, Mandy & Scourfield, 2005) and Social Responsiveness Scale (SRS; Haraguchi et al., 2018) showed smaller effects.

Restricted and repetitive behaviour ATs

By contrast a majority of studies investigating the severity of RRBs over time using a single-trajectory approach reported no change in ATs (Barbaro & Dissanayake, 2017; Chawarska et al., 2007; Haraguchi et al., 2018; Honey et al, 2006 [2 samples]; Joseph et al., 2013; Messinger et al., 2015; Moss et al., 2008; Soke et al., 2011; Starr et al., 2003). Where significant change was observed, there was no consistency to the direction of change, with large effect sizes observed for both decreasing (Charman et al., 2005; McGovern & Sigman, 2005; Pellicano, 2012) and increasing RRB (Honey et al., 2006 [one sample], Moore & Goodson, 2003). The ADI-R again appeared to produce a greater likelihood of variable outcomes over time. The only study to use a detailed measure specific to RRB (Joseph et al., 2013), the Repetitive Behaviors Scale-Revised (RBS-R; Bodfish, Symons, Parker, & Lewis, 2000), found no significant change.

Predictors of AT change

There was some evidence to support that IQ reliably predicted decreases in social ATs, but not RRB (Barbaro & Dissanayake, 2017; Charman et al., 2005; Chawarska et al., 2007), with only one study finding no significant association (Moore & Goodson, 2003). This finding was replicated in Global ATs, with three studies finding IQ predicted decreasing trajectories (Eaves & Ho, 2004; Shumway et al., 2012; Simonoff et al., 2019) and one showing no association (Lord et al., 2006).

Where investigated, gender did not significantly predict change in ATs (Haraguchi et al., 2018; Lord et al., 2006; Messinger et al., 2015). Reporting on the effect of intervention (eg Applied Behaviour Analysis; ABA) on changes in scores was poor, with few studies investigating associations. Where reported, results were inconsistent, with one study finding each of an association with decreasing (Pellicano, 2012) or increasing trajectories (Lord et al., 2006), and three finding no relationship (Constantino et al., 2009; Darrou et al., 2010; Eaves & Ho, 2004).

A sensitivity analysis comparing outcomes by diagnostic groupings did not reveal a consistent pattern of differences linked to diagnostic grouping. However, a comparison between diagnosed autistic samples and non-autistic samples did suggest a difference in patterns of change. Analysis limited to non-autistic or general population samples (Honey et al., 2006; Haraguchi et al., 2018; Messinger et al., 2015 [two samples]; Robinson et al., 2011) and a sizeable proportion of non-autistic participants (Bacon et al., 2017; Eaves & Ho, 2004; Lord et al., 2006; Wagner et al., 2018) revealed non-autistic samples were especially likely to show no significant change over time in any domain. A sensitivity analysis restricted to the highest quality studies did not suggest any difference in patterns of change.

Chronogeneity of ATs

Intra-individual chronogeneity

Firstly, evidence for intra-individual chronogeneity was provided by the apparent separability of social and RRB domains in mean change reported above. Secondly, there is emerging evidence that the RRB domain itself can be sub-divided. Richler and colleagues (2010) applied LCGM to investigate separate factors within the RRB domain - Repetitive Sensorimotor (RSM) and Insistence on Sameness (IS). RSM showed a wider distribution and greater reductions in difficulty, while for IS there was a narrower range of scores and the majority showed increasing difficulty over time. This supports the idea that the RRB subdomain itself is fractionable (Cuccaro et al., 2003) and its apparent stability may mask underlying chronogeneity. Future longitudinal studies should therefore investigate RSM and IS factors as well as overall RRB scores.

Intra-individual chronogeneity was also shown by the disconnection between patterns of change in ATs, language and functioning. Of four single-trajectory studies that found significantly improved daily functioning (Bacon et al., 2017; Giserman-Kiss & Carter, 2019; Postorino, 2015; Szatmari et al., 2009), and five studies that found significantly improved communication (Bacon et al., 2017; Giserman-Kiss & Carter, 2019 Messinger et al., 2015; Postorino et al., 2015; Szatmari et al., 2009), only one showed a decrease in global ATs (Giserman-Kiss & Carter, 2019). However, three of the four studies showing

improved communication showed significant decreases in social ATs (Charman et al., 2005; Moss et al., 2015; Pellicano, 2012; Starr et al., 2003). Therefore, there is evidence to suggest not only that the fractionation of social and non-social domains of autism applies longitudinally, but also that there is a complex longitudinal relationship between ATs, language and functioning. While improved communication appears related to decreasing social difficulties, there is no consistent evidence for a longitudinal relationship between any ATs and daily functioning. Few studies reported outcomes related to daily functioning, and none reported outcomes relating to quality of life and mental health.

Inter-individual chronogeneity

The existence of inter-individual chronogeneity was strongly supported by evidence from single and multiple trajectory studies. Louwerse and colleagues (2015) observed that despite no reliable change across the entire sample, 40% of participants showed a reliable increase and 20% a reliable decrease – meaning that while a global metric of change suggested stability of ATs, a majority of the sample did show reliable change in one direction or another. Across single trajectory studies, it was observed that 12-43% of the overall sample showed a change in diagnostic status during the course of the study, indicating clinically significant change is consistently taking place for a minority of participants, which is not reflected by broad trends across entire samples (Barbaro & Dissanayake, 2017; Eaves & Ho, 2004; Lord et al., 2006; Moore & Goodson, 2003; Pellicano, 2012; Soke et al., 2011).

Some multiple trajectory studies defined *a priori* groups according to diagnostic change (Table 3), and were therefore designed to further elaborate upon the AT trajectories of participants whose diagnostic status altered during the course of the study. Ozonoff and colleagues (2015) used groupings based on diagnostic stability to reveal the presence of subgroups of participants with significantly increasing and decreasing ATs. Similarly, Barbaro and Dissanayake (2017) and Clark, Barbaro and Dissanayake (2017) used the same Social Attention and Communication Study (SACS) dataset to identify a stable diagnostic group (15%) and a “crossover” group with decreasing scores (27%). Decreasing scores therefore appear linked to the well-established finding that a subset of children with autism diagnoses will no longer meet criteria for diagnosis at follow-up (Woolfenden et al., 2012). The link to diagnostic change also lends support to the clinical significance of the decreasing scores. The limitation of such

studies is that they are limited to clinical samples and based on diagnostic convention, and as a result may underestimate increasing trajectories and underrepresent female participants.

Similarly, a cross-sectional latent class approach (Table 3) allowed high and low severity groups to emerge at baseline and follow-up, and compare rates of transition (Darrou et al., 2010; Georgiades et al., 2014). In one study, by follow-up, 27% of participants transitioned from the high to the low severity group and only 1% from the low to the high (Darrou et al., 2010). This is likely to reflect high baseline scores in a diagnostic sample. In another, 65% transitioned from low to high and 18% from high to low, with 9% of the overall sample showing significantly decreased scores with large effect sizes (Georgiades et al., 2014). These studies supported the presence of chronogeneity in the samples, but gender, IQ and other predictors were not assessed. This approach had the relative benefit of allowing high and low groups to emerge from the data, but were not clearly benchmarked against clinical thresholds so the magnitude and clinical significance of change could not be determined.

A longitudinal approach to group comparisons used Reliable Change Index (RCI) scores to allocate individuals to groups that showed reliably increased or decreased scores and those whose trajectories remained stable (Waizbard-Bartvov et al., 2020). This methodology highlighted the chronogeneity within the sample, as almost half of participants showed reliable change in ADOS CSS scores over time. A relative limitation of this approach however is that trajectories are constrained to a maximum of three possible groups, whereas latent class growth modelling (LCGM) by contrast allows multiple groups to emerge based on the best fit to the data.

For this reason, the most successful approach to inter-individual chronogeneity in this review involved LCGM approaches. These statistical techniques use the entire trajectories of individuals over time to put them into groups (Table 3). Eight studies taking this approach were identified (Gotham, Pickles, & Lord, 2012; Lord & Luyster, 2006; Lord, Luyster, Guthrie, & Pickles, 2012; Kim, Macari, Koller, & Chawarska, 2018; Richler, Huerta, Bishop, & Lord, 2010; Szatmari et al., 2015; Venker, Ray-Subramanian, Bolt, & Weismer, 2014; Visser et al., 2017).

In an LCGM approach, chronogeneous data are grouped into relatively homogeneous subgroups, following similar patterns of change over time. Overall, it appeared that a four-class model composed of consistently high, consistently low, increasing and decreasing trajectories most consistently characterised the data. This revealed that there was noteable inter-individual chronogeneity across studies. Indeed, four studies used ADOS CSS scores, supporting the idea that the apparent stability of CSS across large samples obscures significant chronogeneity (Gotham et al., 2012; Kim et al., 2018; Szatmari et al., 2015; Venker et al., 2014). All studies included only participants referred or diagnosed with autism, and characterised change at the severe or clinical end of the AT continuum.

Five studies applying LCGM to ADOS scores found a four-class model best fit the data (Gotham et al., 2012; Lord & Luyster, 2006; Lord et al., 2012; Kim et al., 2018; Venker et al., 2014; see Table 3 for details). Two of these studies used potentially-overlapping Early Diagnosis Cohort data (Gotham et al., 2012; Lord & Luyster, 2006), and Kim and colleagues (2018) extended the analysis of Lord and colleagues (2012) by using 41% overlapping data. All papers observed consistently high and low classes, and increasing and decreasing classes. The confidence with which individuals are allocated to groups is measured using an entropy statistic, and these were all in the good or excellent range (Muthén, 2004). Models were selected using the lowest Bayesian Information Criteria (BIC) value. BIC reflects the log-likelihood of a model and performs well as a basis for model selection (Nylund, Asparouhov, & Muthén, 2007). Of these five studies only Lord and Luyster (2006) did not report their fit indices or information about how the model was selected.

A further two studies found a five-class and a two-class model respectively (Visser et al., 2017; Szatmari et al., 2015), although BIC values did not appear to support these selections. Visser and colleagues (2017) selected a five-class model for ADOS scores but BIC and entropy values also appeared to favour fewer classes. Szatmari and colleagues (2015) selected a two-class model including a stable high and a decreasing class. However, BIC values continued to improve as more classes were added, and fit indices for quadratic models were not reported – more classes appear to have been warranted. Therefore there was not yet any convincing evidence to contradict the four-class model.

One further study applied LCGM to RRB scores (Richler et al., 2010). RRB was separated into Repetitive Sensorimotor (RSM) and Insistence on Sameness (IS) factors, and over time RSM showed a wider distribution and also greater reductions in difficulty (50% were in a decreasing group), while for IS there was a narrower range of scores and the majority were allocated to an increasing score group.

LCGM provided good evidence for the presence of inter-individual chronogeneity, and that this can best be characterised by four trajectories. A potential limitation of LCGM is that large numbers of spurious classes could be extracted, which lack real-world meaning (Bauer and Curran, 2003). In particular, the four-class representation of change that was observed across studies has been found to arise (Sher, Jackson, & Steinley, 2011). For this reason, it is crucial that LCGM studies establish the external validity of the groups by demonstrating associations with extraneous variables (Muthén, 2003). It was therefore helpful that studies reported meaningful diagnostic differences and changes between groups, which helped validate their relation to real-life practice. The finding of a connection to IQ further helped characterise their real-world differences, and the presence of genuine chronogeneity. Future studies should include greater numbers of female participants, to investigate gender differences (in particular, general population studies are needed). They should also investigate behavioural and mental health difficulties and pre-baseline predictors, and continue if possible into adulthood.

Inter-group chronogeneity

Studies explored inter-group chronogeneity using diagnostic, gender and IQ groupings. IQ but not diagnostic grouping was linked to chronogeneity, and gender appeared to interact with baseline severity. Group-level variables showed clearer relationships to trajectories when defined *a priori,* and weaker relationships when investigated as post-hoc variables in LCGM studies. More LCGM studies with wide ranges of predictors in general population samples are needed.

Where diagnostic categories were used as *a priori* grouping variables, it did not reveal chronogeneity. Four papers compared a group with strictly-defined autism diagnoses to a group with broader ASC diagnoses (Chawarska, Klin, Paul, & Volkmar, 2007; Honey et al., 2006; Starr, Szatmari, Bryson, & Zwaigenbaum, 2003; Szatmari et al., 2009) and one to a general population and sibling group

(Messinger et al., 2015). All papers observed comparable patterns of change, albeit with a severity difference that was consistently maintained over time, ie. there was a fixed distinction between groups (cross-sectional heterogeneity), but no influence of time (chronogeneity).

When gender was used as an *a priori* grouping variable, there was some evidence of inter-group chronogeneity. Studies comparing trajectories between genders suggested that female autistic participants' social ATs significantly decreased, while female non-autistic participants' social ATs significantly increased. Therefore, severity of initial ATs appears to interact with gender to predict change over time, but this is limited to a small number of high-quality studies. Wagner and colleagues (2018) observed significant and markedly decreasing social ATs for a female group with ASD diagnoses. Robinson and colleagues (2011) observed that in a large general population dataset, male participants had significantly higher mean social ATs at age 7 than female participants. Mandy and colleagues (2018) extended this study until the age of 16 and observed that by this point female participants' social ATs increased such that there was no longer any significant gender difference. Female gender appears to be related to broadly decreasing social ATs from the clinical end of the broader AT spectrum, but to increasing social ATs during adolescence across the general population. Both studies had large sample sizes, but more evidence is needed to clarify the veracity of these findings.

Studies using latent class approaches allowed groups to emerge from the data itself, then investigated gender as a post-hoc predictor of membership of these groups. When gender was investigated in this fashion, there was little convincing evidence for inter-gender chronogeneity. Two studies used a cross-sectional latent class approach, extracting large numbers of similar latent classes that were of limited utility for investigating predictor variables (Chawarska et al., 2014; Kim, Macari, Koller & Chawarska, 2016). Most of these did not show clear differences in demographics or severity, and were therefore difficult to interpret. There was some evidence however to suggest that increasing scores were linked to baseline characteristics of more severe RRB (Chawarska et al., 2014) and lower adaptive skills but not gender (Kim et al., 2016).

When longitudinal latent class approaches (LCGM) were used, gender again did not appear to significantly predict class membership. Only one (Szatmari et al., 2015) of five studies (Gotham et al., 2012; Kim et al., 2018; Lord et al., 2012; Szatmari et al., 2015; Venker et al., 2014; Visser et al., 2017) found a significant relationship, and observed that boys were more likely to be in a high stable AT group, and girls to be in a decreasing AT group. Importantly a simpler methodology of allocating to groups based on RCI scores replicated this association, as female gender was again significantly associated with membership of the decreasing trajectory group (Waizbard-Bartvov et al., 2020).

It is possible that studies were underpowered to find gender differences in most cases due to relatively low numbers of female participants (13-23%). There is also a possibility of a male-biased diagnostic approach (Dworzynski, Ronald, Bolton, & Happé, 2012; Goldman, 2013). Further, all LCGM studies used global ATs (ADOS CSS or 2007 algorithms), and none investigated social AT domains separately – it may be that female gender is associated with change in social ATs but not RRB, and further research is needed to clarify this relationship. No latent class approaches included general population participants, and so the finding of increasing scores for non-autistic female participants (Mandy et al., 2018) could not be replicated. LCGM studies using general population participants are especially needed to clarify the chronogeneity of ATs.

Although IQ is a continuous variable, it has been used as a grouping variable in the literature and led to an apparent association between higher IQ and both decreasing ATs and faster rates of decrease in autistic (diagnosed) samples. Defining groups *a priori* by IQ led to an observed connection between higher IQ and decreasing ATs across two studies. McGovern and Sigman (2005) observed a significantly greater decrease in both social and RRB ATs for a group with IQ scores greater than 70, while Lord, Bishop and Anderson (2015) observed more rapid social AT decreases in a higher-IQ group. This association held when limited to the verbal IQ domain - Bal, Kim, Fok, and Lord (2018) defined verbal ability at two time points, and found significantly faster decreases in AT scores for the group that was verbal at both time points than for a delayed verbal group. A group that remained minimally verbal showed no change in ATs over time.

For IQ, as with gender, there was some evidence to suggest that this pattern is moderated by baseline severity because in a large (n>6000) general population dataset, Robinson and colleagues (2011) did not find that patterns of change differed for high-IQ subsamples. More evidence from general population (subclinical) AT trajectories in particular is needed to clarify this apparent distinction.

While IQ was used as a grouping variable in the studies above, the findings were supported by the fact that IQ emerged as a consistently significant predictor of group membership in five LCGM studies (Gotham et al., 2012; Kim et al., 2018; Lord et al., 2012; Venker et al., 2014; Visser et al., 2017) and one longitudinal group design using RCI (Waizbard-Bartvov et al., 2020), with only one (Szatmari et al., 2015) finding no relationship. Higher verbal IQ predicted membership in low and decreasing classes, compared to a stable high class (Gotham et al., 2012; Kim et al., 2018; Lord et al., 2012; Venker et al., 2014; Visser et al., 2017). Increasing non-verbal IQ also predicted membership of low and decreasing classes in three studies (Gotham et al., 2012; Kim et al., 2018; Lord et al., 2012), but also of the increasing class in one study (Visser et al., 2017). Lower and slower-growing non-verbal was associated with membership of a high stable class (Venker et al., 2014).

In conclusion, there is evidence for inter-group chronogeneity, with gender and IQ groupings (but not diagnostic categories) demonstrating that extraneous variables are able to predict change over time. The nature of this relationship is complex, and it appears that baseline AT severity and AT subdomains interact with grouping variables. That is to say, levels of chronogeneity interact and are not entirely separable.

Discussion

This article systematically reviewed the literature on AT trajectories, including both single-trajectory studies that characterised mean change and multiple-trajectory studies that were able to describe chronogeneity. There is some limited evidence that ATs tend to reduce over time for autistic children and this reflects change in social but not RRB traits. Scores that were age-adjusted did not show the same pattern of reduction over time, replicating the findings in Bieleninik and colleagues' (2017)

review. In general population, as opposed to clinical, samples convincing evidence for a population-level change over time is lacking.

Chronogeneity was revealed at intra-individual, inter-individual and inter-group levels. There was evidence to support the idea that autism is fractionable (Happé, Ronald, & Plomin, 2006) longitudinally, with apparent disconnections between social and RRB domains, and between ATs, language and functioning. The observation that social ATs tended to decline while RRBs did not, fits with the idea that autism comprises a social and a non-social dimension with partially distinct underlying mechanisms (Mandy & Skuse, 2008; Ronald, Happé, & Plomin, 2005). Further, there was some support for longitudinal differences within the RRB domain itself between higher-order (cognitive – insistence on sameness) and lower-order (behavioural – repetitive stereotyped behaviours) components. These domains are likely to have different genetic and neurological underpinnings (Cuccaro et al., 2003; Turner, 1999). An LCGM study showed that the former tended to stay stable or increase over time, while the latter tended to reduce (Richler et al., 2010).

Around 20% of samples consistently showed each of increasing and decreasing ATs. Female gender and higher IQ were associated with decreasing social ATs in diagnosed samples, and female gender was linked to increasing ATs in general population participants. In summary, while there is limited evidence for population-level change in ATs over time, this obscures the existence of subgroups of participants reliably showing significantly increasing or decreasing trajectories. However, given the current quality of evidence this remains a relatively general finding and a great deal more work is needed to reliably characterise this change and to investigate links to a range of predictors.

Groups of participants with increasing and decreasing ATs over time were revealed by both theory-driven approaches (eg. comparing diagnostic crossover groups) and data-driven methods using latent classes. This variability is likely to reflect the effects of a complex relationship between genetic and environmental influences (Hegarty et al., 2019) as biological and behavioural factors interact across the lifespan in complex feedback and feedforward loops (Elsabbagh, 2020).

A behavioural factor that is likely to be of particular significance is autistic "camouflaging," which has been defined as a collection of behaviours used to compensate for or mask autistic traits in social situations (Hull et al., 2017). Examples of camouflaging include a young autistic person acquiring a repertoire of gestures from watching and imitating class-mates, or learning to approximate flexible eye-contact by making a conscious effort to look between the eyes of others. There is emerging evidence that camouflaging is an important part of the experience of many autistic people, including children and adolescents (Bargiela, Steward & Mandy, 2016; Lai et al., 2017). Some autistic people have reported that they learned over time to camouflage their repetitive stereotyped behaviours, for example by suppressing "stimming" (Mandy & Tchanturia, 2015). Given that such camouflaging behaviours are effortful cognitive strategies (Hull et al., 2017), this could help explain the relationship discovered in this review between higher IQ and decreasing ATs over time. Conversely, the breakdown of simpler childhood camouflaging strategies like social imitation in the increasingly complex social milieu of adolescence may explain the later emergence of increasing ATs for autistic girls (Bargiela et al., 2016; Kopp & Gillberg, 2011; Mandy et al., 2018).

None of the studies in this review examined the influence of camouflaging on AT trajectories, and this relationship is likely to be nuanced and requires further investigation. Camouflaging has been shown to be associated with experiences of exhaustion (Hull et al., 2017; Bargiela et al., 2016) increased risk of depression (Cage, Di Monaco & Newell, 2018) and is a marker for suicidality (Cassidy, Bradley, Shaw & Baron-Cohen, 2018). Conversely, compensation for autistic traits can be an adaptive coping strategy in some social contexts (Livingston, Shah & Happé, 2019), perhaps reflecting the fact that autistic people are often required to make adaptations when the external environment fails to appropriately do so (Mandy, 2019). Recent neuroimaging studies have revealed that compensatory strategies are deeply embedded in autistic people's automatic experiences (Lai et al., 2019) suggesting the autistic experience at a neurocognitive level is profoundly masked by a reduction of visible ATs. In this way the reorganisation of neural networks underlying cognition and behaviour both shape, and are shaped by, learning experiences from the environment (Elsabbagh & Johnson, 2016), highlighting the complex interaction between biological, cognitive and behavioural domains. Alternatively, camouflaging may play a secondary or less prominent role in the development of ATs whose emergence or decline may be determined primarily by maturation and biological processes.

Future studies are needed that track indices of camouflaging longitudinally, perhaps using the Camouflaging Autistic Traits Questionnaire (CAT-Q; Hull, 2019) as well as genetic and neurobiological markers, to clarify the underlying processes of development.

The available literature at the time of this review has not yet been designed to investigate these multiple levels of complexity. Single-trajectory approaches have helped to reveal a general trend at the population level. Meanwhile multiple-trajectory approaches reviewed have included comparisons between groups that are specified *a priori* (following theoretical or diagnostic convention), those defined by cross-sectional latent classes and those revealed by longitudinal latent class methods (LCGM). *A priori* methods were useful in highlighting broad, group level differences but limited by being driven by convention and by limited scope for exploring multiple predictors. By contrast, as a data-driven approach LCGM allowed groups to emerge and is open to a wider range of future developments including a greater number of predictors and increasingly nuanced modelling techniques. A relatively reliable finding was that a four-class model best characterised AT trajectories, including increasing, decreasing, stable high and stable low-moderate groups. The external validity of these groups was supported by reliable relationships with IQ as well as with diagnostic ratios and change. However given the risk of spurious classes (Bauer & Curran, 2003; Sher, Jackson, & Steinley, 2011) further work is needed to replicate and characterise these trajectories. Furthermore, no studies have yet applied LGCM to general population participants, and extant studies were limited to those with clinical diagnoses. This precludes the possibility of detecting individuals with ATs increasing from a subclinical to a clinical range. It also led to inclusion of relatively few female participants, limiting power to detect associations between gender and group membership.

All LCGM studies were limited to childhood, and future studies are needed that extend trajectories into adulthood to explore the influence of predictors across a longer time period, simultaneously clarifying whether childhood trajectories have validity and prognostic value. Furthermore, the role of intervention was not widely investigated as a predictor variable. Future LGCM studies should investigate the role of intervention, perhaps as a time-varying covariate to order to capture its effect on AT trajectories. Additionally, studies reviewed had relatively small sample sizes and there is a need for data-driven approaches such as LGCM to be applied to large datasets including general

population and clinically-diagnosed participants in order to establish reliable relationships between ATs and covariates (Lombardo, Lai & Baron-Cohen, 2019).

Future LGCM studies will be able to draw from a range of available modelling approaches, including Growth Mixture Modelling (GMM), Trajectory Grade of Membership Models and Latent Class Growth Analysis, all of which allow relatively homogeneous groups to emerge from heterogenous data and for inclusion of a range of predictors. In particular, autistic people are at increased risk of a range of co-existing mental health and neurodevelopmental conditions including anxiety disorders, depression, sleep disorder, conduct disorder, psychosis and ADHD (Lai et al., 2019; Simonoff et al., 2008) and negative outcomes including suicide attempts (Cassidy et al., 2014). Yet no studies included in this review modelled ATs alongside mental health outcomes.

Autistic people and their families have identified mental health outcomes and interventions as their top priority for future research (Cusack & Sterry, 2016). As Lounds Taylor (2017) has identified, there is a pressing need to clarify what amounts to a “good outcome” for an autistic person. It is especially important to highlight that there should be no assumption that a reduction of ATs equates to a good outcome. Rather, autism should be considered alongside multiple developmental trajectories across various dimensions including diagnostic features, cognition, adaptive skills, mental health and quality of life (Murphy et al., 2016). There has been a growing emphasis on resilience in the autism literature, and AT trajectories can identify changing points on a pathway towards optimal outcomes (Georgiades & Kasari, 2018). There has been a missed opportunity in the literature to date to accurately describe factors that promote resilience (Szatmari, 2018) and the construct of resilience should also be considered to be multi-factorial (Elsabbagh, 2020).

Future studies might fruitfully apply the Research Domain operational Criteria (RDoc) framework to these multiple levels of complexity (Joyce, Kehagia, Tracy, Proctor, & Shergill, 2017). Firstly, while autism can be conceptualised as a diagnostic category, here it may be modelled in terms of continuously distributed ATs, which are also multi-dimensional (including social and RRB). Secondly, measurement can be considered across multiple domains including genetic, physiological, self-report and behavioural, across emotional, cognitive, social and sensorimotor systems. Thirdly, data-driven

approaches should be used to identify relevant subgroups, and their trajectories considered on an individualised basis to track natural development and response to intervention.

Within this framework, statistical methods such as parallel process GMM may be used to investigate the interaction of multiple trajectories. Techniques such as Mixed Membership Trajectory Models (Manrique-Vallier, 2014) lend themselves to examining both latent trajectory classes *and* individuals' probability of changing between these trajectories over time. For instance, it may be that at crucial junctures individuals may deviate from their previous trajectory to "catapult" onto a new trajectory (Georgiades et al., 2017).

To date, chronogeneity has been revealed across intra-individual, inter-individual and inter-group domains of ATs, and has been linked to gender and IQ. Yet this approach to conceptualising autism is only recently emerging (Georgiades et al., 2017; Lombardo et al., 2019) and hence the evidence base is relatively modest. Future studies are needed that apply data-driven, latent class modelling approaches to large, general population datasets, across multiple domains. ATs should be explored via precise and specific measures of autistic social traits, and RRB should be modelled in terms of IS and RSM factors. Both observational measures (eg. ADOS) and parent/self-report (eg. ADI-R; Diagnostic Interview for Social and Communication Disorders [DISCO; Wing et al., 2002]) should be used. Additional measures of functioning, cognition, language, mental health and quality of life should be modelled as much as possible. Genetic and physiological data where available could be fruitfully incorporated, using the RDoC framework as a guide. For example, Polygenic Risk Scores could be investigated in relation to subgroups defined by LCGM (Plomin, Haworth & Davis, 2009). A wide range of demographic factors where available should be included, including ethnicity and a more nuanced modelling of sex/gender that includes trans and non-binary identities. Trajectories should be followed into adulthood, and turning points in trajectories explored in relation to timing and nature of intervention. In short, data need to be broad (large N), deep (multiple axes) and long (substantial follow-up).

This review was limited by the wide heterogeneity of measurement, ages and follow-up periods in the literature, as well as by relatively limited reporting of statistical output. This meant that conclusions are

necessarily rather general and tentative. Due to the timescale and resources available, it was not possible to contact authors to request effect size data where this was not provided. However, this did not lead to the exclusion of studies, and was an exploratory analysis rather than essential to the research questions. Additionally, the authors acknowledge that it would have been optimal to have double-coded study selection and quality, rating agreement, but the resources of the project did not allow for this. The range of methodologies employed across studies added significant complexity. Therefore perhaps the most important recommendation of this review is to call for more complete reporting of AT data, use of multiple measures and greater use of LCGM methods with full transparency of model selection information such that future reviews will be able to weigh the evidence for chronogeneity with more confidence.

Implications

Clinicians should note that autistic social traits appear to decrease over time amongst autistic children. However, within overall autistic traits and within subdomains there appears to be significant variation, and distinct groups of individuals who show increasing, decreasing or stable traits.

Evidence has accumulated demonstrating that autistic traits are distributed throughout the population (Constantino, 2009) and yet in clinical practice, categorical and diagnostic thinking is still the norm. A more dimensional approach to autism could improve our ability to appreciate individual differences, to understand development toward greater or lesser difficulty over time, and to investigate causation and possible treatment approaches. In particular, a “one size fits all” approach to treatment may serve autistic people especially poorly, and individually-tailored approaches based on predictive data are urgently needed (Shih, Patterson, & Kasari, 2016, p.470). To this end, an understanding of latent trajectory approaches to outcome data helps us investigate individual and contextual predictors of individuals’ likelihood to experience change (Georgiades et al., 2017). This in turn helps us develop adapted intervention approaches that take account of individual development (Almirall & Chronis-Tuscano, 2016).

Acknowledgements and conflict of interest

We are grateful to University College London for providing the resources and facilities to make this review possible. There are no sponsors, funders or conflicts of interest to declare.

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.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books9780890425596

Bacon, E. C., Courchesne, E., Barnes, C. C., Cha, D., Pence, S., Schreibman, L., ... & Pierce, K. (2018). Rethinking the idea of late autism spectrum disorder onset. Development and psychopathology, 30(2), 553-569.

Bal, V. H., Kim, S. H., Fok, M., & Lord, C. (2019). Autism spectrum disorder symptoms from ages 2 to 19 years: Implications for diagnosing adolescents and young adults. Autism Research, 12(1), 89-99.

Barbaro, J., & Dissanayake, C. (2017). Diagnostic stability of autism spectrum disorder in toddlers prospectively identified in a community-based setting: Behavioural characteristics and predictors of change over time. Autism, 21(7), 830-840.

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), 3281-3294.

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.

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.

Bieleninik, Ł., Posserud, M. B., Geretsegger, M., Thompson, G., Elefant, C., & Gold, C. (2017). Tracing the temporal stability of autism spectrum diagnosis and severity as measured by the Autism Diagnostic Observation Schedule: A systematic review and meta-analysis. PLoS One, 12(9), e0183160.

Cage, E., Di Monaco, J., & Newell, V. (2018). Experiences of autism acceptance and mental health in autistic adults. *Journal of Autism and Developmental Disorders*, *48*(2), 473-484.

Cassidy, S., Bradley, P., Robinson, J., Allison, C., McHugh, M., & Baron-Cohen, S. (2014). Suicidal ideation and suicide plans or attempts in adults with Asperger's syndrome attending a specialist diagnostic clinic: a clinical cohort study. *The Lancet Psychiatry*, *1*(2), 142-147.

Cassidy, S., Bradley, L., Shaw, R., & Baron-Cohen, S. (2018). Risk markers for suicidality in autistic adults. *Molecular Autism*, *9*(1), 42.

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.

Chawarska, K., Shic, F., Macari, S., Campbell, D. J., Brian, J., Landa, R., ... & Young, G. S. (2014). 18-month predictors of later outcomes in younger siblings of children with autism spectrum disorder: a baby siblings research consortium study. Journal of the American Academy of Child & Adolescent Psychiatry, 53(12), 1317-1327.

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., & Gruber, C. P. (2012). Social responsiveness scale–second edition (SRS-2). *Torrance: Western Psychological Services*.

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.

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.

Cusack, J., & Sterry, R. (2016). Your questions: shaping future autism research. *London: Autistica*.

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.

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.

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. (2020). Linking risk factors and outcomes in autism spectrum disorder: is there evidence for resilience?. *bmj*, *368*.

Elsabbagh, M., & Johnson, M. H. (2016). Autism and the social brain: the first-year puzzle. *Biological psychiatry*, *80*(2), 94-99.

Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. Pediatric research, 65(6), 591.

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

Georgiades, S., & Kasari, C. (2018). Reframing optimal outcomes in autism. *JAMA pediatrics*, *172*(8), 716-717.

Georgiades, S., Szatmari, P., & Boyle, M. (2013). Importance of studying heterogeneity in autism. Neuropsychiatry, 3(2), 123.

Georgiades, S., Boyle, M., Szatmari, P., Hanna, S., Duku, E., Zwaigenbaum, L., … & Smith, I. (2014). Modeling the phenotypic architecture of autism symptoms from time of diagnosis to age 6. *Journal of autism and developmental disorders, 44*(12), 3045-3055.

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.

Goldman, S. (2013). Opinion: Sex, gender and the diagnosis of autism—A biosocial view of the male preponderance. Research in Autism Spectrum Disorders, 7(6), 675-679.

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. Journal of autism and developmental disorders, 37(4), 613.

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.

Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. Nature neuroscience, 9(10), 1218.

Hegarty, J. P., Pegoraro, L. F., Lazzeroni, L. C., Raman, M. M., Hallmayer, J. F., Monterrey, J. C., ... & Hardan, A. Y. (2019). Genetic and environmental influences on structural brain measures in twins with autism spectrum disorder. *Molecular psychiatry*, 1-11.

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.

Hull, L., Mandy, W., & Petrides, K. V. (2017). Behavioural and cognitive sex/gender differences in autism spectrum condition and typically developing males and females. Autism, 21(6), 706-727.

Hull, L., Mandy, W., Lai, M. C., Baron-Cohen, S., Allison, C., Smith, P., & Petrides, K. V. (2019). Development and validation of the camouflaging autistic traits questionnaire (CAT-Q). *Journal of autism and developmental disorders*, *49*(3), 819-833.

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.

Joyce, D. W., Kehagia, A. A., Tracy, D. K., Proctor, J., & Shergill, S. S. (2017). Realising stratified psychiatry using multidimensional signatures and trajectories. *Journal of Translational Medicine*, *15*(1), 15.

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.

Kim, S. H., Bal, V. H., Benrey, N., Choi, Y. B., Guthrie, W., Colombi, C., & Lord, C. (2018). Variability in Autism Symptom Trajectories Using Repeated Observations From 14 to 36 Months of Age. Journal of the American Academy of Child & Adolescent Psychiatry, 57(11), 837-848.

Koenig, K., & Tsatsanis, K. D. (2005). Pervasive developmental disorders in girls. In Handbook of behavioral and emotional problems in girls (pp. 211-237). Springer, Boston, MA.

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.

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

Lai, M. C., Lombardo, M. V., Ruigrok, A. N., Chakrabarti, B., Auyeung, B., Szatmari, P., ... & MRC AIMS Consortium. (2017). Quantifying and exploring camouflaging in men and women with autism. Autism, 21(6), 690-702.

Lai, M. C., Kassee, C., Besney, R., Bonato, S., Hull, L., Mandy, W., ... & Ameis, S. H. (2019). Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. *The Lancet Psychiatry*, *6*(10), 819-829.

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

Livingston, L. A., Shah, P., & Happé, F. (2019). Compensatory strategies below the behavioural surface in autism: a qualitative study. *The Lancet Psychiatry*, *6*(9), 766-777.

Lombardo, M. V., Lai, M. C., & Baron-Cohen, S. (2019). Big data approaches to decomposing heterogeneity across the autism spectrum. *Molecular psychiatry*, *24*(10), 1435-1450.

Loomes, R., Hull, L., & Mandy, W. (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, 56(6), 466-474.

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

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

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

Lounds Taylor, J. (2017). When is a good outcome actually good? Autism, 21, 918-919.

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.

Mandy, W. (2019). Social camouflaging in autism: Is it time to lose the mask? *Autism,* 23(8), 1879-1881.

Mandy, W. P., & Skuse, D. H. (2008). Research review: What is the association between the social-communication element of autism and repetitive interests, behaviours and activities?. Journal of Child Psychology and Psychiatry, 49(8), 795-808.

Mandy, W., & Tchanturia, K. (2015). Do women with eating disorders who have social and flexibility difficulties really have autism? A case series. Molecular autism, 6(1), 6.

Mandy, W., Pellicano, L., St Pourcain, B., Skuse, D., & Heron, J. (2018). The development of autistic social traits across childhood and adolescence in males and females. Journal of Child Psychology and Psychiatry, 59(11), 1143-1151.

Manrique-Vallier, D. (2014). Mixed membership trajectory models. In: Airoldi, E., Blei, D., Erosheva, E et al., eds. *Handbook of Mixed Membership Models and Their Applications.* New York, NY: Chapman Hall: 173-188.

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.

Mitchell, W., & Beresford, B. (2014). Young people with high-functioning autism and Asperger's syndrome planning for and anticipating the move to college: what supports a positive transition?. British Journal of Special Education, 41(2), 151-171.

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.

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.

Müller, E., Schuler, A., & Yates, G. B. (2008). Social challenges and supports from the perspective of individuals

with Asperger syndrome and other autism spectrum disabilities. Autism, 12(2), 173-190.

Murphy, C. M., Wilson, C. E., Robertson, D. M., Ecker, C., Daly, E. M., Hammond, N., ... & McAlonan, G. M. (2016). Autism spectrum disorder in adults: diagnosis, management, and health services development. *Neuropsychiatric disease and treatment*.

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.

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.

Orinstein, A. J., Suh, J., Porter, K., De Yoe, K. A., Tyson, K. E., Troyb, E., ... & Fein, D. A. (2015). Social function and communication in optimal outcome children and adolescents with an autism history on structured test measures. *Journal of autism and developmental disorders*, *45*(8), 2443-2463.

Ozonoff, S., Gangi, D., Hanzel, E. P., Hill, A., Hill, M. M., Miller, M., ... & Iosif, A. M. (2018). Onset patterns in autism: Variation across informants, methods, and timing. *Autism Research*, *11*(5), 788-797.

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.

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.

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.

Rai, D., Culpin, I., Heuvelman, H., Magnusson, C. M., Carpenter, P., Jones, H. J., ... & Pearson, R. M. (2018). Association of autistic traits with depression from childhood to age 18 years. JAMA psychiatry, 75(8), 835-843.

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.

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.

Rivet, T. T., & Matson, J. L. (2011). Review of gender differences in core symptomatology in autism spectrum disorders. Research in Autism Spectrum Disorders, 5(3), 957-976.

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.

Ronald, A., Happé, F., & Plomin, R. (2005). The genetic relationship between individual differences in social and nonsocial behaviours characteristic of autism. Developmental science, 8(5), 444-458.

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.

Rutter, M. (2012). Resilience as a dynamic concept. *Development and psychopathology*, *24*(2), 335-344.

Rutter, M., Bailey, A., & Lord, C. (2003). SCQ: Social Communication Questionnaire (Western Psychological Services, Los Angeles)

Rutter, M., Caspi, A., & Moffitt, T. E. (2003). Using sex differences in psychopathology to study causal mechanisms: unifying issues and research strategies. Journal of child psychology and psychiatry, 44(8), 1092-1115.

Saito, A., Stickley, A., Haraguchi, H., Takahashi, H., Ishitobi, M., & Kamio, Y. (2017). Association between autistic traits in preschool children and later emotional/behavioral outcomes. Journal of autism and developmental disorders, 47(11), 3333-3346.

Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *CARS: The childhood autism rating scale*. 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.

Sher, K. J., Jackson, K. M., & Steinley, D. (2011). Alcohol use trajectories and the ubiquitous cat's cradle: Cause for concern?. *Journal of abnormal psychology*, *120*(2), 322.

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.

Shumway, S., Farmer, C., Thurm, A., Joseph, L., Black, D., & Golden, C. (2012). The ADOS calibrated severity score: relationship to phenotypic variables and stability over time. Autism Research, 5(4), 267-276.

Sigman, M., & Kim, N. (1999). Continuity and change in the development of children with autism. The changing nervous system: Neurobehavioral consequences of early brain disorders, 274-291.

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(8), 921-929.

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.

St Pourcain, B. S., Mandy, W. P., Heron, J., Golding, J., Smith, G. D., & Skuse, D. H. (2011). Links between co-occurring social-communication and hyperactive-inattentive trait trajectories. Journal of the American Academy of Child & Adolescent Psychiatry, 50(9), 892-902.

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.

Sullivan, S., Rai, D., Golding, J., Zammit, S., & Steer, C. (2013). The association between autism spectrum disorder and psychotic experiences in the Avon longitudinal study of parents and children (ALSPAC) birth cohort. Journal of the American Academy of Child & Adolescent Psychiatry, 52(8), 806-814.

Szatmari, P. (2018). Risk and resilience in autism spectrum disorder: a missed translational opportunity?. *Developmental Medicine & Child Neurology*, *60*(3), 225-229.

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.

Turner, M. (1999). Annotation: Repetitive behaviour in autism: A review of psychological research. The Journal of Child Psychology and Psychiatry and Allied Disciplines, 40(6), 839-849.

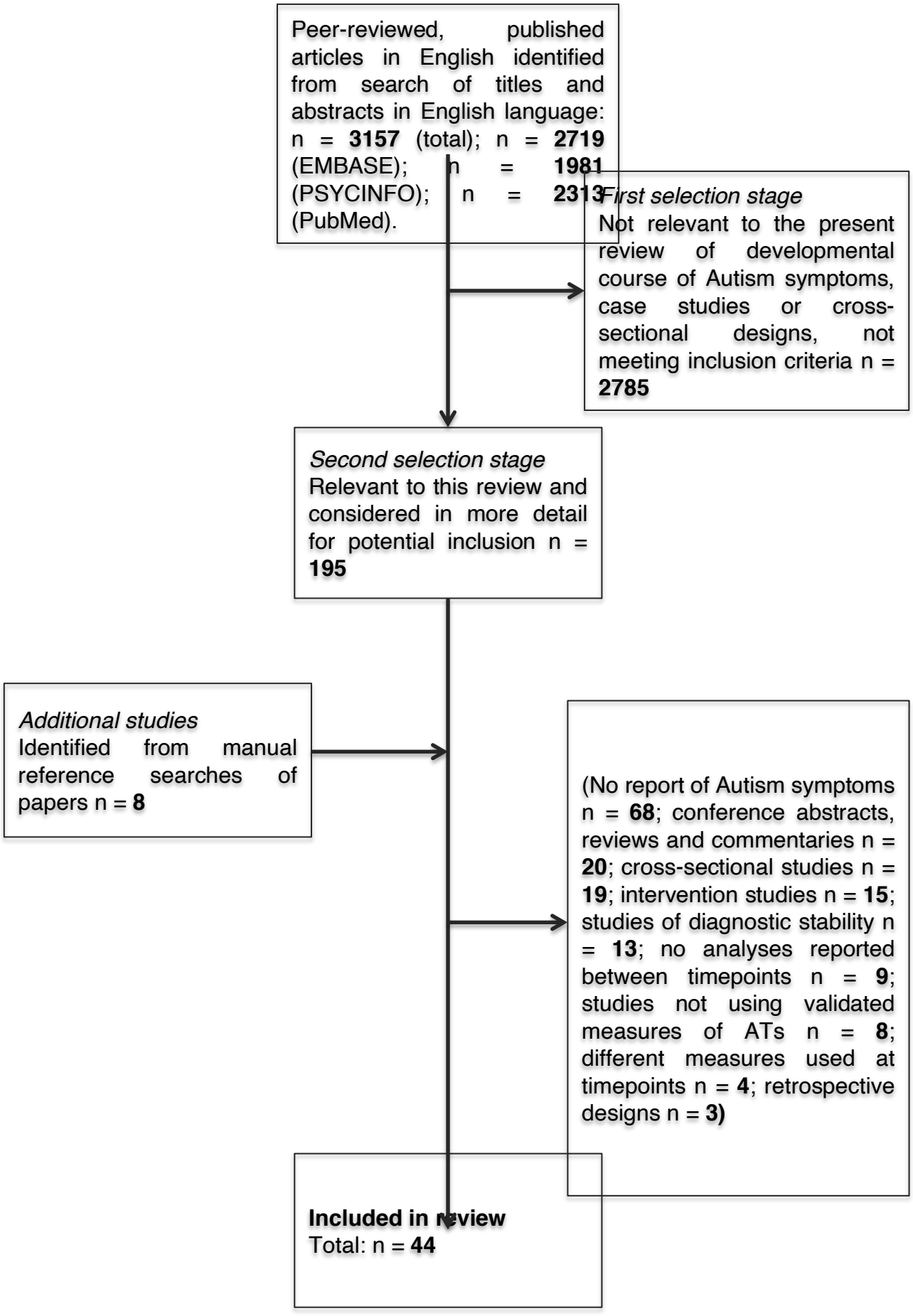
Venker, C. E., Ray-Subramanian, C. E., Bolt, D. M., & Weismer, S. E. (2014). Trajectories of autism severity in early childhood. Journal of autism and developmental disorders, 44(3), 546.

Visser, J. C., Rommelse, N. N., Lappenschaar, M., Servatius-Oosterling, I. J., Greven, C. U., & Buitelaar, J. K. (2017). Variation in the early trajectories of autism symptoms is related to the development of language, cognition, and behavior problems. Journal of the American Academy of Child & Adolescent Psychiatry, 56(8), 659-668.

Wagner, R. E., Zhang, Y., Gray, T., Abbacchi, A., Cormier, D., Todorov, A., & Constantino, J. N. (2019). Autism-related variation in reciprocal social behavior: A longitudinal study. Child development, 90(2), 441-451.

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 disorders: Child and parental contributions. Development and Psychopathology, 27(4), 1045–1057.


Peer-reviewed, published articles in English identified from search of titles and abstracts in English language: n = **3157** (total); n = **2719** (EMBASE); n = **1981** (PSYCINFO); n = **2313** (PubMed).
First selection stage
Not relevant to the present review of developmental course of Autism symptoms, case studies or cross-sectional designs, not meeting inclusion criteria n = **2785**
Second selection stage
Relevant to this review and considered in more detail for potential inclusion n = **195**
Additional studies
Identified from manual reference searches of papers n = **8**
(No report of Autism symptoms n = **68**; conference abstracts, reviews and commentaries n = **20**; cross-sectional studies n = **19**; intervention studies n = **15**; studies of diagnostic stability n = **13**; no analyses reported between timepoints n = **9**; studies not using validated measures of ATs n = **8**; different measures used at timepoints n = **4**; retrospective designs n = **3)**
Included in review
Total: n = **44**