



Cochrane
Library

Cochrane Database of Systematic Reviews

Overall prognosis of preschool autism spectrum disorder diagnoses (Review)

Brignell A, Harwood RC, May T, Woolfenden S, Montgomery A, Iorio A, Williams K

Brignell A, Harwood RC, May T, Woolfenden S, Montgomery A, Iorio A, Williams K.
Overall prognosis of preschool autism spectrum disorder diagnoses.
Cochrane Database of Systematic Reviews 2022, Issue 9. Art. No.: CD012749.
DOI: [10.1002/14651858.CD012749.pub2](https://doi.org/10.1002/14651858.CD012749.pub2).

www.cochranelibrary.com

TABLE OF CONTENTS

ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
SUMMARY OF FINDINGS	4
BACKGROUND	6
OBJECTIVES	8
METHODS	8
RESULTS	12
Figure 1.	13
Figure 2.	16
Figure 3.	18
Figure 4.	19
Figure 5.	20
DISCUSSION	21
AUTHORS' CONCLUSIONS	23
ACKNOWLEDGEMENTS	23
REFERENCES	25
CHARACTERISTICS OF STUDIES	36
ADDITIONAL TABLES	62
APPENDICES	67
Figure 6.	77
Figure 7.	78
Figure 8.	79
Figure 9.	80
Figure 10.	81
Figure 11.	82
Figure 12.	83
Figure 13.	84
Figure 14.	121
HISTORY	122
CONTRIBUTIONS OF AUTHORS	122
DECLARATIONS OF INTEREST	122
SOURCES OF SUPPORT	123
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	123
INDEX TERMS	124

[Prognosis Review]

Overall prognosis of preschool autism spectrum disorder diagnoses

Amanda Brignell^{1,2,3,4}, Rachael C Harwood⁵, Tamara May¹, Susan Woolfenden^{6,7}, Alicia Montgomery⁶, Alfonso Iorio⁸, Katrina Williams^{1,4,9,10}

¹Department of Paediatrics, Monash University, Clayton, Australia. ²Speech and language, Murdoch Children's Research Institute, Parkville, Australia. ³Department of Speech Pathology, Australian Catholic University, Fitzroy, Australia. ⁴Developmental Paediatrics, Monash Children's Hospital, Clayton, Australia. ⁵Community Child Health, Western Health, Melbourne, Australia. ⁶Community Child Health, Sydney Children's Hospital Network, Randwick, Australia. ⁷School of Women's & Children's Health, UNSW Medicine, UNSW Sydney, Kensington, Australia. ⁸Department of Health Research Methods, Evidence and Impact (HEI), McMaster University, Hamilton, Canada. ⁹Department of Paediatrics, The University of Melbourne, Parkville, Australia. ¹⁰Neurodisability and Rehabilitation, Murdoch Children's Research Institute, Melbourne, Australia

Contact: Amanda Brignell, Amanda.brignell@monash.edu.

Editorial group: Cochrane Developmental, Psychosocial and Learning Problems Group.

Publication status and date: New, published in Issue 9, 2022.

Citation: Brignell A, Harwood RC, May T, Woolfenden S, Montgomery A, Iorio A, Williams K. Overall prognosis of preschool autism spectrum disorder diagnoses. *Cochrane Database of Systematic Reviews* 2022, Issue 9. Art. No.: CD012749. DOI: [10.1002/14651858.CD012749.pub2](https://doi.org/10.1002/14651858.CD012749.pub2).

Copyright © 2022 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Autism spectrum disorder is a neurodevelopmental disorder characterised by social communication difficulties, restricted interests and repetitive behaviours. The clinical pathway for children with a diagnosis of autism spectrum disorder is varied, and current research suggests some children may not continue to meet diagnostic criteria over time.

Objectives

The primary objective of this review was to synthesise the available evidence on the proportion of preschool children who have a diagnosis of autism spectrum disorder at baseline (diagnosed before six years of age) who continue to meet diagnostic criteria at follow-up one or more years later (up to 19 years of age).

Search methods

We searched MEDLINE, Embase, PsycINFO, and eight other databases in October 2017 and ran top-up searches up to July 2021. We also searched reference lists of relevant systematic reviews.

Selection criteria

Two review authors independently assessed prospective and retrospective follow-up studies that used the same measure and process within studies to diagnose autism spectrum disorder at baseline and follow-up. Studies were required to have at least one year of follow-up and contain at least 10 participants. Participants were all aged less than six years at baseline assessment and followed up before 19 years of age.

Data collection and analysis

We extracted data on study characteristics and the proportion of children diagnosed with autism spectrum disorder at baseline and follow-up. We also collected information on change in scores on measures that assess the dimensions of autism spectrum disorder (i.e. social communication and restricted interests and repetitive behaviours). Two review authors independently extracted data on study characteristics and assessed risk of bias using a modified quality in prognosis studies (QUIPS) tool. We conducted a random-effects meta-

analysis or narrative synthesis, depending on the type of data available. We also conducted prognostic factor analyses to explore factors that may predict diagnostic outcome.

Main results

In total, 49 studies met our inclusion criteria and 42 of these (11,740 participants) had data that could be extracted. Of the 42 studies, 25 (60%) were conducted in North America, 13 (31%) were conducted in Europe and the UK, and four (10%) in Asia. Most (52%) studies were published before 2014. The mean age of the participants was 3.19 years (range 1.13 to 5.0 years) at baseline and 6.12 years (range 3.0 to 12.14 years) at follow-up. The mean length of follow-up was 2.86 years (range 1.0 to 12.41 years). The majority of the children were boys (81%), and just over half (60%) of the studies primarily included participants with intellectual disability (intelligence quotient < 70). The mean sample size was 272 (range 10 to 8564). Sixty-nine per cent of studies used one diagnostic assessment tool, 24% used two tools and 7% used three or more tools. Diagnosis was decided by a multidisciplinary team in 41% of studies. No data were available for the outcomes of social communication and restricted and repetitive behaviours and interests.

Of the 42 studies with available data, we were able to synthesise data from 34 studies (69% of all included studies; $n = 11,129$) in a meta-analysis. In summary, 92% (95% confidence interval 89% to 95%) of participants continued to meet diagnostic criteria for autism spectrum disorder from baseline to follow-up one or more years later; however, the quality of the evidence was judged as low due to study limitations and inconsistency. The majority of the included studies (95%) were rated at high risk of bias. We were unable to explore the outcomes of change in social communication and restricted and repetitive behaviour and interests between baseline and follow-up as none of the included studies provided separate domain scores at baseline and follow-up. Details on conflict of interest were reported in 24 studies. Funding support was reported by 30 studies, 12 studies omitted details on funding sources and two studies reported no funding support. Declared funding sources were categorised as government, university or non-government organisation or charity groups. We considered it unlikely funding sources would have significantly influenced the outcomes, given the nature of prognosis studies.

Authors' conclusions

Overall, we found that nine out of 10 children who were diagnosed with autism spectrum disorder before six years of age continued to meet diagnostic criteria for autism spectrum disorder a year or more later, however the evidence was uncertain. Confidence in the evidence was rated low using GRADE, due to heterogeneity and risk of bias, and there were few studies that included children diagnosed using a current classification system, such as the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) or the eleventh revision of the International Classification of Diseases (ICD-11). Future studies that are well-designed, prospective and specifically assess prognosis of autism spectrum disorder diagnoses are needed. These studies should also include contemporary diagnostic assessment methods across a broad range of participants and investigate a range of relevant prognostic factors.

PLAIN LANGUAGE SUMMARY

What proportion of preschool aged children diagnosed with autism spectrum disorder retain their diagnosis one or more years later?

Key messages

- Nine out of 10 preschool aged children diagnosed with autism in a research setting may continue to meet diagnostic criteria one or more years later.
- Due to lack of robust evidence, this finding may not be able to be generalised to children outside a research setting, and we were not able to identify any child or research study factors that influenced if a child retained their diagnosis.
- Future research should focus on designing a robust study exploring whether a child retains their autism diagnosis over time in clinical practice and what other factors, if any, may change how likely a child is to retain their diagnosis.

What is autism?

Autism (autism spectrum disorder) is a common neurodevelopmental condition that is generally considered to be lifelong. It is characterised by difficulties in social communication, and restricted interests and repetitive behaviours. How much of a challenge these areas present for each individual is highly variable.

How is autism diagnosed?

Autism is diagnosed by assessing whether an individual meets a set of standardised diagnostic criteria.

In children, an autism diagnostic assessment may involve a paediatrician, child psychiatrist, speech pathologist, occupational therapist and psychologist. One or more of these health professionals may observe and ask questions about a child's social and communication skills, any difficulties in restricted interests and repetitive behaviours, and how they process and respond to sensory information from the world around them. There are diagnostic assessment tools that these professionals can use, alone or in combination, to help make the diagnosis.

What is diagnostic stability, and why is it important?

Diagnostic stability refers to whether an individual retains their diagnosis over time. The diagnostic stability of autism is important to help health professionals, autistic individuals and their families understand how likely it is for a diagnosis of autism spectrum disorder to be lifelong. Additionally, it helps government and community groups to plan what health, education and employment resources are required to support autistic children and their families. Diagnostic stability also helps us to understand whether the characteristics of autistic children and the way that autism spectrum disorder is currently diagnosed influences whether a child continues to meet the criteria for an autism diagnosis over time.

What did we want to find out?

We wanted to find out whether a preschool child who was given a diagnosis of autism spectrum disorder before the age of six years retained their diagnosis at repeat diagnostic assessment one or more years later.

We also wanted to learn more about whether any factors relating to the individual child, the way the child was diagnosed with autism, or the research methods used in the studies, made it more or less likely for the child to continue to meet diagnostic criteria for autism spectrum disorder over time. The factors relating to the individual child included the children's age at the initial and follow-up diagnostic assessments, their intelligence quotient (IQ) score, their ability to complete daily living tasks for a child of their age (adaptive behaviour score), and their ability to communicate with those around them (language score). Factors relating to the way children were diagnosed included the type of tool or criteria used to make the diagnosis, the length of time between diagnostic assessments, and whether the diagnosis was made by a multidisciplinary team. The factors related to the research methods included the year the study was published and the robustness of the evidence.

What did we do?

We searched for studies looking at preschool aged children diagnosed with autism. We then summarised the results, evaluated the evidence and rated our confidence in the evidence based on factors such as study methods and participation.

What did we find?

In total, 49 studies met our inclusion criteria and 42 of these (11,740 children) had data that could be used. The biggest study had 8564 children and the smallest had 11. These studies were from 13 countries, with 16 from the USA. The average age of the children was three years at their first diagnosis and six years at follow-up. The average length of follow-up was 2.86 years.

We found that, in a research setting, nine out of 10 of preschool children diagnosed with autism spectrum disorder may keep their diagnosis one or more years later.

What are the limitations of the evidence?

We have little confidence in the evidence because not all the studies provided data about everything that we were interested in, and the studies were done with different types of people and diagnostic assessments.

For the one in 10 children who no longer met diagnostic criteria for an autism diagnosis at follow-up, we were not able to tell whether they had 'grown out' of their autism because they became more mature over time, or because they had received intervention, or whether the original diagnosis was inaccurate.

How up to date is this evidence?

The evidence is up to date to July 2021.

SUMMARY OF FINDINGS

Summary of findings 1. Summary of findings

Proportion of individuals who have a diagnosis of autism spectrum disorder at baseline and continue not meet diagnostic criteria at follow-up one or more years later

Patient or population: children diagnosed with autism spectrum disorder

Settings: range of settings

Outcomes	Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)	Comments
Proportion with an autism spectrum disorder diagnosis at baseline and follow-up Follow-up: > 12 months	0.92 (0.89 to 0.95)	11,105 (34 studies: 1 intervention trial with 1 arm; 1 RCT ^a ; 2 non-RCTs ^a ; 30 TAU or in the community)	⊕⊕⊕⊕ Low^{b,c}	Limitations (ROB): serious ^b Inconsistency: serious ^c Indirectness: not serious Imprecision: not serious Publication/reporting bias: not serious Effect size: N/A Dose response gradient: N/A Confirmatory evidence: N/A See footnotes below.
Social communication at baseline and follow-up (mean score) Follow-up: > 12 months	See comments			None of the included studies provided separate domain scores at baseline and follow-up
Restricted and repetitive behaviours and interests at baseline and follow-up (mean score) Follow-up: > 12 months	See comments			None of the included studies provided separate domain scores at baseline and follow-up

Definitions of levels of evidence

High: We are very confident that the true prognosis (probability of future events) lies close to that of the estimate

Moderate: We are moderately confident that the true prognosis (probability of future events) is likely to be close to the estimate, but there is a possibility that it is substantially different

Low: Our confidence in the estimate is limited: the true prognosis (probability of future events) may be substantially different from the estimate

Very low: We have very little confidence in the estimate: the true prognosis (probability of future events) is likely to be substantially different from the estimate

CI: Confidence intervals; **N/A:** Not applicable; **RCT(s):** Randomised controlled trial(s); **ROB:** Risk of bias; **TAU:** Treatment as usual.

^aData were taken from the control arm of the study

^bWe downgraded the quality of the evidence by one level for risk of bias due to high risk of bias across most studies: 85% of studies were rated moderate or high risk of bias in study participation, 68% were moderate or high risk of bias in study attrition and 88% were rated moderate or high risk of bias for outcome measurement. Only 5% of studies were rated low risk of bias across all three criteria.

^cWe downgraded the quality of the evidence one level for inconsistency of results (large heterogeneity ($I^2 = 88.71\%$), P value ($P < 0.01$)). The forest plot showed significant variation between point estimates for studies and non-overlapping confidence intervals across many studies. The least and most optimistic point estimates varied considerably (60% to 100%) and each of these estimates were likely to result in different conclusions about the stability of a diagnosis in autism spectrum disorder.

BACKGROUND

Description of the condition

Autism spectrum disorder is a heterogeneous neurodevelopmental disorder. It is currently a clinical diagnosis based on the presence of difficulties in social communication, and restricted interests and repetitive behaviours, which impact on the day-to-day function of the individual, and have been present since early childhood (APA 2013). To make the diagnosis, information is required from more than one setting. A diagnosis of autism spectrum disorder is typically made using criteria from the DSM or the ICD (APA 2013; WHO 1992).

Additionally, three severity levels (requiring support, requiring substantial support and requiring very substantial support) have been introduced for the two main criteria (social communication difficulties; and restricted interests and repetitive behaviours). Children who 'require support' are likely to have difficulty initiating social interactions and have atypical responses to social overtures. The child may also appear to have decreased interest in social interactions. For these children repetitive behaviours and rituals cause significant interference with their daily activities. Children who 'require substantial support' are likely to have marked difficulty with social interactions which are apparent even when allied health supports are in place. These children may also have repetitive behaviours and fixed interests that are obvious to the casual observer and when their interests or repetitive behaviours are interrupted it may cause the child distress. Children who 'require very substantial support' are likely to have great difficulty with verbal and nonverbal social communication that severely impacts their daily functioning. Their repetitive behaviours and fixed interests markedly interfere with their daily activities in all contexts.

A number of other neurodevelopmental conditions are associated with autism spectrum disorder, such as speech and language difficulties, intellectual disability, and attention deficit hyperactivity disorder. In the fifth edition of the Diagnostic Statistical Manual of Mental Disorders (DSM-5; APA 2013), clinicians are encouraged to identify whether individuals have these as co-occurring conditions.

In the past two decades, there have been many changes in the way autistic individuals are diagnosed and cared for. While the DSM-5, published in 2013, now uses the broad term 'autism spectrum disorder' (APA 2013), previous editions of the DSM and the International Classification of Diseases (ICD) have used different criteria and included diagnostic subgroups based on the individual's profile of symptoms (APA 1980; APA 1994; APA 2000; NCHS 2011); see Table 1. In recent years, the United States National Institute of Mental Health has reoriented its focus away from diagnostic categories in mental disorders towards the use of the Research Domain Criteria (RDoC) framework. RDoC aims to address the heterogeneity seen in autism by utilising a biologically-based, rather than symptom-based, research framework. In this approach, dimensions of observable behaviour (e.g. social communication) are integrated with neurobiological, behavioural or environmental components, or both. RDoC contrasts with the categorical approach that is currently being used in most autism research. That is, where the child is described as either having or not having autism.

The diagnostic accuracy and clinical utility of an autism diagnosis is not well described despite the high prevalence of the disorder (CDC 2014; Kim 2011). Current recommendations on how autism spectrum disorder should be diagnosed include a combination of history, observation and application of DSM or ICD criteria, taking into account the overall abilities of the person and ensuring alternative diagnoses are excluded (NICE 2011; Volkmar 2014; WAADF 2012). A number of assessment tools have been published that can assist with diagnosis. These include the Autism Diagnostic Observation Schedule (ADOS; Lord 2000), the Autism Diagnostic Interview – Revised (ADI-R; Le Couteur 2003), the Childhood Autism Rating Scale (CARS; Schopler 1980), and the Diagnostic Interview for Social and Communication Disorders (DISCO; Wing 2002b). It is recommended that these tools are used as part of a multidisciplinary diagnostic assessment, with a team consisting of at least a medical expert, psychologist and speech pathologist, combined with other expertise, if required, depending on the strengths and difficulties of each individual (Randall 2018). In both practice and research, diagnostic process often falls short of this recommendation, and is instead based on an assessment from a sole diagnostician.

The clinical pathway for autistic individuals can be variable, with some individuals showing signs of autism spectrum disorder from as early as one year of age and others being described as having typical development followed by a period of developmental regression or loss of previously acquired skills (Landa 2013). To receive a diagnosis of autism spectrum disorder, symptoms must be present from childhood; however, some individuals with more subtle functional impairment may not receive a diagnosis of autism spectrum disorder until middle childhood, adolescence or adulthood (Van 't Hof 2021), or until demands of the environment exceed the capacity of the individual (APA 2013).

The reported prevalence of autism spectrum disorder diagnoses has increased over the past two decades (Elsabbagh 2012), with estimates from the USA reporting that 1.5% of children aged eight years have been diagnosed with the condition (CDC 2014). Elsabbagh 2012 reported the global median of prevalence estimates of autism spectrum disorder to be 0.62% (range 0.01% to 1.89%). Several factors have been proposed that may have contributed to the increase in prevalence of autism spectrum disorder, including increased community awareness of the condition, administrative factors (e.g. specific funding for a diagnosis of autism spectrum disorder relative to other conditions), a broadening of the criteria informing an autism spectrum disorder diagnosis, and diagnostic substitution with conditions such as intellectual disability (Fombonne 2009; Hansen 2015; King 2009; Wing 2002a). A study from Sweden reported an increase in the diagnosis of autism spectrum disorder over a 10-year period (from 1993 to 2000) but no change in the prevalence of traits of the type seen with autism spectrum disorder (Lundstrom 2015). As such, it remains unclear whether there is a true increase in the prevalence of autism spectrum disorder or whether other factors are influencing the number of people who receive a diagnosis of autism spectrum disorder.

There are numerous causes of autism spectrum disorder, many of which have prenatal roots in genetics and brain development. There are also prenatal, perinatal and postnatal acquired causes. Despite advanced genetic and other medical techniques for investigation, there are still cases for which the underlying

aetiology is unknown. Autism spectrum disorder may have shared aetiological pathways with other neurodevelopmental disabilities, such as intellectual disability, language impairment and attention deficit hyperactivity disorder, and individuals with these conditions may present with overlapping symptoms, behaviours and cognitive deficits that may complicate differential diagnosis and indicate co-occurring conditions that result in impaired functioning (Simonoff 2008). No biological markers for autism spectrum disorder have been identified; hence, autism spectrum disorder is diagnosed based on clinician opinion, through observation of behavioural signs and symptoms, standardised testing and reviewing a combination of parent, guardian, teacher, other informant and self-report questionnaires. There are, however, a cohort of children with an underlying genetic condition, such as Fragile X syndrome and several other genetic syndromes, commonly associated with autism spectrum disorder.

Many clinicians and families believe that autism spectrum disorder is a lifelong disability; however, there is debate in the current literature regarding the permanence of a fitting diagnosis of autism spectrum disorder. Some studies have reported that a significant proportion of previously diagnosed individuals no longer meet the diagnostic criteria for autism spectrum disorder at follow up (Corseello 2013; Daniels 2011; Bopp 2006; Turner 2007), and have also reported a variety of factors found to influence the stability of the diagnosis. Diagnostic stability here refers to whether diagnostic criteria for autism is met at subsequent assessments. Factors relating to diagnostic stability have included: age at diagnosis (Daniels 2011; Turner 2007); milder symptoms of autism spectrum disorder (particularly in the social domain) and higher cognitive scores at two years of age (Turner 2007); the diagnosing clinician, region and a history of regression (Daniels 2011); and maturation, type of diagnosis (autistic disorder versus pervasive developmental disorder — not otherwise specified), and amount of intervention (Bopp 2006). We did not include amount of intervention received in the prognostic factor analysis as many studies did not collect this information, these data can be difficult to compare between studies and our review design was not suited to assessing interventions. Other studies report that autism spectrum disorder can be reliably diagnosed and that few individuals “grow out” of a diagnosis (Barbaro 2016; Guthrie 2013 Jónsdóttir 2007; Ozonoff 2015; Takeda 2005).

Individuals are diagnosed with autism spectrum disorder at varying ages and following different early life experiences. The diagnosis implies that the individual may face challenges that will require intervention and support in addition to that required by their neurotypical peers. These challenges might include difficulty with social relationships and communication, academic difficulties, behavioural difficulties, higher levels of dependence on others, and a poorer quality of life (Howlin 2004; Howlin 2012). The treatment and support that follow a diagnosis of autism spectrum disorder depend on the individual's strengths and difficulties; the parents' or individual's wishes; available and accessible interventions and services; and the individual's functional trajectory from the time of diagnosis. For example, for children in the preschool years, families may choose to pursue a range of different interventions, from complementary to traditional, and their choices may be influenced by what is readily available or promoted in their community (Goin-Kochel 2007; Green 2006). What is currently lacking for preschool-aged children is an evidence base to inform the advice given to

parents about what this diagnosis will mean for their child in the short and long term.

The significant variation in the presentation of autism between individuals and across a broad range of functional domains (e.g. sensory, behavioural and communication) may contribute to the inconsistent evidence about the long-term prognosis in autism spectrum disorder diagnoses. In addition, there may be a number of other factors such as age at diagnosis, presence of intellectual disability, diagnostic subgroup and comorbid diagnoses that contribute to heterogeneity in prognosis.

In recent years there has been increased understanding of neurodiversity and greater recognition of the abilities, differences and strengths of autistic people. There has also been ongoing consideration of the terminology (which is varied) used to refer to autism. In this review we use the term 'autism spectrum disorder' to refer to diagnosis because this is the current diagnostic term stated in the most recent versions of the Diagnostic Statistical Manual of Mental Disorders (DSM) and the International Classification of Diseases (ICD) and this review also primarily focuses on the topic of diagnosis. We use identity first language, including the terms 'autistic children' or 'autistic individuals' to refer to children/individuals (Kenny 2016). An exception is when we present data directly from the included studies where we use the language reported by the study authors. We acknowledge that some people with autism may not wish to reduce their autistic symptoms and do not consider 'no longer meeting diagnostic criteria for autism' a desired outcome. However, it is important to understand the proportions of individuals diagnosed with autism spectrum disorder at baseline who continue to meet diagnostic criteria at follow-up to plan funding, health and community supports, and to assist with prognostication.

Why it is important to do this review

Autism spectrum disorder is a global health issue with far-reaching implications for autistic individuals, their families and support agencies. Substantial economic and social costs are associated with autism spectrum disorder (Ganz 2007; Horlin 2014), with the cost of supporting an autistic individual throughout their lifespan estimated to be around US \$2.4 million (if they have an intellectual disability) or US \$1.4 million (without an intellectual disability) (Beuscher 2014). Support costs may be higher for those diagnosed younger, as an early age of referral or diagnosis has been associated with more severe symptoms, both in autism behaviours and in associated domains such as language and motor skills (Sicherman 2021). In addition to financial costs, an autistic individual may have functional difficulties that result in reduced activity and participation in the community and potential negative impacts on their own and their family's quality of life. As such, there are substantial considerations for policymakers and service providers with regard to the allocation of resources and the planning of future support needs of autistic individuals.

Families, autistic individuals and clinicians require high-quality, reliable information about what proportion of individuals will continue to have a diagnosis of autism spectrum disorder as a crucial first step in understanding diagnostic validity and prognosis. It is also essential information for parents trying to understand their child's strengths and challenges, and plan for their short- and long-term future. Furthermore, information on the proportion of individuals who continue to meet diagnostic criteria

for autism spectrum disorder is important for policymakers and service providers, so they can better plan future support needs for the autistic community. Individuals who do not continue to meet diagnostic criteria for autism spectrum disorder often have other developmental challenges, such as cognitive, language, or attention related conditions that require health and community supports.

There have been three other reviews investigating the stability of autism spectrum disorder diagnoses. One such review, that included a meta-analysis of eight longitudinal studies, found that the proportion of children aged under three years who were still diagnosed with autism disorder at follow-up was 35% if their original diagnosis was PDD-NOS: pervasive developmental disorder (PDD-NOS) and 76% if it was autistic disorder (Rondeau 2010). This study had a number of methodological limitations, such as only searching one database, and it did not assess risk of bias. Another review found that, depending on age, 0% to 30% of children aged less than three years, or 0% to 16% aged less than five years who had been diagnosed with PDD-NOS, no longer met the diagnostic criteria at follow-up, and a similar proportion moved to another diagnostic category (e.g. autistic disorder) (Woolfenden 2012). The proportion that changed diagnosis was higher than many clinicians had anticipated. Variations in the persistence of a diagnosis according to age group, diagnostic subgroup and intelligence quotient (IQ) were also reported. Most studies included in the review were found to be at high risk of bias. The third systematic review (Bieleninik 2017) included 44 studies (n = 40 in a meta-analysis) with 5771 participants, however, this review only included studies that had used a specific measure of autism symptoms (i.e. the ADOS), rather than the full range of tools that are used to diagnose autism. This meant studies that presented data on diagnostic stability using other well known autism diagnostic assessment tools were excluded. Bieleninik 2017 found no significant change in total ADOS scores over time, but a small change was found in the social affect domain over time. No change was reported in the restricted interests and repetitive behaviour domain. In this study, 18% of participants shifted from meeting the cut-off for autism on the ADOS to meeting the cut-off for autism spectrum disorder, however, the overall number meeting criteria for autism spectrum disorder remained unchanged over time. This review did not complete an overall assessment of the quality of the evidence and the search was completed in 2015, so an update of the evidence is indicated.

Much work has been published since these prior reviews. An update is needed to determine whether higher-quality evidence is now available, and to assess whether current information about the proportion of individuals who continue to meet diagnostic criteria for autism spectrum disorder at follow-up is sufficient to inform individually tailored decision-making.

In this Cochrane Review, we included studies in which diagnostic practices reflected current commonly used standards for research or clinical care. Additionally, we investigated whether the baseline diagnostic approach or diagnostic tools used (including age- or ability-modified versions of those tools), alone or in combination, and consistency between tools at baseline and follow-up, contributed to differences in the proportions of children who continued to meet diagnostic criteria for autism spectrum disorder. However, we would not be able to determine definitively whether the individual had 'grown out' of autism spectrum disorder

due to maturation or intervention, or if the original diagnosis was inaccurate. This systematic review provides information for clinicians and families about the likelihood of a preschool-aged child with an autism spectrum disorder diagnosis retaining this diagnosis in one or more years. This review may also serve as an exemplar for prognosis methods.

OBJECTIVES

The primary objective of this review was to synthesise the available evidence on the proportion of preschool children who have a diagnosis of autism spectrum disorder at baseline (diagnosed before six years of age) who continue to meet diagnostic criteria at follow-up one or more years later (up to 19 years of age).

The secondary objectives of this review were to investigate whether there are differences in the proportions of preschool children diagnosed with autism spectrum disorder who maintain a diagnosis at follow-up dependent on:

- use of the different classification systems (i.e. DSM or ICD criteria) and their revisions;
- age;
- language level (verbal/non-verbal; standard score ≤ 70 or > 70);
- intelligence quotient (IQ) (≤ 70 or > 70);
- adaptive behaviour (standard score ≤ 70 or > 70); and
- diagnostic subgroups (Asperger's syndrome/disorder, autistic disorder, childhood autism, pervasive developmental disorder — not otherwise specified, atypical autism, pervasive developmental disorder and autism spectrum disorder).

METHODS

Criteria for considering studies for this review

Types of studies

We included published reports of prospective and retrospective longitudinal studies investigating the prognosis of autism spectrum disorder in preschool aged children that used the same measure to diagnose autism spectrum disorder at baseline and follow-up. Studies were required to have at least one year of follow-up and contain at least 10 participants. The decision to use 10 as the minimum number of participants was made in conjunction with a statistician and is consistent with prior methods used in other studies (e.g. Magiati 2014).

Studies may or may not have included a comparison group observed over the same time period, the characteristics of which were assessed in the same manner. Randomised controlled trials (RCTs) were eligible for inclusion but only data from the control group were extracted. We excluded studies from our review if the follow-up of autistic children or young adults was incidental to another syndrome, or if the outcomes were not appropriately measured (i.e. studies where information on diagnosis at follow-up was not provided).

Types of participants

Participants were diagnosed with autism spectrum disorder, pervasive developmental disorder, pervasive developmental disorder — not otherwise specified, atypical autism, pervasive developmental disorder — unspecified, Asperger's syndrome/disorder, autism, autistic disorder or childhood autism at baseline.

Diagnosis must have been made using a standardised diagnostic tool (see 'Types of outcome measures' below for eligible tools) or by using established diagnostic criteria (e.g. criteria from the third edition (APA 1980), fourth edition (APA 1994), fourth edition-text revision (APA 2000), and fifth edition (APA 2013) of the DSM (DSM-III, DSM-IV, DSM-IV-TR and DSM-5, respectively); or from the ninth revision (WHO 1979) and tenth revision (WHO 1992) of the ICD (ICD-9 and ICD-10, respectively). As per the objectives of this review, only children initially diagnosed before the age of six years and followed up before the age of 19 years were eligible for inclusion. We included children with a dual diagnosis (for example, a diagnosis of Asperger's syndrome/disorder and attention deficit hyperactivity disorder), given the high proportion of autistic children who have co-occurring conditions. We included children with medical aetiologies, such as Fragile X syndrome and tuberous sclerosis, only if these medical conditions occurred with a diagnosis of autism spectrum disorder and were not the focus of the sample. We excluded studies of children with Rett syndrome as this is no longer considered part of autism spectrum disorder in DSM-5. Children with Rett syndrome also have a well-described and different developmental trajectory to children with idiopathic autism. Participants who were in the intervention arm of a randomised controlled trial were excluded in keeping with established prognosis study methods.

Types of prognostic factors

We did not analyse prognostic factors that had been reported and analysed within the included studies in this review.

Types of outcome measures

Our review focused only on diagnostic stability. There are other important outcomes for autistic individuals (e.g. adaptive behaviour); however, reporting on these was beyond the scope of this review.

Primary outcome

The primary outcome was the proportion of preschool children who have a diagnosis of autism spectrum disorder at baseline and who continue to meet diagnostic criteria at follow-up one or more years later.

Diagnosis at follow-up must have been made using DSM or ICD criteria, or a DSM- or ICD-compatible standardised tool. Tools accepted for diagnosis include: the Autism Diagnostic Interview Revised (ADI-R) (Le Couteur 2003), CARS (Schopler 1980), ADOS (Lord 2000), Diagnostic Interview for Social and Communication Disorders (DISCO; Wing 2002b), the Gilliam Autism Rating Scale (GARS; Gilliam 1995), and the Developmental, Dimensional and Diagnostic Interview (3di; Skuse 2004). For each tool, individuals must have met the published cut-off for a diagnosis of autism spectrum disorder. Most studies presented this outcome as a dichotomous variable (i.e. diagnosis or no diagnosis). Additionally, we required the same diagnostic criteria, tools or combination of both to be used at baseline and follow-up. This was to ensure that the type of diagnostic tool or criteria used had minimal impact on whether the individual met the diagnostic criteria for autism spectrum disorder (Randall 2018). Studies were still eligible for inclusion if they used different editions of the same tool (e.g. CARS and CARS-2). If studies did not provide the required data we contacted the authors to request the data.

Secondary outcomes

We assessed the outcomes below, as measured by the diagnostic classification system or diagnostic tool used for the primary outcome, providing the data were available separately. When data were available, we compiled the data and provided a narrative description of the results.

1. Social communication
2. Restricted interests and repetitive behaviours

We included these two secondary outcomes in consideration of the reorientation towards RDoC, with future studies likely to report on clinical presentation/characterisation in more dimensional ways.

We grouped outcome data into three time periods for analysis purposes: short-term (up to two years), medium-term (two to five years) and long-term follow-up (six to 17 years).

We included all outcomes in the [Summary of findings 1](#).

Search method for identification of studies

Electronic searches

We ran the first searches in October 2017 and top-up searches in July 2021. We searched the following electronic databases.

1. MEDLINE Ovid (R) (1946 to June Week 4 2021); searched 5 July 2021.
2. MEDLINE In-Process & Other Non-Indexed Citations Ovid in MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations, Daily and Versions(R) (1946 to July 02, 2021); searched 5 July 2021.
3. MEDLINE Epub Ahead of Print Ovid in MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations, Daily and Versions(R) (1946 to July 02, 2021); searched 5 July 2021.
4. Embase Ovid (1974 to 2021 July 02); searched 5 July 2021.
5. CINAHL Plus EBSCOhost (Cumulative Index to Nursing and Allied Health Literature; 1937 to 6 July 2021).
6. APA PsycINFO Ovid (1967 to June Week 4 2021); searched 6 July 2021.
7. Conference Proceedings Citation Index - Science Web of Science Clarivate (CPCI-S; 1990 to 6 July 2021).
8. Conference Proceedings Citation Index - Social Science & Humanities Web of Science Clarivate (CPCI-SS&H; 1990 to 6 July 2021).
9. *Cochrane Database of Systematic Reviews* (CDSR; 2021, Issue 7) in the Cochrane Library; searched 6 July 2021.
10. Database of Abstracts of Reviews of Effects (DARE; 2015, Issue 2) in the Cochrane Library; searched 12 October 2017. No new content was added to DARE after this issue.
11. Epistemonikos (www.epistemonikos.org; all available years); searched 6 July 2021

We report the strategies used for each source in [Appendix 1](#).

Searching other resources

We identified additional studies by contacting known experts in the field, and by searching the reference lists of relevant reports identified by the electronic searches, including the reference lists

of relevant systematic reviews (reviews of outcomes linked to diagnosis such as language, epilepsy and mortality). We also used Web of Science (Clarivate) to perform forward citation searches of any included studies, and searched the UK National Institute for Health and Research (www.nihr.ac.uk/), and SciELO (Science Electronic Library Online scielo.org/en/) (Appendix 1). These searches were conducted on 17 November 2021.

Data collection and analysis

Selection of studies

Four review authors (AB, RH, SW, KW) independently screened records by title and abstract, removing those that did not meet the criteria listed above. Two review authors were required to screen each record. We advanced records that collected information on diagnosis of autism spectrum disorder, followed individuals for one or more years and had a sample size of more than 10 to the next stage. We then obtained the full texts of potentially relevant reports for review, including those where we considered the inclusion criteria to be unclear. At this stage, authors review authors AB, RH, SW and AM screened reports for type of diagnostic assessment tool or criteria used to confirm that the diagnosis was made, whether diagnosis was made at baseline or prior to the start of the study and that cut-offs for autism spectrum disorder on the relevant tools or criteria were met. If the studies met these criteria on full-text review, they were then advanced for data extraction. Disagreements were resolved by discussion between two initial assessors. If the disagreement could not be resolved a third review author (KW), who was not one of the two initial assessors, acted as an arbiter. The selection process was recorded in a PRISMA diagram (Moher 2009), which was generated using RevMan Web 2020.

Data extraction and management

Using the spreadsheet in Appendix 2, review authors (AB, RH, TM, SW, AM) independently extracted data on: participant characteristics (e.g. language level, mean age, proportion male/female), study characteristics (e.g. country and year of publication), study population type (e.g. clinical or population based) and size, follow-up period, diagnostic classification system (e.g. DSM or ICD criteria) or diagnostic tools used (or both), diagnosis, study attrition, study outcome and change in diagnosis. Two review authors were required to extract data from each record. We also collected information on the version of diagnostic tool or classification system used in each study and noted whether a different version of a tool or classification system was used at baseline and follow-up, as differences in versions of tools or classification systems could impact study findings. In addition, we extracted clinical information needed for prognostic factor analyses (autism spectrum disorder diagnostic groups, IQ, language, adaptive behaviour level, whether diagnosis was multidisciplinary or not, decade of publication, age of inception cohort (i.e. mean age of participants when they entered the study) and at age follow-up), as well as data on duration of follow-up as a possible study factor that influenced the proportion of individuals who remained diagnosed with autism spectrum disorder at follow-up. Disagreements were resolved by discussion. If the disagreement could not be resolved, a third review author who was not one of the two initial assessors, acted as an arbiter. The types of data reported included numbers and percentages.

Assessment of risk of bias in included studies

Two review authors (of AB, RH, AM, SW and TM) independently assessed the risk of bias in each report by examining study participation, study attrition and outcome measurement. Any conflicts that required arbitration were resolved by review author KW. We modified this approach from current literature that addresses the assessment of quality in prognostic systematic reviews (QUIPS; Hayden 2006; Hayden 2013; Hayden 2019). We modified the QUIPS by removing three criteria that were not applicable to this review since we did not extract prognostic factor analyses data from studies (i.e. prognostic factor measurement, study confounding and statistical analysis and reporting). Details regarding the coding of risk of bias are provided in Appendix 3.

Each of the included studies were rated across 18 criteria at low, moderate or high risk of bias. These criteria were then summarised into three domains (study participation, study attrition and outcome measurement) by combining the individual item ratings to provide a risk of bias rating for each summary domain. For study participation, we prioritised ratings for the 'participation in the study by all eligible' and 'study recruitment' criteria. Poor participation and retrospective studies were marked at high risk of bias. Those studies with good participation and prospective recruitment were marked at low risk of bias, which was then graded down to moderate risk of bias if they had one other high risk of bias criterion or three moderate risk of bias criteria across the remainder of the study participation domain criteria. In the study attrition domain, 'loss to follow-up' criteria were prioritised for determining the domain rating. Those with greater than or equal to 85% of the study participants retained at follow-up received a rating of low risk of bias for the study attrition domain rating, except in the case of retrospective studies, where the loss to follow-up is determined by the selection of participants retrospectively, based on data availability. For the outcome measurement domain rating, we prioritised blinding of the study, that is, diagnosticians completing follow-up diagnoses being unaware of the child's diagnostic status at baseline. If the study was unblinded, it was given a rating of high risk of bias for the domain rating. If the blinding was unclear and the remainder of the criteria were at low risk of bias, then the study was rated at moderate risk of bias; however, if blinding was unclear and there was at least one other criteria in the domain rated at moderate or high risk of bias, then that study was rated high risk of bias for the outcome measure domain. Lastly, we provided one overall risk of bias rating for each study, that was either high or low risk of bias. We rated studies to have an overall low risk of bias if all three summary domains were rated at low or moderate risk of bias. Those rated to be at overall high risk of bias were those where one or more summary domains were rated at high risk of bias. As we were not assessing prognostic factors (predictors of outcome) in this review, we did not conduct an analysis of confounders.

If the information required to make an assessment of risk of bias was not available, we emailed the authors of studies published after 2010 to ask for further information, as done in a previous review (see Woolfenden 2012). If the study authors were unwilling or unable to give us the additional information, we documented that we attempted to contact the study authors and marked the risk of bias as unclear. If the minimum necessary information required for inclusion was not available, we excluded the study from the relevant analyses. We used the risk of bias ratings to inform our rating of the quality of the evidence of included studies.

Measures of association

We did not extract measures of association as we were not analysing prognostic factors in this review.

Unit of analysis issues

We collected and analysed study level data from studies included in this review. Some studies used relevant characteristics as eligibility criteria and, as such, reported them at the study level; for example, intelligence, age of participants and duration of follow-up. We extracted the number of individuals with and without a diagnosis of autism spectrum disorder at the time of follow-up and calculated the percentage if it was not presented in the paper. We also extracted these data for prognostic factor analysis. If studies reported data for subgroups (e.g. autistic or autism spectrum disorder; male or female) we calculated a composite mean score, if this was meaningful. Individual participant data meta-analyses were outside the scope of this review.

Dealing with missing data

We included studies that followed up preschool aged children diagnosed with autism spectrum disorder for one or more years after entry, and reported the proportion who still had the same diagnosis at follow-up even if there were missing data. When necessary, we contacted study authors to obtain further information. If authors were unwilling or unable to provide this additional information on missing data, we analysed the available data only (rather than imputing data), as autistic children are very heterogeneous. We documented missing data and considered the possible impact of missing data on each study, in terms of risk of bias, and on the overall review, in terms of quality of evidence. We only included studies when baseline and follow-up data were provided, detailing the number of autistic children, and where the method of diagnosis was explicitly provided. For studies where we could not extract data on the primary outcome, we compared the characteristics of studies included and excluded from the meta-analysis and reported any differences in study samples.

Assessment of heterogeneity

We assessed clinical heterogeneity by comparing important participant factors at a study level, and methodological heterogeneity by comparing the risk of bias of studies, taking into account study participation, participant attrition and outcome measurement factors across the studies (see [Appendix 3](#)). We assessed statistical heterogeneity by inspecting forest plots, and used the I^2 statistics to estimate the total variation across studies due to heterogeneity. When we found high levels of heterogeneity ($I^2 > 50\%$) for the primary outcome, we explored possible sources of heterogeneity using the prognostic factor analyses described below, as required by our secondary objectives.

Reporting bias

We attempted to obtain the results of unpublished studies by contacting study authors. We were able to pool 10 or more studies, so we examined publication bias and other small-study effects, using a funnel plot and Eggers test in STATA ([StataCorp 2019](#)).

Data synthesis

We conducted meta-analyses since data were available from two or more sufficiently homogeneous studies. Currently, it is not

possible to use [RevMan Web 2020](#) to complete the meta-analyses of proportions. Therefore, [StataCorp 2019](#) was used to pool the data, perform statistical analyses and generate forest plots. The Stata command `Metaprop` was used to derive pooled proportions. Since there was heterogeneity in this population, we used a random-effects, generic inverse variance meta-analysis model in [StataCorp 2019](#). We summarised the meta-analysis using the pooled estimate, its 95% confidence interval (CI), and the estimate of between-study variance using I^2 . Where it was not appropriate to combine results using meta-analysis (in the case of heterogeneity or a small number of studies or data extraction problems), we provided a narrative description of the results.

Assessment of quality of the evidence

We used the GRADE framework for prognosis ([Iorio 2015](#)). We judged and reported the overall quality of evidence for all our outcomes using this GRADE approach. Two authors (AB, TM) rated the overall strength of evidence, considering risk of bias, inconsistency, indirectness, imprecision, publication bias, effect size and dose-response gradient (see [Appendix 4](#)). We ranked the quality of the evidence as high, moderate, low or very low (see [Appendix 5](#)). Two review authors (AI, KW) who are experts in prognosis methods when completing GRADE assessments, were consulted if there was any uncertainty.

Prognostic factor analysis

There are a number of potential sources of heterogeneity in studies in autistic individuals, such as the use of different tools or classification systems to diagnose autism spectrum disorder, different diagnostic classifications and types of participants (e.g. with or without language delay or intellectual disability; level of adaptive behaviour). We assessed the primary outcome only in the prognostic factor analyses. We examined the effects of prognostic factors on overall prognosis by visual inspections of confidence intervals for the following.

Study factors:

1. duration of follow-up: short term (up to 2 years), medium term (2 to 5 years), and long term (6 to 17 years) follow-up;
2. decade of publication: 1960 to 1969; 1970 to 1979; 1980 to 1989; 1990 to 1999; 2000 to 2009; 2010 to 2019; 2020 to 2029;
3. studies that use the same version of the diagnostic tool or a different version of the diagnostic tool at baseline and follow-up (e.g. ADOS ([Lord 2000](#)) and ADOS-2 ([Lord 2012](#)));
4. type of diagnostic approach at baseline: multidisciplinary or not multidisciplinary (i.e. included two or more different professionals making the diagnosis).

Child factors:

1. age at baseline: < 2 years; 2 to 3 years; 4 to 6 years;
2. age at follow-up: 2 to 3 years; 4 to 6 years; 7 to 12 years; 13 to 18 years;
3. intelligence: mean IQ \leq 70; mean IQ $>$ 70; or more than 70% of the cohort has IQ \leq 70;
4. adaptive behaviour: mean standard score \leq 70; mean standard score $>$ 70; or $>$ 70% of the cohort has mean standard score \leq 70; and
5. language: $>$ 70% verbal; $>$ 70% non-verbal (i.e. use $<$ 15 words); mean standardised language score $<$ 70; mean standardised

language score ≥ 70 ; or $> 70\%$ of the cohort has mean language score < 70 .

We did not include amount of intervention received in each study in the prognostic factor analysis as many studies did not collect this information and these data can be difficult to compare across studies and hence can be unreliable.

We accepted non-overlapping confidence intervals to indicate a statistically significant difference between the factors that modify overall prognosis. We conducted analyses using [StataCorp 2019](#).

Additional, planned but unused methods for prognostic factor analyses are available in [Appendix 6](#) and [Brignell 2017](#).

Prognostic factor analyses did not reduce heterogeneity of results. Forest plots for the prognostic factor analyses are included in [Appendix 7](#).

Sensitivity analysis

We were unable to complete the planned sensitivity analysis because only two of the 41 included studies with available data for analysis were rated overall at low risk of bias. Planned methods for sensitivity analysis are available in [Appendix 6](#) and [Brignell 2017](#).

RESULTS

Results of the search

The database search (12 October 2017; updated 18 July 2021) identified 44,750 records. We identified an additional 1023 records

from other sources (990 from citation searches, 30 from websites and three from hand searching), making a total of 45,773 records. We removed 20,525 duplicates, leaving 25,248 records. Of these, 24,314 records were excluded based on title and abstract, and 934 records were assessed for eligibility at the full-text level, including the three records that were sourced through reference lists of included studies. We excluded 819 reports. Of these, 21 studies (from 43 reports) appeared to meet inclusion criteria but on close inspection had used different diagnostic tools or criteria at baseline and follow-up. Studies were also excluded where the authors only followed up children who continued to meet diagnostic criteria for autism spectrum disorder (see [Characteristics of excluded studies](#) tables). In addition, 22 studies are awaiting classification. We could not obtain the full text for 20 of these records to enable assessment and two of the records (from one study; [Mosconi 2009](#)) appeared eligible for inclusion but had insufficient information in the full text article to determine eligibility (see [Studies awaiting classification](#) tables). For one study ([Selvakumar 2018](#)), we were unable to determine from the full text whether the sample overlapped with an included study. We contacted the authors of these records but were unable to obtain the information necessary to classify these studies.

In total, 49 studies (from 92 reports) met the inclusion criteria (see [Figure 1](#)). From each of the 12 studies with multiple publications, we selected one as the primary publication. The primary publication was the one that best represented the study or had available data. [Table 2](#) shows the primary studies with multiple publications.

Figure 1. Study flow diagram.

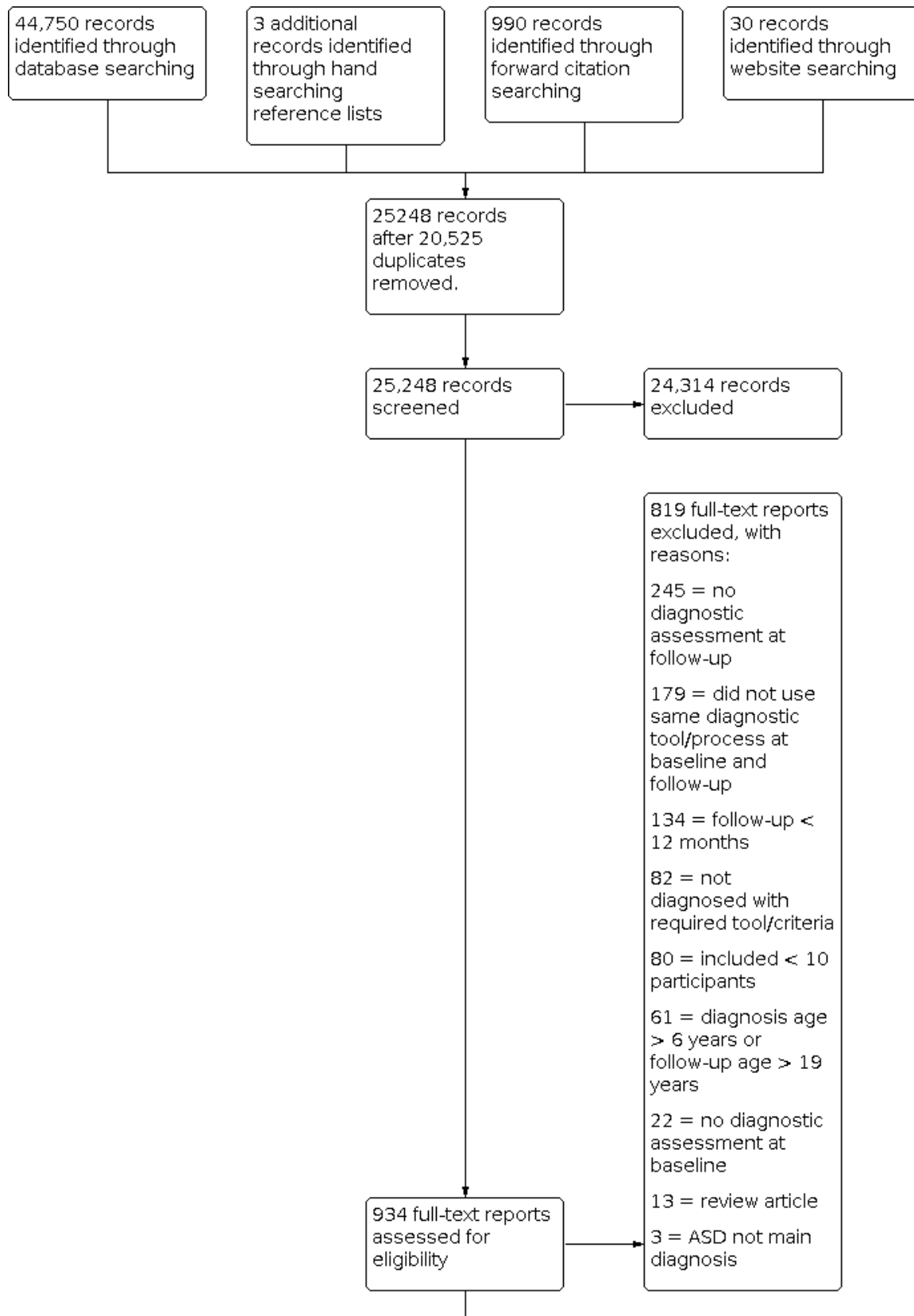
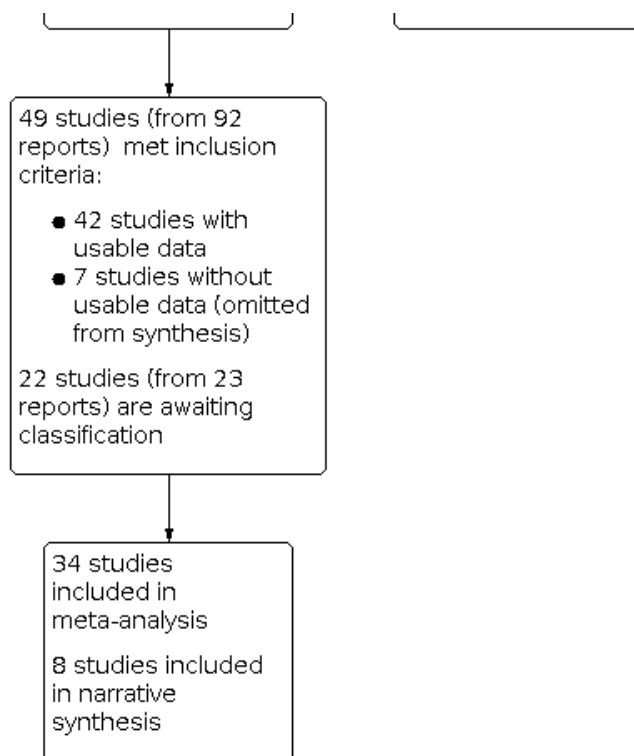


Figure 1. (Continued)



Included studies

Forty-nine studies collected information on autism spectrum disorder diagnosis at baseline and follow-up, using the same diagnostic criteria and tools, and were eligible for inclusion. Of these studies, seven met the inclusion criteria, but data relevant to our aims could not be extracted, and therefore their data have not been included in the review (Anderson 2009; Dietz 2007; Gabriels 2007; Lombardo 2015; Naigles 2016; Neuhaus 2016; Martin-Borreguero 2021). The primary reasons data for these studies were not extractable were: (1) the cohort included some participants that did not have autism spectrum disorder and the authors did not present data separately for those diagnosed with autism spectrum disorder; and (2) mean scores on a diagnostic tool (e.g. ADOS or CARS) were reported but not the diagnostic status at an individual participant level. We contacted the authors for these data but were not able to access any information beyond what was available in the published papers. Where available, we extracted relevant data and included it in the [Characteristics of included studies](#) tables. For studies where we could not extract data, we provide specific reasons in these tables.

Of the remaining 42 studies, 34 (N = 11,129; 81% male) had data that could be included in a meta-analysis (see [Appendix 8](#)), and eight studies (N = 537; 81% male) were suitable for narrative synthesis (Bopp 2006; Chu 2017; DeWaay 2010; Thomas 2009; Haglund 2020; Rivard 2019; Smith 2019; Szatmari 2021). In all eight studies, an acceptable diagnostic tool was used (e.g. ADOS, CARS; GARS; ADI-R); however, the authors reported scores on these tools, rather than the proportion of children diagnosed with autism spectrum disorder at baseline who continued to meet diagnostic criteria at follow-up.

Of the 42 studies with data extracted, 21 (50%) studies were published more than eight years ago. The sample size ranged from 11 (Sheinkopf 1998) to 8564 (Wu 2016) participants, with a mean of 272 and median of 43. Twenty-nine studies (69%) used one diagnostic assessment tool, 10 studies (24%) used two tools, and three studies (7%) used three or more tools. Tools used to make a diagnosis included the Autism Diagnostic Observation Schedule (n = 17), the Childhood Autism Rating Scale (n = 13), the Autism Diagnostic Interview (n = 6), and the Gilliam Autism Rating Scale (n = 2). Thirteen of the 32 studies (41%) that presented relevant data used a multidisciplinary team make the diagnosis. Forty studies (95%) included children with a diagnosis of autism spectrum disorder, two studies (5%) included children with a diagnosis of autistic disorder, and zero studies included children with a diagnosis of childhood autism or Asperger's syndrome. Four studies (10%) were from a population base and 30 (71%) were prospective in design. Nineteen of 32 studies (60%) that reported on IQ primarily included participants with intellectual disability (IQ < 70). The mean age of participants at baseline was 3.19 years (range 1.13 to 5.0 years) and at follow-up was 6.12 years (range 3.0 to 12.14 years). The mean length of follow-up was 2.86 years (range 1 to 12.41 years). Studies were conducted in the following countries: USA (n = 16), Canada (n = 9), Italy (n = 4), Sweden (n = 3), UK (n = 2), China (n = 1), Denmark (n = 1), Japan (n = 1), France (n = 1), India (n = 1), Norway (n = 1), Switzerland (n = 1) and Taiwan (n = 1). Funding support was reported by 30 studies, 12 studies omitted details on funding sources and two studies reported no funding support. Of those that reported their funding source, 12 reported funding from a government organisation, two reported funding from a non-government organisation or charity, two reported funding from a university, and the remaining 12 reported funding from a combination of these sources. None were industry funded. We provide a more detailed description of the

characteristics of the included studies in the [Characteristics of included studies](#) tables and in [Table 3](#). We present a summary of findings for studies included in the meta-analysis in [Summary of findings 1](#).

The characteristics of the included studies were similar to those that had data not suitable for synthesis in some areas but not in others (see [Appendix 8](#) for details). Studies not included in the meta-analysis included: a higher proportion of participants with intellectual disability; a higher proportion of children who did not have a multidisciplinary team make the diagnosis, and a higher proportion of children that only used one tool to make the diagnosis. Children not included in the meta-analysis were, on average, around 10 months older at baseline and were followed up around 1.7 years longer. Studies in the meta-analysis had a higher proportion of studies that were retrospective, a higher proportion that were published more than five years ago and a higher proportion that were derived from a population sample.

Excluded studies

In total, 811 reports were excluded at full-text level (see [Figure 1](#) for exclusion reasons). Of these, 21 studies (from 43 reports) appeared eligible for inclusion but on closer inspection did not meet criteria. Eighteen studies did not use the same diagnostic criteria, tools or combination of both at baseline and follow-up, two studies only included data from the children who retained their diagnosis over time, and one study did not include a baseline diagnostic assessment. Three studies only included children who continued to meet diagnostic criteria for autism spectrum disorder at follow-up, and the authors could not provide data on the numbers of children who did not continue to meet diagnostic criteria for autism spectrum disorder. Three studies only diagnosed children at follow-up. One study followed children for less than 12 months and in one study children were > six years at baseline. See characteristics of [Excluded studies](#) tables for further details.

Studies awaiting classification

We were not able to obtain the full text for 22 studies and had inadequate detail from the abstract or full text to determine

whether they were eligible for inclusion. We made multiple attempts to obtain these data (including contacting authors via email and extensive searches through libraries and online sources). A large proportion of these studies were conference abstracts ($n=9$) where data had not been published beyond the abstract. We were unable to obtain data or essential information from the authors of four studies to determine eligibility ([Millikovsky-Ayalon 2012](#); [Muratori 2002](#); [Perucchini 2005](#); [Selvakumar 2018](#)).

Risk of bias assessment of included studies

We assessed the risk of bias across 18 criteria for 41 of the 42 included studies with usable data, using a modified version of the QUIPS tool ([Hayden 2006](#); [Hayden 2013](#)). One study, [Santocchi 2012](#), had only a conference abstract available with data to allow inclusion in the meta-analysis but insufficient information for risk of bias rating. We summarised the 18 criteria into three domain risk of bias ratings (study participation, study attrition and outcome measurement), as well as giving an overall risk of bias rating for each study. The overall risk of bias was rated at high for 39 of 41 studies (95%) and low for two studies (5%). In total, only one of 41 studies (2.4%) was rated at low risk of bias across all three summary domain criteria; one (2.4%) was rated at low risk of bias for two domains, and moderate for the third domain; and 17 (41.5%) were rated at high risk of bias for all three domains. The remaining 22 studies (53.7%) had a high risk of bias rating for at least one domain with the remaining two domains rated at a combination of low, moderate or high risk of bias. For risk of bias in the study participation criteria five studies were rated low, eight were rated moderate, and 28 were rated high. Study attrition criteria risk of bias was rated low in 11 studies and high in the remaining 30. For risk of bias in the outcome measurement criteria, four studies were rated low, nine were rated moderate and 28 were rated high. [Figure 2](#) presents information about our risk of bias judgements for each of the three domains for each included study with useable data, and [Figure 3](#) provides a summary of the judgements across each domain. [Figure 4](#) shows the detail of each rating across the 18 domains, and [Appendix 9](#) presents the supporting evidence for each judgement in detail.

Figure 2. Risk of bias ratings on the QUIPS tool (40 studies). Green is low risk of bias, orange is moderate risk of bias and red is high risk of bias. Summary risk of bias ratings for provided for each QUIPS domain (i.e.study participation, study attrition, outcome measurement). See [Appendix 9](#) for a figure showing all criteria that were rated for each domain. Studies were rated to have an overall low risk of bias if all three summary domains were

rated low or moderate risk of bias. Studies were rated to have an overall high risk of bias if all three summary domains were rated low or moderate risk of bias.

	Study Participation	Study Attrition	Outcome Measures
Baghdadli 2012	+	+	+
Benedetto 2021	+	+	+
Bopp 2006	+	+	+
Brian 2016	+	+	+
Chu 2017	+	+	+
Demb 1989	+	+	+
DeWaay 2010	+	+	+
Eaves 2004	+	+	+
Elmose 2014	+	+	+
Flanagan 2010	+	+	+
Freeman 2004	+	+	+
Gillberg 1990	+	+	+
Giserman-Kiss 2020	+	+	+
Gonzales 1993	+	+	+
Haglund 2020	+	+	+
Hinnebusch 2017	+	+	+
Kim 2016	+	+	+
Klintwall 2015	+	+	+
Mahli 2011	+	+	+
Moore 2003	+	+	+
Moss 2008	+	+	+
Ozonoff 2015	+	+	+
Paul 2008	+	+	+
Qian 2018	+	+	+
Rivard 2019	+	+	+
Robain 2020	+	+	+
Sheinkopf 1998	+	+	+
Smith 2019	+	+	+
Soke 2011	+	+	+
Solomon 2014	+	+	+
Solomon 2016	+	+	+
Spjut Jansson 2016	+	+	+
Sullivan 2010	+	+	+
Szatmari 2021	+	+	+
Takeda 2007	+	+	+
Thomas 2009	+	+	+
Venker 2014	+	+	+
Wu 2016	+	+	+
Zappella 1990	+	+	+
Zappella 2010	+	+	+
Zwaigenbaum 2015	+	+	+

Figure 3. Risk of bias graph: review authors' judgements about each risk of bias item for each included study presented as percentages across all included studies (41 studies).

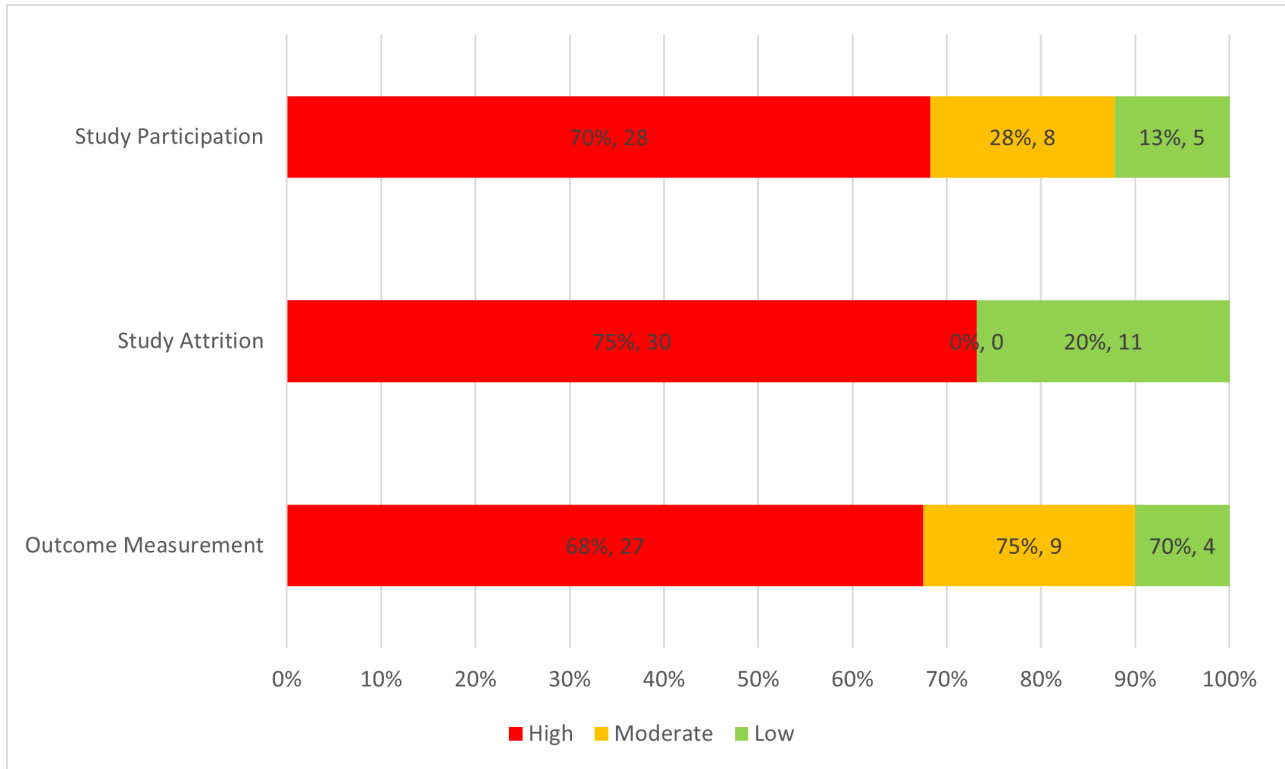


Figure 4. Risk of bias ratings for each included study for each of the 18 criteria. Red indicates high, orange indicates moderate, green indicates low and yellow indicates unclear risk of bias.

	Sample Described	Description of sampling frame	Description of baseline study sample	Description of inclusion or exclusion criteria	Adequacy of participation in study by all eligible	Recruitment	Loss to follow up	Descriptive attempts to collect information on those lost to follow up	Reasons for loss to follow up provided	Reasons for loss to follow up linked to outcome	Description of lost to follow up participants	Analysis: important differences between lost to follow up and non-lost to follow up participants in the study	Blinding	Clear definition of diagnosis provided at follow-up	Same diagnosis outcome tool for all	Valid and reliable tool	Method and setting of outcome measurements same for all participants	Completeness of outcome measure
Baghdadi 2012	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Benedetto 2021	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bopp 2006	+	+	+	+	+	+	+	N/A	N/A	N/A	N/A	N/A	+	+	+	+	+	
Brian 2016	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Chu 2017	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Demb 1989	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DeWaay 2010	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Eaves 2004	+	+	+	+	+	+	+	N/A	N/A	N/A	N/A	N/A	+	+	+	+	+	
Elmose 2014	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Flanagan 2010	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Freeman 2004	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Gillberg 1990	+	+	+	+	+	+	+	N/A	N/A	N/A	N/A	N/A	+	+	+	+	+	+
Giserman-Kiss 2020	+	+	+	+	+	+	+	N/A	N/A	N/A	N/A	N/A	+	+	+	+	+	+
Gonzales 1993	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Haglund 2020	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Hinnebusch 2017	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Kim 2016	+	+	+	+	+	+	+	N/A	N/A	N/A	N/A	N/A	+	+	+	+	+	+
Klintwall 2015	+	+	+	+	+	+	+	N/A	N/A	N/A	N/A	N/A	+	+	+	+	+	+
Mahli 2011	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Moore 2003	+	+	+	+	+	+	+	N/A	N/A	N/A	N/A	N/A	+	+	+	+	+	+
Moss 2008	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ozonoff 2015	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Paul 2008	+	+	+	+	+	+	+	N/A	N/A	N/A	N/A	N/A	+	+	+	+	+	+
Qian 2018	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Rivard 2019	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Robain 2020	+	+	+	+	+	+	+	N/A	N/A	N/A	N/A	N/A	+	+	+	+	+	+
Sheinkopf 1998	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Smith 2019	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Soke 2011	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Solomon 2014	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Solomon 2016	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Sjput Jansson 2016	+	+	+	+	+	+	+	N/A	N/A	N/A	N/A	N/A	+	+	+	+	+	+
Sullivan 2010	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Szatmari 2021	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Takeda 2007	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Thomas 2009	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Venker 2014	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Wu 2016	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Zappella 1990	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Zappella 2010	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Zwaigenbaum 2015	+	+	+	+	+	+	+	N/A	N/A	N/A	N/A	N/A	+	+	+	+	+	+

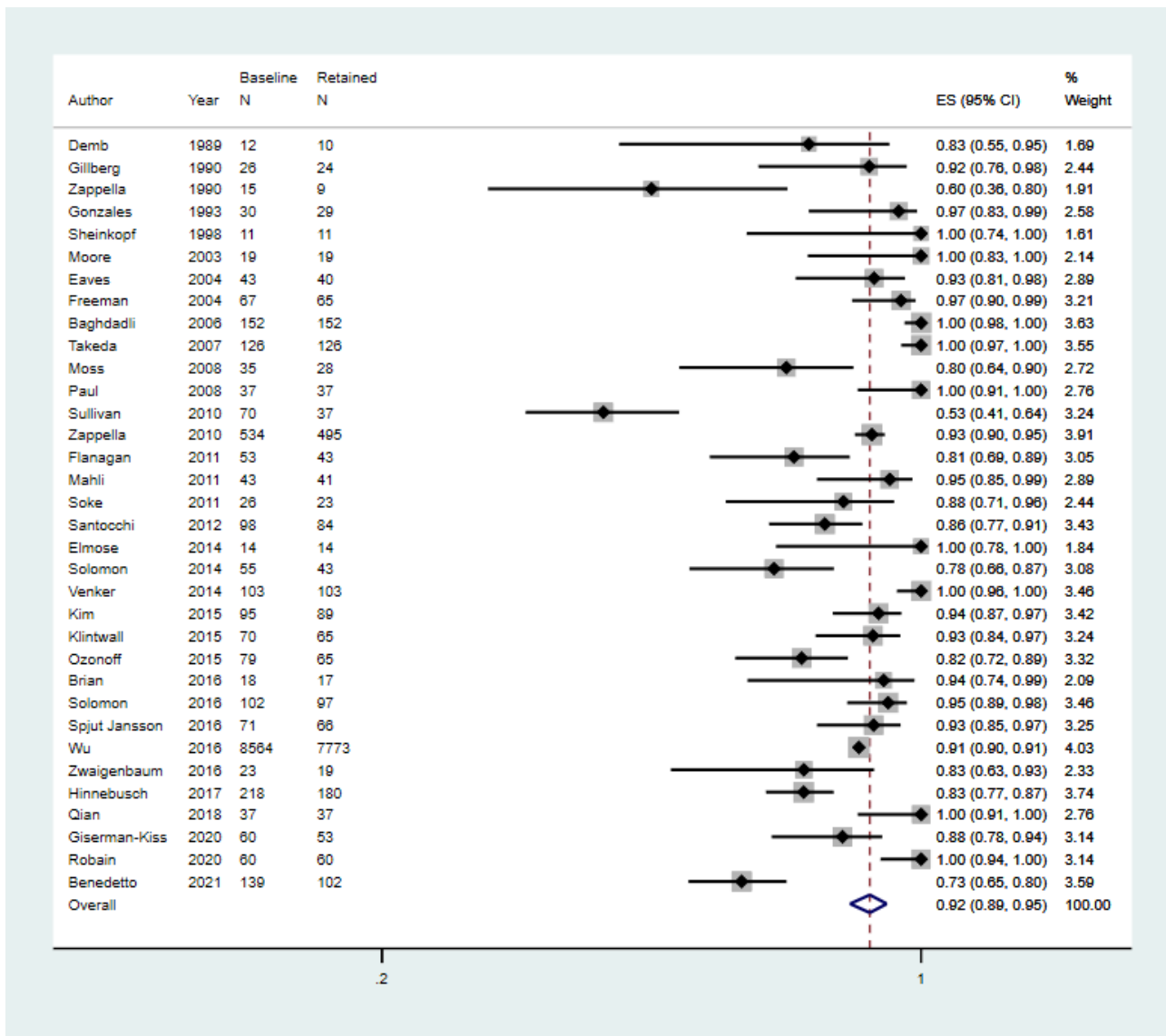
Findings

Primary outcomes

Thirty-four studies (N = 11,129) provided sufficient data on the primary outcome to be included in our meta-analysis (Figure 5). The pooled proportion of children who continue to meet diagnostic criteria for autism spectrum disorder from the 34 included studies was 0.92 (95% CI 0.89 to 0.95; range 0.36 to 1.0), with a high inconsistency index ($I^2 = 88.71\%$), indicating substantial heterogeneity across studies. The diagnosis of autism spectrum disorder was generally stable, with the majority of children who continue to meet diagnostic criteria for autism spectrum disorder from baseline to one or more year(s) follow-up. Overall, we judged

the quality of the evidence to be low on the outcome of the proportion who continued to meet diagnostic criteria for autism spectrum disorder at follow-up, based on GRADE (see [Summary of findings 1](#) for the primary outcome). We downgraded the evidence one level for high risk of bias because 5% of studies were rated low risk of bias across the main three risk of bias criteria (85% of studies were rated moderate or high risk of bias in study participation; 68% in study attrition and 88% in outcome measurement). We downgraded the evidence one level for inconsistency because there was large variability in point estimates which ranged from 60 to 100% who continued to meet diagnostic criteria for autism spectrum disorder, confidence intervals were non-overlapping and there was a high I^2 statistic.

Figure 5. Forest plot of proportion of children that retained their autism diagnosis *Footnote*
CI: confidence interval; ES: effect size; N: number in sample



We assessed publication bias by visual examining asymmetry using a funnel plot that was produced in STATA (StataCorp 2019; Appendix 10). The funnel plot did not show any evidence of reporting bias (i.e. large studies did not provide a different conclusion to

small studies); however, the funnel plot should be interpreted with caution. The primary outcome presented by studies was the proportion that continued to meet diagnostic criteria for autism spectrum disorder, which is likely to have a ceiling effect of

proportions (i.e. studies could not exceed 100%). Furthermore, conventional funnel plots have been found to be a less accurate method of assessing publication bias in proportion studies (Hunter 2014). Egger's test of small-study effects was performed in STATA (StataCorp 2019); the results indicated no evidence of small study effects: $z = -0.05$, $P = 0.59$.

For those eight studies not included in the meta-analysis, the mean sample size of participants was 67 (range 13 to 272), with 81% male. Participants had a baseline mean age of 3.81 years and were followed up for an average of 4.24 years. Three studies presented scores over two time points on the ADOS (Chu 2017; Haglund 2020; Szatmari 2021), two studies on the CARS (Bopp 2006; Thomas 2009), two on the GARS (DeWaay 2010; Rivard 2019), and one on the ADI-R (Smith 2019). On all four tools, higher scores indicate more autism spectrum disorder characteristics. In all but one study (Haglund 2020), the mean scores reported in these studies decreased from baseline to follow-up, indicating that in most cases, the characteristics of autism spectrum disorder reduced over time. On the ADOS, scores decreased in two studies from 7.68 (SD 1.68) to 6.81 (SD 2.62) (Szatmari 2021) and 16.2 (SD 4.7) to 13.9 (SD 3.9) (Chu 2017). On the CARS, scores decreased from 36.14 (SD 5.95) to 34.34 (SD 7.9) (Bopp 2006) and from 35.76 (SD 9.55) to 30.39 (SD 8.11) (Thomas 2009). On the GARS scores decreased in two studies from 86.13 (SD 14.22) to 84.28 (21.19) (Rivard 2019) and 46.46 (SD 24.69) to 33.15 (SD 17.24) (DeWaay 2010). On the ADI-R, scores decreased from 31.5 (SD 4) to 14.9 (SD 10.8) (Smith 2019). Only one study showed a slight increase in scores on the ADOS from 13 (SD 4.5) to 13.1 (SD 5.3) (Haglund 2020).

Secondary outcomes

We were not able to extract any usable data on the two domains (social communication; restricted interests and repetitive behaviours) separately. None of the included studies repeated the same tool (and version) at baseline and follow-up or presented scores separately for the two domains.

Prognostic factor analyses

We conducted prognostic factor analyses to investigate the association between participant and study characteristics and the proportion of children who continued to meet diagnostic criteria for autism spectrum disorder at follow-up. There was no significant association between each of the following factors and the proportion of children who continued to meet diagnostic criteria for autism spectrum disorder: study design factors including duration of follow-up, decade of publication, whether the child had a multidisciplinary diagnosis at baseline; and child factors including age at baseline and follow-up, intelligence, level of adaptive behaviour, language ability (Table 4).

Prognostic factor analyses did not reduce heterogeneity of results. Forest plots for the prognostic factor analyses are included in Appendix 7.

Sensitivity analyses

The planned sensitivity analysis was to compare low versus high risk of bias studies; however, this analysis could not be performed as only two of the 41 studies were rated low risk of bias and therefore the analysis would not be valid.

DISCUSSION

Summary of main results

Our systematic review provides evidence on overall prognosis of preschool children who have been diagnosed with autism spectrum disorder, focusing on the proportion who continue to meet diagnostic criteria for autism spectrum disorder one or more years later. It also provides an exemplar for the implementation of systemic review methods in prognosis. Our review included 42 studies with extractable data, of which, 34 were suitable for meta-analysis. The results suggest that overall, approximately 92% of preschool children diagnosed with autism spectrum disorder keep their diagnosis at a second diagnostic assessment one or more years later, however, the quality of evidence was low, due to high heterogeneity and risk of bias. Furthermore, the proportion of individuals who continue to meet diagnostic criteria for autism spectrum disorder one or more years later varied widely between studies. Most participants in the included studies were aged two to three years old at baseline, were followed up for one to two years, and were studied in the last decade.

We did not identify any child factors (such as age, intelligence quotient, adaptive function, language), study factors (such as duration of follow-up) or any other diagnostic process factors (such as multidisciplinary team conducting the diagnosis, or changes to diagnostic criteria over time) that were associated with continuing to meet diagnostic criteria for autism spectrum disorder at follow-up. As described, there were a majority of studies in one group for age at baseline, year of publication and one to two years of follow-up, which could reduce the opportunity to identify between-group differences, if they exist. Furthermore, a number of studies reported perfect diagnostic accuracy (Baghdadli 2012; Elmoose 2014; Guthrie 2013; Moore 2003; Paul 2008; Sheinkopf 1998), which means there may have been ceiling effects, resulting in biased inferences (Šimkovic 2019).

Some clinically important factors were not well reported: of the 34 studies included in the meta-analysis, 26 reported data on intelligence quotient, 13 reported data on adaptive function and only nine reported data on language level. Additionally, many of the factors interacted; for example, children with an older age at baseline were linked to a longer follow-up duration and earlier decade of study publication. Of note, studies often did not report the inception year of their cohort or other important clinical information, such as scores for autism spectrum disorder diagnostic tools and tools used to measure cognition, language skills and adaptive function. Without these data we are unable to tailor prognostic information for parents and clinicians to inform decision-making. Some studies, such as Sutera 2010, have looked closely at children who are no longer diagnosed with an autism spectrum disorder over time. This systematic review was not able to explore factors that have been identified but were not reported in the included studies.

Risk of bias was high across the included studies, mainly due to the number of retrospective studies, unblinded studies, inadequate participation in the study by all eligible, and high proportion of participants lost to follow-up. There was only one of 41 studies able to be rated for risk of bias that was graded low at all three risk of bias domains (study participation, study attrition and outcome measurement). Seventeen studies were graded high in all three domains and a further 13 were graded high in two of the three

domains. The remainder had moderate or high risk of bias in at least one domain. Study attrition was the domain with the greatest number of studies at low risk of bias, with 11 of 41 rated low risk of bias.

There was heterogeneity across the studies in both clinical and methodological aspects. This included type of study, sample size, criteria or tool used for diagnosis, year of publication, duration of follow-up, and child factors (e.g. IQ). We carefully designed our inclusion criteria to minimise the effects of the heterogeneity on our end results. However, despite rigorous inclusion criteria, heterogeneity was high. The high heterogeneity and risk of bias impacted on the quality of the evidence available, ultimately resulting in the findings of our review needing to be interpreted with caution.

Recently, the United States National Institute of Mental Health has reoriented its focus away from diagnostic categories in mental disorders towards the use of the Research Domain Criteria (RDoC) framework. This research framework focuses on the dimensions of functioning that underlie human behaviour. It is not yet known whether the RDoC approach is changing clinical assessment and diagnosis, or service and funding access requirements. This will not impact already published studies that have used autism spectrum disorder diagnostic labels, but it may influence the way future studies are structured and the types of outcomes they report on, particularly those studies conducted in the USA. While the current study explored categorical diagnosis, future studies could explore the stability of RDoC domains related to autism.

Quality of evidence available

Overall, there was low quality of the evidence for the stability of a preschool autism spectrum disorder diagnosis. The quality of the evidence was downgraded because most (between 68% to 85%, depending on the criteria) studies were rated moderate or high risk of bias. There was also high inconsistency across studies with a substantial I^2 value and wide variance of point estimates across studies with non-overlapping confidence intervals. The lowest and highest point estimates varied considerably (60% to 100%) and each of these estimates were likely to result in different conclusions about the stability of a diagnosis in autism spectrum disorder. We considered that there was no serious indirectness, imprecision, or publication bias present. See [Summary of findings 1](#) for further details.

Strengths and weaknesses of the review

Our review has many strengths. A protocol paper was written and published prior to commencing the review, with clearly defined selection criteria ([Brignell 2017](#)), thus ensuring that decisions made in the process of conducting the review were not data driven. We conducted an extensive search of the literature using multiple electronic databases and to make sure that all relevant studies were identified and a broad range of literature was identified, including grey literature. We also sought additional papers by contacting known experts in the field, searching the reference lists of relevant publications, performing forward citation searches of any included studies, and searching other relevant sites such as those of the UK National Institute for Health and Research, and SciELO (Science Electronic Library Online). We also included studies in language other than English to ensure the findings were as generalisable as possible and that all possible studies were included in the results.

Data requirements meant fewer studies than were included in the review could be included in the meta-analysis, which limited the reliability of the meta-analysis. To compensate for this we used both meta-analysis and narrative synthesis. This allowed us to carefully interpret additional evidence that was not suitable for meta-analysis, which strengthened the validity of our overall findings. We also judged the overall quality of the available evidence. We methodically defined characteristics potentially related to heterogeneity and explored the impact of these with prognostic factor analysis. It would have been valuable to include the amount of intervention as a prognostic factor however, this information was not presented consistently or reliably by included studies, so we did not feel this analysis would be valid. We planned to conduct sensitivity analyses; however, we judged only two of 41 studies at overall low risk of bias, and therefore we were unable to conduct the sensitivity analysis.

Currently, a standardised tool to assess risk of bias for overall prognosis studies does not exist. We utilised an adapted approach from that reported in the literature ([Hayden 2006](#); [Hayden 2013](#)), but subsequently identified issues with assessing retrospective studies. To mitigate this, we further adapted the risk of bias tool, so that retrospective studies were scored at high risk of bias for study attrition. Our methodological approach to rating risk of bias ensured the included studies were rated for bias consistently, which allowed assessment of the quality of the evidence. Further development of a robust risk of bias tool is needed.

For those children who no longer met diagnostic criteria at follow-up, we were unable to determine whether this was due to them being inaccurately diagnosed to begin with, or whether the child's autistic symptoms changed over time, so they no longer met the criteria. We tried to counter this by only including studies with same diagnostic processes and tools at baseline and follow-up. Studies that varied in their diagnostic tools, criteria or combination of both were excluded. We were not able to complete prognostic factor analyses on studies that used the same or different versions or editions of the diagnostic tool at baseline and follow-up, or autism spectrum disorder subgroups (i.e. autistic disorder versus pervasive developmental disorder - not otherwise specified (PDD-NOS), due to the small number of studies that had presented data for these areas.

Despite extensive searches of the literature, there were 21 records for which we were unable to determine eligibility for inclusion due to a lack of access to data from the study, which could not be obtained either through contacting the authors. We have listed these records in the [Studies awaiting classification](#) section of the review. While 21 studies is small relative to the number identified in the search ($n = 24,131$), there is a possibility we have missed some relevant data, particularly data from grey literature, which may impact our findings. However, the number of studies included in the meta-analysis is relatively large ($n = 34$) and studies were published a range of countries ($n = 13$) which minimises bias.

Applicability of findings to clinical practice and policy

The majority of the included studies were conducted in the USA and Canada (60%), with only 10% conducted in non-Western countries (one each for India, Japan, China and Taiwan). This limits the applicability of the prognostic data across a wide range of ethnicities and socioeconomic populations globally. It is unclear between countries, and sometimes within them, whether services

are seeing similar children. Information about whether a healthcare service is tertiary, secondary or primary would assist applicability, along with consistent reporting of intelligence, language level and adaptive behaviour of participants. Additionally, we note that the majority of study participants were boys and further explorations on diagnostic stability by gender will be important.

Although the results of this review need to be interpreted with caution, due to the low quality of evidence, they provide an overall estimate of persistence of diagnosis. This may be useful for policy and service developers, considering autism spectrum disorder is typically considered to be a persistent condition with continuing service needs. However, it is important to note that this is likely to be variable for each child.

Agreements and disagreements with other studies or reviews

We are aware of three previous systematic reviews that have been published and which investigated diagnostic stability in autism spectrum disorder (Bieleninik 2017; Rondeau 2010; Woolfenden 2012). Rondeau 2010 and Woolfenden 2012 found that children diagnosed with the subtype PDD-NOS were more likely to lose their diagnosis over time compared with children with a diagnosis of autistic disorder. Bieleninik 2017 found that some children diagnosed with autism shifted their diagnosis to autism spectrum disorder but the overall prevalence of autism spectrum disorder (which encompasses autism) remained constant. Given the changes to diagnostic criteria and labelling, our review was not able to assess whether the type of diagnosis (e.g. pervasive developmental disorder — not otherwise specified, or autistic disorder or autism) was associated with different proportions of children who continue to meet diagnostic criteria for autism spectrum disorder over time. Our review is more up to date and includes a larger volume of studies than the above reviews. We also used stricter inclusion criteria, so that included studies needed to use the same tools or criteria at baseline and follow-up, and we included a wide range of standardised and commonly utilised diagnostic tools and criteria.

Our systematic review explored autism spectrum disorder diagnosis in preschool children whereas other systematic reviews considered the diagnostic stability in older children and adolescents (Rondeau 2010; Woolfenden 2012). Furthermore, Rondeau 2010 and Woolfenden 2012 included a comparatively smaller samples ($n = 444$ and $n = 1363$, respectively) in their meta-analysis, suggesting a substantial number of studies eligible for inclusion in our review were not included in their analysis. The narrower focus of our review (i.e. only including preschool aged children at baseline and studies that had used the same diagnostic tools or criteria at baseline and follow-up), may have contributed to some differing findings from the prior reviews, namely that the proportion that were diagnosed at follow-up was slightly higher in our review.

Bieleninik 2017 found that specific interventions predicted change in total ADOS score. Intervention for autism spectrum disorder was not the focus of our review, and within the cohorts in the included studies, there was a marked heterogeneity in interventions being offered as treatment as usual. For intervention studies, we only included studies with a treatment-as-usual comparison group, and did not include data from the intervention arms. As such, we did not

seek to investigate whether the interventions made a difference to diagnostic stability.

AUTHORS' CONCLUSIONS

Autism spectrum disorder is an increasingly diagnosed neurodevelopmental disorder that can require significant support, depending on the individual, their characteristics and their environment. This systematic review brings together the existing evidence base of prognosis for autism spectrum disorder and provides an exemplar for implementation of systematic review methods in prognosis. Our review found low-quality evidence that nine out of 10 preschool children diagnosed with autism spectrum disorder may continue to have a diagnosis after one or more years of follow-up.

Future prospective studies, at low risk of bias, that are specifically designed to assess prognosis of autism spectrum disorder diagnoses are needed. These studies will ideally be embedded in clinical settings and so representative of current clinical practice rather than research based diagnostic processes. They will ideally provide information about the cohort for known factors of clinical importance including: intelligence quotient (IQ), language ability, gender, and co-occurring disorders. They will perform clinically comparable and replicable diagnostic processes at initial and follow-up time points using multidisciplinary teams. This will improve the quality of the evidence, allow rapid application to improve care and add certainty to the evidence base upon which parents, clinicians and policymakers can make decisions.

ACKNOWLEDGEMENTS

We would like to thank the editorial team of Cochrane Developmental, Psychosocial and Learning Problems (DPLP) for their advice and support during the preparation of this review, particularly Margaret Anderson who contributed to the development of the search strategy and Joanne Duffield who also provided support through the process. We also wish to thank the editors, external referees and statistician who commented on earlier drafts. We would like to thank Natalia Albein Urios for her contribution to the published protocol, Ruth Braden for her assistance with data extraction and risk of bias ratings and Andrew Hayen for advice on the statistical analyses. We would also like to acknowledge Charissa Ying Zhen Chan and Francesca Lami for the translation of some articles that were not written in English.

The CRG Editorial Team are grateful to the following reviewers for their time and comments: Dr Farid Foroutan, Ted Rogers Centre for Heart Research, Toronto, Ontario, Canada; and Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada; Dr Kristelle Hudry, Department of Psychology, Counselling and Therapy, School of Psychology and Public Health, La Trobe University, Australia; Nuala Livingstone, Cochrane Editorial and Methods Department, UK; and Zoe Thomas, Australia; as well as three reviewers who chose not to be publicly acknowledged. They are also grateful to Heather Maxwell for copyediting the review.

Cochrane DPLP supported the authors in the development of this prognosis review.

The following people conducted the editorial process for this article:

- Sign-off Editor (final editorial decision): Professor Geraldine Macdonald, Cochrane DPLP; University of Bristol, UK;
- Managing Editor (selected peer reviewers, collated peer-reviewer comments, provided editorial guidance to authors, edited the article): Dr Joanne Duffield, Cochrane DPLP; Queen's University Belfast, NI;
- Deputy Managing Editor (conducted editorial policy checks, provided editorial guidance to authors, edited the article): Dr Sarah Davies, Cochrane DPLP; University of Bristol, UK;
- Copy Editor (copy editing and production): Heather Maxwell;
- Peer-reviewers (provided comments): Dr Farid Foroutan, Ted Rogers Centre for Heart Research, Toronto, Ontario, Canada; and Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada (methods review)^a, Dr Kristelle Hudry, Department of Psychology, Counselling and Therapy, School of Psychology and Public Health, La Trobe University, Australia (clinical/content review), Dr Nuala Livingstone, Cochrane Editorial and

Methods Department, UK (methods review)^a; Zoe Thomas, Australia (consumer review), and Margaret Anderson, Cochrane DPLP; Queen's University Belfast, NI (search review). One additional peer reviewer provided clinical peer review, and two provided statistical peer review^{a,b}, but chose not to be publicly acknowledged.

^a*Dr Farid Foroutan is a member of Cochrane Metabolic and Endocrine Disorders Group and the Prognosis Methods Group, Dr Nuala Livingstone is a member of the Cochrane Editorial and Methods Department, as is one peer reviewer who chose to remain anonymous. All provided peer-review comments on this article and approved it for copyediting and publication, but otherwise were not involved in the editorial process or decision making for this article.

^bOne peer reviewer who chose to remain anonymous is a member of Cochrane DPLP, and provided peer-review comments on this article, but was not otherwise involved in the editorial process or decision making for this article.

REFERENCES

References to studies included in this review

Anderson 2009 {published data only} [10.1007/s10802-009-9326-0](#)

Anderson DK, Lord C, Risi S, DiLavore PS, Shulman C, Thurm A, et al. Patterns of growth in verbal abilities among children with autism spectrum disorder. *Journal of Consulting and Clinical Psychology* 2007;**75**(4):594-604. [DOI: [10.1037/0022-006X.75.4.594](#)] [PMID: 17663613]

* Anderson DK, Oti RS, Lord C, Welch K. Patterns of growth in adaptive social abilities among children with autism spectrum disorders. *Journal of Abnormal Child Psychology* 2009;**37**(7):1019-34. [DOI: [10.1007/s10802-009-9326-0](#)] [PMCID: PMC2893550] [PMID: 19521762]

Bedford R, Pickles A, Lord C. Early gross motor skills predict the subsequent development of language in children with autism spectrum disorder. *Autism Research* 2016;**9**(9):993-1001. [DOI: [10.1002/aur.1587](#)] [PMCID: PMC5031219] [PMID: 26692550]

Gotham K, Pickles A, Lord C. Trajectories of autism severity in children using standardized ADOS scores. *Pediatrics* 2012;**130**(5):e1278-84. [DOI: [10.1542/peds.2011-3668](#)] [PMCID: PMC3483889] [PMID: 23090336]

Gotham KO. Defining and quantifying severity of impairment in autism spectrum disorders across the lifespan. *Doctoral Dissertation, The University of Michigan* 2011;**71**(11-B):1-171.

Hus V, Taylor A, Lord C. Telescoping of caregiver report on the Autism Diagnostic Interview - Revised. *Journal of Child Psychology and Psychiatry* 2011;**52**(7):753-60. [DOI: [10.1111/j.1469-7610.2011.02398.x](#)] [PMCID: PMC3549439] [PMID: 21410473]

Lord C, Luyster R, Guthrie W, Pickles A. Patterns of developmental trajectories in toddlers with autism spectrum disorder. *Journal of Consulting and Clinical Psychology* 2012;**80**(3):477-89. [DOI: [10.1037/a0027214](#)] [PMID: 22506796] [PMID: PMC3365612]

Lord C, Shulman C, DiLavore P. Regression and word loss in autistic spectrum disorders. *Journal of Child Psychology and Psychiatry* 2004;**45**(5):936-55. [DOI: [10.1111/j.1469-7610.2004.t01-1-00287.x](#)] [PMID: 15225337]

Lord C. Follow-up of two-year-olds referred for possible autism. *Journal of Child Psychology and Psychiatry* 1995;**36**(8):1365-82. [DOI: [10.1111/j.1469-7610.1995.tb01669.x](#)] [PMID: 8988272]

Luyster R, Qui S, Lopez K, Lord C. Predicting outcomes of children referred for autism using the MacArthur-Bates Communicative Development Inventory. *Journal of Speech, Language, and Hearing Research* 2007;**50**(3):667-81. [DOI: [10.1044/1092-4388\(2007/047\)](#)] [PMID: 17538108]

Pickles A, Anderson DK, Lord C. Heterogeneity and plasticity in the development of language: a 17-year follow-up of children referred early for possible autism. *Journal of Child Psychology and Psychiatry* 2014;**55**(12):1354-62. [DOI: [10.1111/jcpp.12269](#)] [PMID: 24889883]

Richler J, Huerta M, Bishop SL, Lord C. Developmental trajectories of restricted and repetitive behaviors and interests in children with autism spectrum disorders. *Developmental Psychopathology* 2010;**22**(1):55-69. [DOI: [10.1017/S0954579409990265](#)] [PMCID: PMC2893549] [PMID: 20102647]

Thurm A, Lord C, Lee L-C, Newschaffer C. Predictors of language acquisition in preschool children with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2007;**37**(9):1721-34. [DOI: [10.1007/s10803-006-0300-1](#)] [PMID: 17180717]

Baghdadli 2012 {published data only}

* Baghdadli A, Assouline B, Sonié S, Pernon E, Darrou C, Michelon C, et al. Developmental trajectories of adaptive behaviors from early childhood to adolescence in a cohort of 152 children with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2012;**42**(7):1314-25. [PMID: 10.1007/s10803-011-1357-z] [PMID: 21928042]

Baghdadli A, Michelon C, Pernon E, Picot M C, Miot S, Sonie S, Rattaz C, Mottron L. Adaptive trajectories and early risk factors in the autism spectrum: A 15-year prospective study. *Autism Research* 2018;**11**(11):1455-1467.

Baghdadli A, Picot MC, Michelon C, Bodet J, Pernon E, Burstejzn C, et al. What happens to children with PDD when they grow up? Prospective follow-up of 219 children from preschool age to mid-childhood. *Acta Psychiatrica Scandinavica* 2007;**115**(5):403-12. [DOI: [10.1111/j.1600-0447.2006.00898.x](#)] [PMID: 17430419]

Baghdadli A, Picot MC, Pry R, Michelon C, Burzstejn C, Lazartigues A, et al. What factors are related to a negative outcome of self-injurious behaviour during childhood in pervasive developmental disorders? *Journal of Applied Research in Intellectual Disabilities* 2008;**21**(2):142-9. [DOI: [10.1111/j.1468-3148.2007.00389.x](#)]

Darrou C, Pry R, Pernon E, Michelon C, Aussilloux C, Baghdadli A. Outcome of young children with autism: does the amount of intervention influence developmental trajectories? *Autism* 2010;**14**(6):663-77. [DOI: [10.1177/1362361310374156](#)] [PMID: 21149421]

Pry R, Petersen AF, Baghdadli A. In search of prediction factors for autism spectrum disorders: an impossible task? *Psychology* 2012;**3**(11):997-1003. [DOI: [10.4236/psych.2012.311150](#)]

Pry R, Petersen AF, Baghdadli AM. On general and specific markers of lexical development in children with autism from 5 to 8 years of age. *Research in Autism Spectrum Disorders* 2011;**5**(3):1243-52. [DOI: [10.1016/j.rasd.2011.01.014](#)]

Benedetto 2021 {published data only}

Benedetto L, Cucinotta F, Maggio R, Germano E, De Raco R, Alquino A, et al. One-year follow-up diagnostic stability of autism spectrum disorder diagnosis in a clinical sample of children and toddlers. *Brain Sciences* 2021;**11**(1):1-15.

Bopp 2006 {published data only}

Bopp KD, Mirenda P, Zumbo BD. Behavior predictors of language development over 2 years in children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research* 2009;**52**(5):1106-20. [DOI: [10.1044/1092-4388\(2009/07-0262\)](https://doi.org/10.1044/1092-4388(2009/07-0262))] [PMID: 19797136]

* Bopp KD. Behaviour predictors of child development and parenting stress trajectories of children with autism [Doctoral dissertation]. Vancouver (BC): University of British Columbia, 2006.

Smith V, Mirenda P, Zaidman-Zait A. Predictors of expressive vocabulary growth in children with autism. *Journal of Speech, Language, and Hearing Research* 2007;**50**(1):149-60. [DOI: [10.1044/1092-4388\(2007/013\)](https://doi.org/10.1044/1092-4388(2007/013))] [PMID: 17344556]

Brian 2016 {published data only}

Brian J, Bryson SE, Smith IM, Roberts W, Roncadin C, Szatmari P, et al. Stability and change in autism spectrum disorder diagnosis from age 3 to middle childhood in a high-risk sibling cohort. *Autism* 2016;**20**(7):888-92. [DOI: [10.1177/1362361315614979](https://doi.org/10.1177/1362361315614979)] [PMID: 26685198]

Chu 2017 {published data only}

Chu C-L, Chiang C-H, Wu C-C, Hou Y-M, Liu J-H. Service system and cognitive outcomes for young children with autism spectrum disorders in a rural area of Taiwan. *Autism* 2017;**21**(5):581-91. [DOI: [10.1177/1362361316664867](https://doi.org/10.1177/1362361316664867)] [PMID: 28610539]

Demb 1989 {published data only}

Demb HB, Weintraub AG. A five-year follow-up of preschool children diagnosed as having an atypical pervasive developmental disorder. *Journal of Developmental & Behavioral Pediatrics* 1989;**10**(6):292-8. [PMID: 2600185]

DeWaay 2010 {published data only}

DeWaay RJ. Parents' perceptions of treatment effectiveness in a DIR/Floortime home intervention [Doctoral dissertation]. Pasadena (CA): Fuller Theological Seminary, 2010. [UMI NUMBER: 3485984]

Dietz 2007 {published data only}

Dietz C, Swinkels SH, Buitelaar JK, Van Daalen E, Van Engeland H. Stability and change of IQ scores in preschool children diagnosed with autistic spectrum disorder. *European Child & Adolescent Psychiatry* 2007;**16**(6):405-10. [DOI: [10.1007/s00787-007-0614-3](https://doi.org/10.1007/s00787-007-0614-3)] [PMID: 7401607]

Eaves 2004 {published data only}

Eaves LC, Ho HH. The very early identification of autism: outcome to age 4 1/2-5. *Journal of Autism and Developmental Disorders* 2004;**34**(4):367-78. [DOI: [10.1023/b:jadd.0000037414.33270.a8](https://doi.org/10.1023/b:jadd.0000037414.33270.a8)] [PMID: 15449513]

Elmose 2014 {published data only}

Elmose M, Trillingsgaard A, Jørgensen M, Nielsen A, Bruhn SS, Sørensen EU. Follow-up at mid-school age (9-13 years) of children assessed for autism spectrum disorder before the age

of four. *Nordic Journal of Psychiatry* 2014;**68**(5):362-8. [DOI: [10.3109/08039488.2013.846411](https://doi.org/10.3109/08039488.2013.846411)] [PMID: 24199947]

Flanagan 2010 {published data only}

Flanagan HE, Perry A, Freeman NL. Effectiveness of large-scale community-based intensive behavioral intervention: a waitlist comparison study exploring outcomes and predictors. *Research in Autism Spectrum Disorders* 2012;**6**(2):673-82. [DOI: [10.1016/j.rasd.2011.09.011](https://doi.org/10.1016/j.rasd.2011.09.011)]

* Flanagan HE. The Impact of Community-Based Intensive Behavioural Intervention [PhD thesis]. Toronto (ON): York University, 2010.

Freeman 2004 {published data only}

Freeman LJ. Functional Impairment in PDD-NOS: Predicting Outcome at a Two-Year Follow-up [PhD thesis]. Windsor (ON): University of Windsor, 2004.

Gabriels 2007 {published data only}

Gabriels RL, Ivers BJ, Hill DE, Agnew JA, McNeill J. Stability of adaptive behaviors in middle-school children with autism spectrum disorders. *Research in Autism Spectrum Disorders* 2007;**1**(4):291-303. [DOI: [10.1016/j.rasd.2006.11.004](https://doi.org/10.1016/j.rasd.2006.11.004)]

Gillberg 1990 {published data only}

Gillberg C, Ehlers S, Schaumann H, Jakobsson G, Dahlgren SO, Lindblom R, et al. Autism under age 3 years: a clinical study of 28 cases referred for autistic symptoms in infancy. *Journal of Child Psychology and Psychiatry* 1990;**31**(6):921-34. [DOI: [10.1111/j.1469-7610.1990.tb00834.x](https://doi.org/10.1111/j.1469-7610.1990.tb00834.x)] [PMID: 2246342]

Giserman-Kiss 2020 {published data only}

* Giserman-Kiss I, Carter AS. Stability of autism spectrum disorder in young children with diverse backgrounds. *Journal of Autism & Developmental Disorders* 2020;**50**(9):3263-75.

Giserman-Kiss I. Diagnostic stability of autism spectrum disorder in young children with diverse backgrounds [Doctoral Dissertation]. Boston: University of Massachusetts, 2018. [PROQUEST NUMBER: 10743610]

Gonzalez 1993 {published data only}

Gonzalez NM, Alpert M, Shay J, Campbell M, Small AM. Autistic children on follow up - change of diagnosis. *Psychopharmacology Bulletin* 1993;**29**(3):353-8. [PMID: 8121963]

Haglund 2020 {published data only}

Haglund N, Dahlgren S, Rastam M, Gustafsson P, Kallen K. Improvement of autism symptoms after comprehensive intensive early interventions in community settings. *Journal of the American Psychiatric Nurses Association* 2020;**00**(0):1-13. [DOI: [10.1177/1078390320915257](https://doi.org/10.1177/1078390320915257)]

Hinnebusch 2017 {published data only}

Hinnebusch AJ, Miller LE, Fein DA. Autism spectrum disorders and low mental age: diagnostic stability and developmental outcomes in early childhood. *Journal of Autism and Developmental Disorders* 2017;**47**(12):3967-82. [DOI: [10.1007/s10803-017-3278-y](https://doi.org/10.1007/s10803-017-3278-y)] [PMCID: PMC5845818] [PMID: 28861732]

Kim 2016 {published data only}

Kim SH, Macari S, Koller J, Chawarska K. Examining the phenotypic heterogeneity of early autism spectrum disorder: subtypes and short-term outcomes. *Journal of Child Psychology and Psychiatry* 2016;**57**(1):93-102. [DOI: [10.1111/jcpp.12448](https://doi.org/10.1111/jcpp.12448)] [PMID: 26264996] [PMID: PMC6852790]

Klintwall 2015 {published data only}

Klintwall L, Macari S, Eikeseth S, Chawarska K. Interest level in 2-year-olds with autism spectrum disorder predicts rate of verbal, nonverbal, and adaptive skill acquisition. *Autism* 2015;**19**(8):925-33. [DOI: [10.1177/1362361314555376](https://doi.org/10.1177/1362361314555376)] [PMCID: PMC4878117] [PMID: 25398893]

Lombardo 2015 {published data only}

Lombardo MV, Pierce K, Eyer LT, Carter Barnes C, Ahrens-Barbeau C, Solso S, et al. Different functional neural substrates for good and poor language outcome in autism. *Neuron* 2015;**86**(2):567-77. [DOI: [10.1016/j.neuron.2015.03.023](https://doi.org/10.1016/j.neuron.2015.03.023)] [PMID: 25864635] [PMID: PMC4610713]

Malhi 2011 {published data only}

Malhi P, Singhi P. Follow up of children with autism spectrum disorders: stability and change in diagnosis. *Indian Journal of Pediatrics* 2011;**78**(8):941-5. [DOI: [10.1007/s12098-011-0370-8](https://doi.org/10.1007/s12098-011-0370-8)] [PMID: 21318394]

Martin-Borreguero 2021 {published data only}

Martin-Borreguero P, Gomez-Fernandez AR, De La Torre-Aguilar MJ, Gil-Campos M, Flores-Rojas K, Perez-Navero JL. Children with autism spectrum disorder and neurodevelopmental regression present a severe pattern after a follow-up at 24 months. *Frontiers in Psychiatry* 2021;**12**:644324.

Moore 2003 {published data only}

Moore V, Goodson S. 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* 2003;**7**(1):47-63. [DOI: [10.1177/1362361303007001005](https://doi.org/10.1177/1362361303007001005)] [PMID: 12638764]

Moss 2008 {published data only}

Magiati I, Charman T, Howlin P. A two-year prospective follow-up study of community-based early intensive behavioural intervention and specialist nursery provision for children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry* 2007;**48**(8):803-12. [DOI: [10.1111/j.1469-7610.2007.01756.x](https://doi.org/10.1111/j.1469-7610.2007.01756.x)] [PMID: 17683452]

Magiati I, Moss J, Charman T, Howlin P. Patterns of change in children with autism spectrum disorders who received community based comprehensive interventions in their pre-school years: a seven year follow-up study. *Research in Autism Spectrum Disorders* 2011;**5**(3):1016-27. [DOI: [10.1016/j.rasd.2010.11.007](https://doi.org/10.1016/j.rasd.2010.11.007)]

Magiati I, Moss J, Yates R, Charman T, Howlin P. Is the Autism Treatment Evaluation Checklist a useful tool for monitoring progress in children with autism spectrum disorders? *Journal of Intellectual Disability Research* 2011;**55**(3):302-12. [DOI: [10.1111/j.1365-2788.2010.01359.x](https://doi.org/10.1111/j.1365-2788.2010.01359.x)] [PMID: 21199043]

* Moss J, Magiati I, Charman T, Howlin P. 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* 2008;**38**(6):1081-91. [DOI: [10.1007/s10803-007-0487-9](https://doi.org/10.1007/s10803-007-0487-9)] [PMID: 18058215]

Naigles 2016 {published data only}

Naigles LR, Cheng M, Rattansone NX, Tek S, Khetrupal N, Fein D, et al. "You're telling me!" The prevalence and predictors of pronoun reversals in children with autism spectrum disorders and typical development. *Research in Autism Spectrum Disorders* 2016;**27**:11-20. [PMCID: PMC4834724] [PMID: 10.1016/j.rasd.2016.03.008] [PMID: 27103941]

Neuhaus 2016 {published data only}

Neuhaus E, Jones EJ, Barnes K, Sterling L, Estes A, Munson J, et al. The relationship between early neural responses to emotional faces at age 3 and later autism and anxiety symptoms in adolescents with autism. *Journal of Autism and Developmental Disorders* 2016;**46**(7):2450-63. [DOI: [10.1007/s10803-016-2780-y](https://doi.org/10.1007/s10803-016-2780-y)] [PMCID: PMC5305034] [PMID: 27055415]

Ozonoff 2015 {published data only}

Ozonoff S, Young GS, Landa RJ, Brian J, Bryson S, Charman T, et al. Diagnostic stability in young children at risk for autism spectrum disorder: a baby siblings research consortium study. *Journal of Child Psychology and Psychiatry* 2015;**56**(9):988-98. [DOI: [10.1111/jcpp.12421](https://doi.org/10.1111/jcpp.12421)] [PMCID: PMC4532646] [PMID: 25921776]

Paul 2008 {published data only}

Paul R, Chawarska K, Cicchetti D, Volkmar F. Language outcomes of toddlers with autism spectrum disorders: a two year follow-up. *Autism Research* 2008;**1**(2):97-107. [DOI: [10.1002/aur.12](https://doi.org/10.1002/aur.12)] [PMCID: PMC2946084] [PMID: 19360656]

Qian 2018 {published data only}

Ke X, Qian L. Alteration of hub organization in the white matter structural network in toddlers with autism spectrum disorder: a two-year follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry* 2017;**56**(10):S260.

Li Y, Zhou Z, Chang C, Qian L, Li C, Xiao T, et al. Anomalies in uncinate fasciculus development and social defects in preschoolers with autism spectrum disorder. *BMC Psychiatry* 2019;**19**(1):399. [DOI: [10.1016/j.neurenf.2012.04.383](https://doi.org/10.1016/j.neurenf.2012.04.383)]

* Qian L, Wang Y, Chu K, Li Y, Xiao C, Xiao T, et al. Alterations in hub organization in the white matter structural network in toddlers with autism spectrum disorder: A 2-year follow-up study. *Autism Research* 2018;**11**(9):1218-28.

Rivard 2019 {published data only}

Mello C, Rivard M, Terroux A, Mercier C. Differential responses to early behavioural intervention in young children with autism spectrum disorders as a function of features of intellectual disability. *Journal on Developmental Disabilities* 2018;**23**(3):5-17.

* Rivard M, Morin M, Mello C, Terroux A, Mercier C. Follow-Up of children with autism spectrum disorder 1 year after early behavioral intervention. *Behavior Modification* 2019;**43**(4):490-517.

Robain 2020 {published data only}

Robain F, Franchini M, Kojovic N, Wood de Wilde Hi, Schaer M. Predictors of treatment outcome in preschoolers with autism spectrum disorder: an observational study in the greater Geneva area, Switzerland. *Journal of Autism & Developmental Disorders* 2020;**50**(11):3815-30.

Santocchi 2012 {published data only}

Santocchi E, Prosperi M, Tancredi R, Muratori F. Diagnostic stability of autism in preschooler age. *Conference: 20th World Congress of the International Association for Child and Adolescent Psychiatry and Allied Professions* 2012;**60**(5):S198. [DOI: [10.1016/j.neurenf.2012.04.383](https://doi.org/10.1016/j.neurenf.2012.04.383)]

Sheinkopf 1998 {published data only}

Sheinkopf SJ, Siegel B. Home-based behavioral treatment of young children with autism. *Journal of Autism and Developmental Disorders* 1998;**28**(1):15-23. [DOI: [10.1023/a:1026054701472](https://doi.org/10.1023/a:1026054701472)] [PMID: 9546298]

Smith 2019 {published data only}

Smith D P, Hayward D W, Gale C M, Eikeseth S, Klintwall L. Treatment gains from early and intensive behavioral intervention (EIBI) are maintained 10 years later. *Behavior Modification* 2019;**45**(4):581-601. [DOI: [.00r.g1/107.171/0771/405145445451591988822895](https://doi.org/10.1177/107.171/0771/405145445451591988822895)]

Soke 2011 {published data only}

Soke GN, Philofsky A, Diguseppi C, Lezotte D, Rogers S, Hepburn S. 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* 2011;**15**(5):545-62. [DOI: [10.1177/1362361309358332](https://doi.org/10.1177/1362361309358332)] [PMCID: PMC4426200] [PMID: 21586639]

Solomon 2014 {published data only}

Mahoney G, Solomon R. Mechanism of developmental change in the PLAY Project Home Consultation program: evidence from a randomized control trial. *Journal of Autism and Developmental Disorders* 2016;**46**(5):1860-71. [DOI: [10.1007/s10803-016-2720-x](https://doi.org/10.1007/s10803-016-2720-x)] [PMID: 26830414]

* Solomon R, Van Egeren LA, Mahoney G, Quon Huber MS, Zimmerman P. PLAY Project Home Consultation intervention program for young children with autism spectrum disorders: a randomized controlled trial. *Journal of Developmental & Behavioral Pediatrics* 2014;**35**(8):475-85. [DOI: [10.1097/DBP.000000000000096](https://doi.org/10.1097/DBP.000000000000096)] [PMCID: PMC4181375] [PMID: 25264862]

Solomon 2016 {published data only}

* Solomon M, Iosif A-M, Nordahl C, Libero L, Li D, Ghetti S, et al. W33. IQ-based developmental phenotypes of autism spectrum disorder in early childhood and their correlates. *Neuropsychopharmacology* 2016;**41**:S476. [DOI: [10.1038/npp.2016.242](https://doi.org/10.1038/npp.2016.242)] [URL: www.nature.com/articles/npp2016242.pdf]

Solomon M, Iosif AM, Reinhardt VP, Libero LE, Nordahl CW, Ozonoff S, et al. What will my child's future hold? phenotypes of

intellectual development in 2-8-year-olds with autism spectrum disorder. *Autism Research* 2018;**11**(1):121-32.

Waizbard-Bartov E, Ferrer E, Young GS, Heath B, Rogers S, Wu Nordahl C, et al. Trajectories of autism symptom severity change during early childhood. *Journal of Autism & Developmental Disorders* 2021;**51**(1):227-42.

Spjut Jansson 2016 {published data only}

Spjut Jansson B, Miniscalco C, Westerlund J, Kantzer A-K, Fernell E, Gillberg C. Children who screen positive for autism at 2.5 years and receive early intervention: a prospective naturalistic 2-year outcome study. *Neuropsychiatric Disease and Treatment* 2016;**12**:2255-63. [DOI: [10.2147/NDT.S108899](https://doi.org/10.2147/NDT.S108899)] [PMCID: PMC5012621] [PMID: 27621636]

Sullivan 2010 {published data only}

Sullivan A. Developmental trajectories of young children with autism enrolled in an intensive behaviour intervention program: what the ablls can tell us about their progress [Doctoral dissertation]. Toronto (ON): York University, 2010. [IBSN: 978-0-494-64911-4]

Szatmari 2021 {published data only}

Baribeau DA, Vigod S, Pullenayegum E, Kerns CM, Mirenda P, Smith IM, et al. Co-occurring trajectories of anxiety and insistence on sameness behaviour in autism spectrum disorder. *British Journal of Psychiatry* 2021;**218**(1):20-7.

Baribeau DA, Vigod S, Pullenayegum E, Kerns CM, Mirenda P, Smith IM, et al. Repetitive behavior severity as an early indicator of risk for elevated anxiety symptoms in autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry* 2020;**59**(7):890-9.

Bennett TA, Szatmari P, Georgiades K, Hanna S, Janus M, Georgiades S, et al. Do reciprocal associations exist between social and language pathways in preschoolers with autism spectrum disorders? *Journal of Child Psychology and Psychiatry* 2015;**56**(8):874-83. [DOI: [10.1111/jcpp.12356](https://doi.org/10.1111/jcpp.12356)] [PMID: 25376440]

Bennett TA, Szatmari P, Georgiades K, Hanna S, Janus M, Georgiades S, et al. Language impairment and early social competence in preschoolers with autism spectrum disorders: a comparison of DSM-5 profiles. *Journal of Autism and Developmental Disorders* 2014;**44**(11):2797-808. [DOI: [10.1007/s10803-014-2138-2](https://doi.org/10.1007/s10803-014-2138-2)] [PMID: 24865586]

Courchesne, V, Bedford R, Pickles A, Duku E, Kerns C, Mirenda P, et al, Pathways Team. Non-verbal IQ and change in restricted and repetitive behavior throughout childhood in autism: a longitudinal study using the Autism Diagnostic Interview-Revised. *Molecular Autism* 2021;**12**(1):10. [DOI: <https://doi.org/10.1186/s13229-021-00461-7>]

Georgiades S, Boyle M, Szatmari P, Hanna S, Duku E, Zwaigenbaum L, et al. Modeling the phenotypic architecture of autism symptoms from time of diagnosis to age 6. *Journal of Autism and Developmental Disorders* 2014;**44**(12):3045-55. [DOI: [10.1007/s10803-014-2167-x](https://doi.org/10.1007/s10803-014-2167-x)] [PMID: 24958435]

Georgiades S, Tait PA, McNicholas PD, Duku E, Zwaigenbaum L, Smith IM, et al. Trajectories of symptom severity in children

with autism: variability and turning points through the transition to school. *Journal of Autism and Developmental Disorders*. 2021; **Epub ahead of print**:no pagination. [DOI: [0.1007/s10803-021-04949-2](https://doi.org/10.1007/s10803-021-04949-2)]

* Szatmari P, Cost KT, Duku E, Bennett T, Elsabbagh M, Georgiades S, et al. Association of child and family attributes with outcomes in children with autism. *JAMA Network Open* 2021; **4**(3):e212530.

Szatmari P, Georgiades S, Duku E, Bennett TA, Bryson S, Fombonne E, et al. Developmental trajectories of symptom severity and adaptive functioning in an inception cohort of preschool children with autism spectrum disorder. *JAMA Psychiatry* 2015; **72**(3):276-83. [DOI: [10.1001/jamapsychiatry.2014.2463](https://doi.org/10.1001/jamapsychiatry.2014.2463)] [PMID: 25629657]

Takeda 2007 {published data only}

Takeda T, Koyama T, Kurita H. Comparison of developmental/intellectual changes between autistic disorder and pervasive developmental disorder not otherwise specified in preschool years. *Psychiatry and Clinical Neurosciences* 2007; **61**(6):684-6. [DOI: [10.1111/j.1440-1819.2007.01740.x](https://doi.org/10.1111/j.1440-1819.2007.01740.x)] [PMID: 18081633]

Thomas 2009 {published data only}

Thomas CJ. Analyses of five years of results from project data (developmentally appropriate treatment for autism), a high-quality early childhood special education program for students with autism spectrum disorders [Doctoral dissertation]. Washington (DC): University of Washington, 2009. [UMI NUMBER: 3393990]

Venker 2014 {published data only}

Davidson Mm, Ellis Weismer S. A discrepancy in comprehension and production in early language development in ASD: is it clinically relevant? *Journal of Autism AND Developmental Disorders* 2017; **47**(7):2163-75. [DOI: [10.1007/s10803-017-3135-z](https://doi.org/10.1007/s10803-017-3135-z)] [PMCID: PMC5812677] [PMID: 28447305]

Ellis Weismer S, Kover ST. Preschool language variation, growth, and predictors in children on the autism spectrum. *Journal of Child Psychology and Psychiatry* 2015; **56**(12):1327-37. [DOI: [10.1111/jcpp.12406](https://doi.org/10.1111/jcpp.12406)] [PMCID: PMC4565784] [PMID: 25753577]

Ray-Subramanian CE, Ellis Weismer S. Receptive and expressive language as predictors of restricted and repetitive behaviors in young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2012; **42**(10):2113-20. [DOI: [10.1007/s10803-012-1463-6](https://doi.org/10.1007/s10803-012-1463-6)] [PMCID: PMC3422597] [PMID: 22350337]

Venker CE, Kover ST, Ellis Weismer S. Brief report: fast mapping predicts differences in concurrent and later language abilities among children with ASD. *Journal of Autism and Developmental Disorders* 2016; **46**(3):1118-23. [DOI: [10.1007/s10803-015-2644-x](https://doi.org/10.1007/s10803-015-2644-x)] [PMCID: PMC4747812] [PMID: 26572655]

* Venker CE, Ray-Subramanian CE, Bolt DM, Ellis Weismer S. Trajectories of autism severity in early childhood. *Journal of Autism and Developmental Disorders* 2014; **44**(3):546-63. [DOI: [10.1007/s10803-013-1903-y](https://doi.org/10.1007/s10803-013-1903-y)] [PMCID: PMC3909724] [PMID: 23907710]

Wu 2016 {published data only}

Wu Y-T, Maenner MJ, Wiggins LD, Rice CE, Bradley CC, Lopez ML, et al. Retention of autism spectrum disorder diagnosis: the role of co-occurring conditions in males and females. *Research in Autism Spectrum Disorders* 2016; **25**:76-86. [DOI: [10.1016/j.rasd.2016.02.001](https://doi.org/10.1016/j.rasd.2016.02.001)] [PMCID: PMC5603237] [PMID: 28936232]

Zappella 1990 {published data only}

Zappella M. Young autistic children treated with ethologically oriented family therapy. *Family Systems Medicine* 1990; **8**(1):14-27. [DOI: [10.1037/h0089267](https://doi.org/10.1037/h0089267)]

Zappella 2010 {published data only}

Zappella M. Autistic regression with and without EEG abnormalities followed by favourable outcome. *Brain & Development* 2010; **32**(9):739-45. [DOI: [10.1016/j.braindev.2010.05.004](https://doi.org/10.1016/j.braindev.2010.05.004)] [PMID: 20708360]

Zwaigenbaum 2015 {published data only}

Zwaigenbaum L, Bryson SE, Brian J, Smith IM, Roberts W, Szatmari P, et al. Stability of diagnostic assessment for autism spectrum disorder between 18 and 36 months in a high-risk cohort. *Autism Research* 2016; **9**(7):790-800. [DOI: [10.1002/aur.1585](https://doi.org/10.1002/aur.1585)] [PMID: 26613202]

References to studies excluded from this review

Bacon 2018 {published data only}

Bacon EC, Courchesne E, Barnes CC, Cha D, Pence S, Schreibman L, et al. Rethinking the idea of late autism spectrum disorder onset. *Development and Psychopathology* 2018; **30**(2):553-69. [DOI: [10.1017/S0954579417001067](https://doi.org/10.1017/S0954579417001067)] [PMID: 28803559]

Bacon EC, Moore A, Lee Q, Carter BC, Courchesne E, Pierce K. Identifying prognostic markers in autism spectrum disorder using eye tracking. *Autism: The International Journal of Research & Practice* 2020; **24**(3):658-69.

Pierce K, Gazestani VH, Bacon E, Barnes CC, Cha D, Nalabolu S, et al. Evaluation of the diagnostic stability of the early autism spectrum disorder phenotype in the general population starting at 12 months. *JAMA Pediatrics* 2019; **173**(6):578-87.

Bal 2019 {published data only}

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

Berry 2009 {published data only}

* Berry LN. Early treatments associated with optimal outcome in children with autism spectrum disorders [Doctoral dissertation]. Connecticut, New England: University of Connecticut, 2009. [UMI NUMBER: 3377035]

Boorstein HC. Regressive and early onset autism spectrum disorders: a comparison of developmental trajectories, autistic behaviors, and medical histories [Doctoral dissertation]. Connecticut, New England: University of Connecticut, 2010. [UMI NUMBER: 3475517]

Kleinman JM, Ventola PE, Pandey J, Verbalis AD, Barton M, Hodgson S, et al. Diagnostic stability in very young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2008;**38**(4):606-15. [DOI: [10.1007/s10803-007-0427-8](https://doi.org/10.1007/s10803-007-0427-8)] [PMCID: PMC3625643] [PMID: 17924183]

Miller LE, Burke JD, Troyb E, Knoch K, Herlihy LE, Fein DA. Preschool predictors of school-age academic achievement in autism spectrum disorder. *Clinical Neuropsychologist* 2017;**31**(2):382-403. [DOI: [10.1080/13854046.2016.1225665](https://doi.org/10.1080/13854046.2016.1225665)] [PMCID: PMC5464727] [PMID: 27705180]

Moulton E, Barton M, Robins DL, Abrams DN, Fein D. Early characteristics of children with ASD who demonstrate optimal progress between age two and four. *Journal of Autism and Developmental Disorders* 2016;**46**(6):2160-73. [DOI: [10.1007/s10803-016-2745-1](https://doi.org/10.1007/s10803-016-2745-1)] [PMCID: PMC4860351] [PMID: 26895327]

Verbalis AD. Longitudinal changes in the expression of autism spectrum disorders in girls and boys [Doctoral dissertation]. Connecticut, New England: University of Connecticut, 2010. [UMI NUMBER: 3451418]

Canal-Bedia 2016 {published data only}

Canal-Bedia R, Magan-Maganto M, Bejarano-Martin A, De Pablos-De la Morena A, Bueno-Carrera G, Manso-De Dios S, et al. Early detection and stability of diagnosis in autism spectrum disorders. *Revista de Neurología* 2016;**62**(Suppl 1):S15-20. [PMID: 26922953]

Charman 2005 {published data only}

* Charman T, Taylor E, Drew A, Cockerill H, Brown J-A, Baird G. 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* 2005;**46**(5):500-13. [DOI: [10.1111/j.1469-7610.2004.00377.x](https://doi.org/10.1111/j.1469-7610.2004.00377.x)] [PMID: 15845130]

Cox A, Klein K, Charman, T, Baird G, Baron-Cohen S, Swettenham J, et al. Autism spectrum disorders at 20 and 42 months of age: Stability of clinical and ADI-R diagnosis. *Journal of Child Psychology and Psychiatry* 1999;**40**(5):719-32. [DOI: <https://doi.org/10.1111/1469-7610.00488>]

Drew A, Baird G, Baron-Cohen S, Cox A, Slonims V, Wheelwright S, et al. A pilot randomised control trial of a parent training intervention for pre-school children with autism: Preliminary findings and methodological challenges. *European Child and Adolescent Psychiatry* 2002;**11**:266-72. [DOI: [10.1007/s00787-002-0299-6](https://doi.org/10.1007/s00787-002-0299-6)]

Chawarska 2007 {published data only}

Chawarska K, Klin A, Paul R, Macari S, Volkmar F. A prospective study of toddlers with ASD: short-term diagnostic and cognitive outcomes. *Journal of Child Psychology and Psychiatry* 2009;**50**(10):1235-45. [DOI: [10.1111/j.1469-7610.2009.02101.x](https://doi.org/10.1111/j.1469-7610.2009.02101.x)] [PMCID: PMC4878113] [PMID: 19594835]

* Chawarska K, Klin A, Paul R, Volkmar F. Autism spectrum disorder in the second year: stability and change in syndrome expression. *Journal of Child Psychology and Psychiatry* 2007;**48**(2):128-38. [DOI: [10.1111/j.1469-7610.2006.01685.x](https://doi.org/10.1111/j.1469-7610.2006.01685.x)] [PMID: 17300551]

Chawarska K, Paul R, Klin A, Hannigen S, Dichtel LE, Volkmar F. Parental recognition of developmental problems in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2007;**37**(1):62-72. [DOI: [10.1007/s10803-006-0330-8](https://doi.org/10.1007/s10803-006-0330-8)] [PMID: 17195921]

Clark 2017 {published data only}

Barbaro J, Dissanayake C. Diagnostic stability of autism spectrum disorder in toddlers prospectively identified in a community-based setting: behavioural characteristics and predictors of change over time. *Autism* 2017;**21**(7):830-40. [DOI: [10.1177/1362361316654084](https://doi.org/10.1177/1362361316654084)] [PMID: 27474118]

* Clark ML, Barbaro J, Dissanayake C. Continuity and change in cognition and autism severity from toddlerhood to school age. *Journal of Autism and Developmental Disorders* 2017;**47**(2):328-39. [DOI: [10.1007/s10803-016-2954-7](https://doi.org/10.1007/s10803-016-2954-7)] [PMID: 27848051]

Clark ML, Vinen Z, Barbaro J, Dissanayake C. School age outcomes of children diagnosed Early and later with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 2018;**48**(1):92-102. [DOI: [10.1007/s10803-017-3279-x](https://doi.org/10.1007/s10803-017-3279-x)] [PMID: 28905160]

De Giacomo 2009 {published data only}

De Giacomo A, Lamanna AL, Laforteza E, Lecce P, Martinelli D, Margari L. Diagnostic stability and early developmental trajectory of pervasive developmental disorder using the Autism Diagnostic Interview - Revised. *Italian Journal of Psychopathology* 2009;**15**(4):336-42.

Guthrie 2013 {published data only}

Guthrie W, Swineford LB, Nottke C, Wetherby AM. Early diagnosis of autism spectrum disorder: stability and change in clinical diagnosis and symptom presentation. *Journal of Child Psychology and Psychiatry* 2013;**54**(5):582-90. [DOI: [10.1111/jcpp.12008](https://doi.org/10.1111/jcpp.12008)] [PMCID: PMC3556369] [PMID: 23078094]

Hedvall 2014 {published data only}

Barnevik Olsson M, Lundström S, Westerlund J, Giacobini MB, Gillberg C, Fernell E. Preschool to school in autism: neuropsychiatric problems 8 years after diagnosis at 3 years of age. *Journal of Autism and Developmental Disorders* 2016;**46**(8):2749-55. [DOI: [10.1007/s10803-016-2819-0](https://doi.org/10.1007/s10803-016-2819-0)] [PMID: 27230761]

Fernell E, Hedvall A, Westerlund J, Hoglund Carlsson L, Eriksson M, Barnevik Olsson M, et al. Early intervention in 208 Swedish preschoolers with autism spectrum disorder. A prospective naturalistic study. *Research in Developmental Disabilities* 2011;**32**(6):2092-101. [DOI: [10.1016/j.ridd.2011.08.002](https://doi.org/10.1016/j.ridd.2011.08.002)] [PMID: 21985993]

* Hedvall A, Westerlund J, Fernell E, Holm A, Gillberg C, Billstedt E. Autism and developmental profiles in preschoolers: stability and change over time. *Acta Paediatrica* 2014;**103**(2):174-81. [DOI: [10.1111/apa.12455](https://doi.org/10.1111/apa.12455)] [PMID: 24237479]

Hedvall A, Westerlund J, Fernell E, Norrelgen F, Kjellmer L, Olsson MB, et al. Preschoolers with autism spectrum disorder followed for 2 years: those who gained and those who lost the most in terms of adaptive functioning outcome. *Journal of*

Autism and Developmental Disorders 2015;**45**(11):3624-33. [DOI: [10.1007/s10803-015-2509-3](https://doi.org/10.1007/s10803-015-2509-3)] [PMID: 26123008]

Norrelgen F, Fernell E, Eriksson M, Hedvall A, Persson C, Sjölin M, et al. Children with autism spectrum disorders who do not develop phrase speech in the preschool years. *Autism* 2015;**19**(8):934-43. [DOI: [10.1177/1362361314556782](https://doi.org/10.1177/1362361314556782)] [PMID: 25488002]

Jónsdóttir 2007 {published data only}

Jónsdóttir SL, Saemundsen E, Asmundsdóttir G, Hjartardóttir S, Asgeirsdóttir BB, Smáradóttir HH, et al. Follow-up of children diagnosed with pervasive developmental disorders: stability and change during the preschool years. *Journal of Autism and Developmental Disorders* 2007;**37**(7):1361-74. [DOI: [10.1007/s10803-006-0282-z](https://doi.org/10.1007/s10803-006-0282-z)] [PMID: 17146706]

Kadam 2021 {published data only}

Kadam A, Patni B, Pandit A, Patole S. Stability of the initial diagnosis of autism spectrum disorder by DSM-5 in children: a short-term follow-up study. *Journal of Tropical Pediatrics* 2020;**00**(0):1-6. [DOI: [10.1093/tropej/fmaa104](https://doi.org/10.1093/tropej/fmaa104)]

Kantzer 2018 {published data only}

* Kantzer A-K, Fernell E, Westerlund J, Hagberg B, Gillberg C, Miniscalco C. Young children who screen positive for autism: stability, change and "comorbidity" over two years. *Research in Developmental Disabilities* 2018;**72**:297-307. [DOI: [10.1016/j.ridd.2016.10.004](https://doi.org/10.1016/j.ridd.2016.10.004)] [PMID: 27818061]

Thompson L, Gillberg C, Landberg S, Kantzer AK, Miniscalco C, Barnevik OM, et al. Autism with and without regression: a two-year prospective longitudinal study in two population-derived Swedish cohorts. *Journal of Autism & Developmental Disorders* 2019;**49**(6):2281-290.

Ozonoff 2018 {published data only}

Ozonoff S, Young GS, Brian J, Charman T, Shephard E, Solish A, et al. Diagnosis of autism spectrum disorder after age 5 in children evaluated longitudinally since infancy. *Journal of the American Academy of Child & Adolescent Psychiatry* 2018;**57**(11):849-57.

Shumway 2012 {published data only}

Shumway S, Farmer C, Thurm A, Joseph L, Black D, Golden C. The ADOS calibrated severity score: relationship to phenotypic variables and stability over time. *Autism Research* 2012;**5**(4):267-76. [DOI: [10.1002/aur.1238](https://doi.org/10.1002/aur.1238)] [PMCID: PMC3422401] [PMID: 22628087]

Stone 1999 {published data only}

* Stone WL, Lee EB, Ashford L, Brissie J, Hepburn SL, Coonrod EE, et al. Can autism be diagnosed accurately in children under 3 years? *Journal of Child Psychology and Psychiatry* 1999;**40**(2):219-26. [PMID: 10188704]

Turner LM, Stone WL, Pozdol SL, Coonrod EE. Follow-up of children with autism spectrum disorders from age 2 to age 9. *Autism* 2006;**10**(3):243-65. [DOI: [10.1177/1362361306063296](https://doi.org/10.1177/1362361306063296)] [PMID: 16682397]

Turner LM, Stone WL. Variability in outcome for children with an ASD diagnosis at age 2. *Journal of Child Psychology and Psychiatry* 2007;**48**(8):793-802. [DOI: [10.1111/j.1469-7610.2007.01744.x](https://doi.org/10.1111/j.1469-7610.2007.01744.x)] [PMID: 17683451]

Sutera 2010 {published data only}

Sutera S, Pandey J, Esser EL, Rosenthal MA, Wilson LB, Barton M, et al. Predictors of optimal outcome in toddlers diagnosed with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2007;**37**(1):98-107. [DOI: [10.1007/s10803-006-0340-6](https://doi.org/10.1007/s10803-006-0340-6)]

* Sutera S. Predictors of optimal outcome in children with an autism spectrum disorder. *Dissertation Abstracts International: Section B: The Sciences and Engineering* 2010;**70**(9-B):5849.

Thurm 2015 {published data only}

Farmer C, Swineford L, Swedo SE, Thurm A. Classifying and characterizing the development of adaptive behavior in a naturalistic longitudinal study of young children with autism. *Journal of Neurodevelopmental Disorders* 2018;**10**(1):No pagination reported. [DOI: [10.1186/s11689-017-9222-9](https://doi.org/10.1186/s11689-017-9222-9)]

Thurm A, Manwaring SS, Swineford L, Farmer C. Longitudinal study of symptom severity and language in minimally verbal children with autism. *Journal of Child Psychology and Psychiatry* 2015;**56**(1):97-104. [DOI: [10.1111/jcpp.12285](https://doi.org/10.1111/jcpp.12285)] [PMCID: PMC4581593] [PMID: 24961159]

Tunc 2021 {published data only}

Tunc B, Pandey J, St John T, Meera SS, Maldarelli J E, Zwaigenbaum L, et al. Diagnostic shifts in autism spectrum disorder can be linked to the fuzzy nature of the diagnostic boundary: a data-driven approach. *Journal of Child Psychology and Psychiatry* 2021;**62**(10):1236-45. [DOI: [10.1111/jcpp.13406](https://doi.org/10.1111/jcpp.13406)]

Van Daalen 2009 {published data only}

Van Daalen E, Kemner C, Dietz C, Swinkels SH, Buitelaar JK, Van Engeland H. Inter-rater reliability and stability of diagnoses of autism spectrum disorder in children identified through screening at a very young age. *European Child & Adolescent Psychiatry* 2009;**18**(11):663-74. [DOI: [10.1007/s00787-009-0025-8](https://doi.org/10.1007/s00787-009-0025-8)] [PMCID: PMC2762529] [PMID: 19421728]

Venter 1992 {published data only}

Venter A, Lord C, Schopler E. A follow-up study of high functioning autistic children. *Journal of Child Psychology and Psychiatry* 1992;**33**(3):489-507. [DOI: [10.1111/j.1469-7610.1992.tb00887.x](https://doi.org/10.1111/j.1469-7610.1992.tb00887.x)] [PMID: 1577895]

References to studies awaiting assessment

Anglim 2012 {published data only}

Anglim M, Ackermann P, Barry M, Kashif M, Moran A, O'Connell A, et al. Stability and change in a clinical sample of preschool children with autistic spectrum disorder. *Neuropsychiatrie de l'Enfance et de l'Adolescence. Conference Abstract: 20th World Congress of the International Association for Child and Adolescent Psychiatry and Allied Professions, IACAPAP 2012. Paris France. 2012*;**60**(5 Supplement):S215-S216. [DOI: [10.1016/j.neurenf.2012.04.472](https://doi.org/10.1016/j.neurenf.2012.04.472)]

Boi 2017 {published data only}

Boi J, Donno F, Petza S, Cera F, Balia C, Carucci S, Zuddas A. Medium-term efficacy data of medications in children and adolescents with autism spectrum disorder: An 18 months retrospective follow up study. *European Neuropsychopharmacology* 2017;**27** (Supplement 4):S1112.

Brown 1997 {published data only}

Brown H M, N Amer Riding Handicapped Assoc. Post-therapy follow up of the effects on autism of equine-based therapy. Denver: North American Riding Handicapped Association Inc, 1997.

Chang 2017 {published data only}

Chang Tzu-Ling, Chen Chia-ling, Chung Chia-Ying. Poster 135: Developmental Outcomes in Children with Autism Spectrum Disorder of Different Cognitive Functions. *Physical Medicine and Rehabilitation* 2017;**9**:S176-S177. [DOI: [10.1016/j.pmrj.2017.08.078](https://doi.org/10.1016/j.pmrj.2017.08.078)]

Chang 2017a {published data only}

Chang C, Qiu NN, Xiao T, Xiao X, Chu KK, Li Y, et al. [Structural change of the corpus callosum fibers in toddlers with autism spectrum disorder: two-year follow-up]. *Zhonghua Er Ke Za Zhi* 2017;**55**(12):920-5.

Da Silva 2003 {published data only}

Da Silva PC, Eira C, Pombo J, Silva A, Da Silva LC, Martins F, et al. Clinical program for treatment of difficulties with relating and communicating, based on the D.I.R. Model. *Analise Psicológica* 2003;**21**(1):3139.

Eapen 2019 {published data only}

Eapen V, Mathew N, Mazzoni A. Subtyping autism: can we predict treatment response in Autism Spectrum Disorder? *IBRO Reports* 2019;**6** (Supplement):S358.

Faroghizadeh 2021 {published data only}

Faroghizadeh K, Ziaian T. Effectiveness of applied behavioral analysis method on autism symptoms. *International Journal of Pharmaceutical Research* 2021;**13**(1):5710-6.

Gabis 2011 {published data only}

Gabis V L, Maayan M, Rivka S, Aya SH, Marcy Y. Preschool diagnostic process and changes in diagnosis of autism spectrum disorder. *Annals of Neurology. Conference: 40th Annual Child Neurology Society Meeting Scientific Program. Savannah, GA United States* 2011;**70**(15):S132.

Ghamari Kivi 2012 {published data only}

Ghamari Kivi H, Agh A, Nasoudi R. Efficacy of applied behavioral analysis in reducing symptoms of stereotyped behavior, interaction and communicational problems in autistic children. *Iranian Journal of Psychiatry. Supplement Abstracts of the 5th International Congress of Iranian Association of Child and Adolescents Psychiatry*. 2012;**7**(4):106-7.

Jimenez-Martinez 2018 {published data only}

Jimenez-Martinez M, Nunez-Rodriguez A, Guzman G. Analysis of applied behavior treatment for children with autism spectrum disorder. *European Psychiatry* 2018;**48** (Supplement 1):S473.

Melville 1987 {published data only}

Melville LC. Douglass Developmental Disabilities Center: A Follow-Up Study [PhD thesis]. New Brunswick (NJ): Rutgers University, 1987.

Millikovsky-Ayalon 2012 {published data only}

Millikovsky-Ayalon M, Sofrin R, Raz R, Shilon-Hadass A, Yehuda M, Mukamel M, et al. Preschool diagnostic process and changes in diagnosis of autism spectrum disorder. *Harefuah* 2012;**151**(3):150-4, 190. [PMID: 22519262]

Mohanta 2019 {published data only}

Mohanta A, Mittal V K, Ieee. Acoustic features for characterizing speech of children affected with ASD. *2019 IEEE 16th India Council International Conference* 2019;**00**(0):1-4. [DOI: [10.1109/INDICON47234.2019.9029043](https://doi.org/10.1109/INDICON47234.2019.9029043)]

Mosconi 2009 {published data only}

* Mosconi MW, Cody-Hazlett H, Poe MD, Gerig G, Gimpel-Smith R, Piven J. Longitudinal study of amygdala volume and joint attention in 2- to 4-year-old children with autism. *Archives of General Psychiatry* 2009;**66**(5):509-16. [DOI: [10.1001/archgenpsychiatry.2009.19](https://doi.org/10.1001/archgenpsychiatry.2009.19)] [PMCID: PMC3156446] [PMID: 19414710]

Mosconi MW, Reznick JS, Mesibov G, Piven J. The Social Orienting Continuum and Response Scale (SOC-RS): a dimensional measure for preschool-aged children. *Journal of Autism & Developmental Disorders* 2009;**39**(2):24-50. [DOI: [10.1007/s10803-008-0620-4](https://doi.org/10.1007/s10803-008-0620-4)]

Muratori 2002 {published data only}

Muratori F, Dini P, Cosenza A, Parrini B, Fascetti L, Vanni F. A controlled study on therapeutic effects of day-hospital for autistic children [Studio controllato sugli effetti terapeutici del day-hospital nell'autismo infantile]. *Imago* 2002;**9**(2):133-42. [URL: Available at www.scopus.com/record/display.uri?eid=2-s2.0-0036322970&origin=inward&txGid=089debeeb957531e3d1767500217f]

Ozyurt 2020 {published data only}

Ozyurt G, Elikucuk C D. Augmentative and alternative communication for children with autism spectrum disorder: a randomised study of awareness and developmental language interventions. *Hong Kong Journal of Paediatrics* 2020;**25**(2):79-88.

Perucchini 2005 {published data only}

Perucchini P, Muratori F, Parrini B. Theory of mind, gesture and autism. *Giornale Italiano di Psicologia* 2005;**32**(4):799-817.

Selvakumar 2018 {published data only}

Selvakumar L, Malhi P, Singhi P. Stability and change in Diagnosis of Autism Spectrum Disorder over time among toddlers. *International Journal of Medical Research & Health Sciences* 2018;**7**(3):4045.

Takesada 1992 {published data only}

Takesada M, Naruse H, Nagahata M, Kazamatsuri H, Nakane Y, Yamazaki K, et al. An Open Clinical-Study of Apropterin Hydrochloride (R-Tetrahydrobiopterin, R-Thbp) in Infantile-

Autism - Clinical Effects and Long-Term Follow-Up. 965 edition. Amsterdam: Elsevier Science Publ B V, 1992.

Zhang 2019 {published data only}

Zhang L, Liu Y, Zhou Z, Wei Y, Wang J, Yang Ji, et al. A follow-up study on the long-term effects of rehabilitation in children with autism spectrum disorders. *NeuroRehabilitation* 2019;**44**(1):1-7.

Zirakashvili 2018 {published data only}

Zirakashvili M, Gabunia M, Tatishvili N, Lomidze G, Janelidze M. Predictors of better outcome in children with autism spectrum disorders: a pilot study in Georgia. *Developmental Medicine and Child Neurology* 2018;**60** (Supplement 2):23-4.

Additional references

APA 1980

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Third Edition (DSM-III). Washington (DC): American Psychiatric Association, 1980.

APA 1994

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). Washington (DC): American Psychiatric Association, 1994.

APA 2000

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR). Washington (DC): American Psychiatric Association, 2000.

APA 2013

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th edition. (DSM-5). Arlington (VA): American Psychiatric Publishing, 2013.

Barbaro 2016

Barbaro J, Dissanayake C. Diagnostic stability of autism spectrum disorder in toddlers prospectively identified in a community-based setting: behavioural characteristics and predictors of change over time. *Autism* 2016 Jul 28 [Epub ahead of print]. [DOI: [10.1177/1362361316654084](https://doi.org/10.1177/1362361316654084)] [PMID: 27474118]

Beuscher 2014

Buescher AV, Cidav Z, Knapp M, Mandell DS. Costs of autism spectrum disorder in the United Kingdom and the United States. *JAMA Pediatrics* 2014;**168**(8):721-8. [DOI: [10.1001/jamapediatrics.2014.210](https://doi.org/10.1001/jamapediatrics.2014.210)] [PMID: 24911948]

Bieleninik 2017

Bieleninik L, Posserud M-B, Geretsegger M, Thompson G, Elefant C, Gold C. 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* 2017;**12**(9):e0183160. [DOI: [10.1371/journal.pone.0183160](https://doi.org/10.1371/journal.pone.0183160)] [PMCID: PMC5608197] [PMID: 28934215]

Borenstein 2009

Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to Meta-Analysis. West Sussex (UK): John Wiley & Sons, Ltd, 2009.

CDC 2014

Centers for Disease Control and Prevention. Prevalence of autism spectrum disorder among children aged 8 years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2010. www.cdc.gov/mmwr/preview/mmwrhtml/ss6302a1.htm (accessed 3 October 2016).

Corsello 2013

Corsello CM, Akshoomoff N, Stahmer AC. Diagnosis of autism spectrum disorders in 2-year-olds: a study of community practice. *Journal of Child Psychology and Psychiatry* 2013;**54**(2):178-85. [DOI: [10.1111/j.1469-7610.2012.02607.x](https://doi.org/10.1111/j.1469-7610.2012.02607.x)] [PMC3505251] [PMID: 22905987]

Daniels 2011

Daniels AM, Rosenberg RE, Law JK, Lord C, Kaufmann WE, Law PA. Stability of initial autism spectrum disorder diagnoses in community settings. *Journal of Autism and Developmental Disorders* 2011;**41**(1):110-21. [DOI: [10.1007/s10803-010-1031-x](https://doi.org/10.1007/s10803-010-1031-x)] [PMID: 20473589]

Deeks 2022

Deeks JJ, Higgins JP, Altman DG. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor(s). *Cochrane Handbook for Systematic Reviews of Interventions* Version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook..

Elsabbagh 2012

Elsabbagh M, Divan G, Koh YJ, Kim YS, Kauchali S, Marcin C, et al. Global prevalence of autism and other pervasive developmental disorders. *Autism Research* 2012;**5**(3):160-79. [DOI: [10.1002/aur.239](https://doi.org/10.1002/aur.239)] [PMC3763210] [PMID: 22495912]

Fombonne 2009

Fombonne E. Epidemiology of pervasive developmental disorders. *Pediatric Research* 2009;**65**(6):591-8. [DOI: [10.1203/PDR.0b013e31819e7203](https://doi.org/10.1203/PDR.0b013e31819e7203)] [PMID: 19218885]

Ganz 2007

Ganz ML. The lifetime distribution of the incremental societal costs of autism. *Archives of Paediatrics & Adolescent Medicine* 2007;**161**(4):343-9. [DOI: [10.1001/archpedi.161.4.343](https://doi.org/10.1001/archpedi.161.4.343)] [PMID: 17404130]

Gilliam 1995

Gilliam JE. Gilliam Autism Rating Scale. Austin (TX): Pro-Ed, 1995.

Goin-Kochel 2007

Goin-Kochel RP, Myers BJ, Mackintosh VH. Parental reports on the use of treatments and therapies for children with autism spectrum disorders. *Research in Autism Spectrum Disorders* 2007;**1**(3):195-209. [DOI: <http://dx.doi.org/10.1016/j.rasd.2006.08.006>]

Green 2006

Green VA, Pituch KA, Itchon J, Choi A, O'Reilly M, Sigafoos J. Internet survey of treatments used by parents of children with autism. *Research in Developmental Disabilities* 2006;**27**(1):70-84. [DOI: [10.1016/j.ridd.2004.12.002](https://doi.org/10.1016/j.ridd.2004.12.002)] [PMID: 15919178]

Guyatt 2011

Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *Journal of Clinical Epidemiology* 2011;**64**(4):380-2. [DOI: [10.1016/j.jclinepi.2010.09.011](https://doi.org/10.1016/j.jclinepi.2010.09.011)] [PMID: 21185693]

Hansen 2015

Hansen SN, Schendel DE, Parner ET. Explaining the increase in the prevalence of autism spectrum disorders: the proportion attributable to changes in reporting practices. *JAMA Pediatrics* 2015;**169**(1):56-62. [DOI: [10.1001/jamapediatrics.2014.1893](https://doi.org/10.1001/jamapediatrics.2014.1893)] [PMID: 25365033]

Hayden 2006

Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Annals of Internal Medicine* 2006;**144**(6):427-37. [PMID: 16549855]

Hayden 2013

Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Annals of Internal Medicine* 2013;**158**(4):280-6. [DOI: [10.7326/0003-4819-158-4-201302190-00009](https://doi.org/10.7326/0003-4819-158-4-201302190-00009)] [PMID: 23420236]

Hayden 2019

Hayden JA, Wilson MN, Riley RD, Iles R, Pincus T, Ogilvie R. Individual recovery expectations and prognosis of outcomes in non-specific low back pain: prognostic factor review. *Cochrane Database of Systematic Reviews* 2019, Issue I1. Art. No: CD011284. [DOI: [10.1002/14651858.CD011284.pub2](https://doi.org/10.1002/14651858.CD011284.pub2)]

Horlin 2014

Horlin C, Falkmer M, Parsons R, Albrecht MA, Falkmer T. The cost of autism spectrum disorders. *PLoS One* 2014;**9**(9):e106552. [DOI: [10.1371/journal.pone.0106552](https://doi.org/10.1371/journal.pone.0106552)]

Howlin 2004

Howlin P, Goode S, Hutton J, Rutter M. Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry* 2004;**45**(2):212-29. [PMID: 14982237]

Howlin 2012

Howlin P, Moss P. Adults with autism spectrum disorders. *Canadian Journal of Psychiatry* 2012;**57**(5):275-83.

Hunter 2014

Hunter JP, Saratzis A, Sutton AJ, Boucher RH, Sayers RD, Bown MJ. In meta-analyses of proportion studies, funnel plots were found to be an inaccurate method of assessing publication bias. *Journal of Clinical Epidemiology* 2014;**67**(8):897-903. [DOI: [doi:10.1016/j.jclinepi.2014.03.003](https://doi.org/10.1016/j.jclinepi.2014.03.003)]

Iorio 2015

Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ* 2015;**350**:h870. [DOI: [10.1136/bmj.h870](https://doi.org/10.1136/bmj.h870)]

Kenny 2016

Kenny L, Hattersley C, Molins B, Buckley C, Povey C, Pellicano E. Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism* 2016;**20**(4):442-62. [DOI: [10.1177/1362361315588200](https://doi.org/10.1177/1362361315588200)] [PMID: 26134030]

Kim 2011

Kim YS, Leventhal BL, Koh YJ, Fombonne E, Laska E, Lim EC, et al. Prevalence of autism spectrum disorders in a total population sample. *The American Journal of Psychiatry* 2011;**168**(9):904-12. [DOI: [10.1176/appi.ajp.2011.10101532](https://doi.org/10.1176/appi.ajp.2011.10101532)] [PMID: 21558103]

King 2009

King M, Bearman P. Diagnostic change and the increased prevalence of autism. *International Journal of Epidemiology* 2009;**38**(5):1224-34. [DOI: [10.1093/ije/dyp261](https://doi.org/10.1093/ije/dyp261)] [PMC2800781]

Landa 2013

Landa Rj, Gross AL, Stuart EA, Faherty A. Developmental trajectories in children with and without autism spectrum disorders: the first 3 years. *Child Development* 2013;**84**(2):429-42. [DOI: [10.1111/j.1467-8624.2012.01870.x](https://doi.org/10.1111/j.1467-8624.2012.01870.x)] [PMC4105265] [PMID: 23110514]

Le Couteur 2003

Le Couteur A, Rutter M, Lord C. Autism Diagnostic Interview - Revised. Los Angeles (CA): Western Psychological Services, 2003.

Lord 2000

Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, et al. The Autism Diagnostic Observation Schedule - Generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders* 2000;**30**(3):205-23. [PMID: 11055457]

Lord 2012

Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, Bishop S. Autism Diagnostic Observation Schedule. 2nd edition. Torrance (CA): Western Psychological Services, 2012.

Lundstrom 2015

Lundström S, Reichenberg A, Anckarsäter H, Lichtenstein P, Gillberg C. Autism phenotype versus registered diagnosis in Swedish children: prevalence trends over 10 years in general population samples. *BMJ* 2015;**350**:h1961. [DOI: <http://dx.doi.org/10.1136/bmj.h1961>]

Magiati 2014

Magiati I, Tay XW, Howlin P. Cognitive, language, social and behavioural outcomes in adults with autism spectrum disorders: a systematic review of longitudinal follow-up studies

in adulthood. *Clinical Psychology Review* 2014;**34**(1):73-86. [10.1016/j.cpr.2013.11.002] [PMID: 24424351]

Moher 2009

Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;**339**:b2535. [DOI: [10.1136/bmj.b2535](https://doi.org/10.1136/bmj.b2535)] [PMCID: PMC2714657] [PMID: 19622551]

NCCHS 2011

National Center for Health Statistics. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). www.cdc.gov/nchs/icd/icd9cm.htm (accessed 7 July 2017).

NICE 2011

National Institute for Health and Care Excellence. Autism spectrum disorder in under 19s: recognition, referral and diagnosis (CG128). www.nice.org.uk/guidance/cg128/resources/autism-in-under-19s-recognition-referral-and-diagnosis-35109456621253 (accessed 20 May 2015).

Randall 2018

Randall M, Egberts KJ, Samtani A, Scholten RJ, Hooft L, Livingstone N, et al. Diagnostic tests for autism spectrum disorder (ASD) in preschool children. *Cochrane Database of Systematic Reviews* 2018, Issue 7. Art. No: CD009044. [DOI: [10.1002/14651858.CD009044.pub2](https://doi.org/10.1002/14651858.CD009044.pub2)] [PMCID: PMC6513463] [PMID: 30075057]

RevMan Web 2020 [Computer program]

The Cochrane Collaboration Review Manager Web (RevMan Web). Version 1.22.0. The Cochrane Collaboration, 2020. Available at revman.cochrane.org.

Rondeau 2010

Rondeau E, Klein LS, Masse A, Bodeau N, Cohen D, Guilé JM. Is pervasive developmental disorder not otherwise specified less stable than autistic disorder? A meta-analysis. *Journal of Autism and Developmental Disorders* 2011;**41**(9):1267-76. [10.1007/s10803-010-1155-z] [PMID: 21153874]

Schopler 1980

Schopler E, Reichler RJ, DeVellis RF, Daly K. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of Autism and Developmental Disorders* 1980;**10**(1):91-103. [PMID: 6927682]

Schünemann 2013

Schünemann H, Brožek J, Guyatt G, Oxman A, editor(s). GRADE Handbook (updated October 2013). Available from gdt.guidelinedevelopment.org/app/handbook/handbook.html (accessed 24 July 2017).

Sicherman 2021

Sicherman N, Charite J, Eyal G, Janecka M, Loewenstein G, Law K, et al. Clinical signs associated with earlier diagnosis of children with autism Spectrum disorder. *BMC Pediatrics* 2021;**21**(1):96. [DOI: doi.org/10.1186/s12887-021-02551-0]

Šimkovic 2019

Šimkovic M, Träuble B. Robustness of statistical methods when measure is affected by ceiling and/or floor effect. *PLOS One* 2019;**14**(8):e0220889. [DOI: [10.1371/journal.pone.0220889](https://doi.org/10.1371/journal.pone.0220889)] [PMCID: PMC6699673] [PMID: 31425561]

Simonoff 2008

Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G. 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 and Adolescent Psychiatry* 2008;**47**(8):921-9. [DOI: [10.1097/CHI.0b013e318179964f](https://doi.org/10.1097/CHI.0b013e318179964f)] [PMID: 18645422]

Skuse 2004

Skuse D, Warrington R, Bishop D, Chowdhury U, Lau J, Mandy W, et al. The developmental, dimensional and diagnostic interview (3di): a novel computerized assessment for autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry* 2004;**43**(5):548-58. [DOI: [10.1097/00004583-200405000-00008](https://doi.org/10.1097/00004583-200405000-00008)] [PMID: 15100561]

StataCorp 2019 [Computer program]

Stata Statistical Software. StataCorp, Version Version 16. College Station (TX): StataCorp, 2019. Available at www.stata.com.

Takeda 2005

Takeda T, Koyama T, Kanai C, Kurita H. Clinical variables at age 2 predictive of mental retardation at age 5 in children with pervasive developmental disorder. *Psychiatry and Clinical Neurosciences* 2005;**59**(6):717-22. [DOI: [10.1111/j.1440-1819.2005.01442.x](https://doi.org/10.1111/j.1440-1819.2005.01442.x)] [PMID: 16401249]

Turner 2007

Turner LM, Stone WL. Variability in outcome for children with an ASD diagnosis at age 2. *Journal of Child Psychology and Psychiatry* 2007;**48**(8):793-802. [DOI: [10.1111/j.1469-7610.2007.01744.x](https://doi.org/10.1111/j.1469-7610.2007.01744.x)] [PMID: 17683451]

Van 't Hof 2021

Van 't Hof M, Tisseur C, Van Berckeleer-Onnes I, Van Nieuwenhuyzen A, Daniels AM, Deen M, et al. Age at autism spectrum disorder diagnosis: a systematic review and meta-analysis from 2012 to 2019. *Autism* 2021;**25**(4):862-73. [DOI: [10.1177/1362361320971107](https://doi.org/10.1177/1362361320971107)] [PMID: 33213190]

Volkmar 2014

Volkmar F, Siegel M, Woodbury-Smith M, King B, McCracken J, State M, et al. Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* 2014;**53**(2):237-57. [DOI: [10.1016/j.jaac.2013.10.013](https://doi.org/10.1016/j.jaac.2013.10.013)] [PMID: 24472258]

WAADF 2012

Western Australian Autism Diagnosticians' Forum. Autism Spectrum Disorders: Assessment Process in Western Australia. www.waadf.org.au/WAADF_ASD_Assessment_Process_2012.pdf (accessed 20 May 2015).

WHO 1979

World Health Organization. International Classification of Diseases, Ninth Revision. Geneva (Switzerland): World Health Organization, 1979.

WHO 1992

World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders. Clinical Descriptions and Diagnostic Guidelines. Geneva (Switzerland): World Health Organization, 1992.

Wing 2002a

Wing L, Potter D. The epidemiology of autism spectrum disorders: is prevalence rising? *Mental Retardation and Developmental Disabilities Research Reviews* 2002;**8**(3):151-61. [DOI: [10.1002/mrdd.10029](https://doi.org/10.1002/mrdd.10029)] [PMID: 12216059]

Wing 2002b

Wing L, Leekam SR, Libby SJ, Gould J, Locombe M. The Diagnostic Interview for Social and Communication Disorders:

background, inter-rater reliability and clinical use. *Journal of Child Psychology and Psychiatry* 2002;**43**(3):307-25. [PMID: 11944874]

Woolfenden 2012

Woolfenden S, Sarkozy V, Ridley G, Williams K. A systematic review of the diagnostic stability of autism spectrum disorder. *Research in Autism Spectrum Disorders* 2012;**6**(1):345-54. [DOI: <http://dx.doi.org/10.1016/j.rasd.2011.06.008>]

References to other published versions of this review
Brignell 2017

Brignell A, Albein-Urios N, Woolfenden S, Hayen A, Iorio A, Williams K. Overall prognosis of preschool autism spectrum disorder diagnoses. *Cochrane Database of Systematic Reviews* 2017, Issue 8. Art. No: CD012749. [DOI: [10.1002/14651858.CD012749](https://doi.org/10.1002/14651858.CD012749)]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES
Characteristics of included studies [ordered by study ID]

Anderson 2009
Study characteristics

Methods	Design: prospective Setting: cohort study Sample: clinical, drawn from broad community base Location (country): USA Length of follow-up (years): 11 (seen at 2, 3, 5, 9 & 13 years) Diagnostic tool (multidisciplinary or not): DSM-IV (N)
Population	Sample size (% male): 192 (84) Diagnosis type: autism, PDD-NOS and non-spectrum developmental delay Mean age at baseline (years): 2.4 (0.43) IQ (mean standard score): < 70 Language (mean standard score): - Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: not reported Funding: not reported Note(s): sample included children without a diagnosis of ASD and we were unable to extract data on the children with ASD separately as we were unable to obtain this information from study authors

Baghdadli 2012
Study characteristics

Methods	Design: retrospective Setting: cohort study Sample: clinical Location (country): France Length of follow-up (years): 3.0
---------	---

Overall prognosis of preschool autism spectrum disorder diagnoses (Review)

Copyright © 2022 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Baghdadli 2012 *(Continued)*
Diagnostic tool (multidisciplinary or not): ICD-10 & CARS (Y)

Population	Sample size (% male): 152 (82) Diagnosis type: ASD Mean age at baseline (years): 4.9 IQ (mean standard score): < 70 Language (mean standard score): < 70 Adaptive behaviour (mean standard score): < 70
Notes	Conflict of interest: not reported Funding: Programme Hospitalier de Recherche Clinique (PHRC) and Orange Foundation, France Note(s): none

Benedetto 2021
Study characteristics

Methods	Design: prospective Setting: cohort study Sample: clinical with broad community base Location (country): Italy Length of follow-up (years): 1 Diagnostic tool (multidisciplinary or not): DSM-5 and ADOS (Y)
Population	Sample size (% male): 147 (80) Diagnosis type: ASD Mean age at baseline (years): 2.3 IQ (mean standard score): - Language (mean standard score): - Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: the authors declared they have no COI. Funding: no funding Note(s):

Bopp 2006
Study characteristics

Methods	Design: prospective Setting: cohort study Sample: clinical Location (country): Canada Length of follow-up (years): 2.0 Diagnostic tool (multidisciplinary or not): CARS (U)
Population	Sample size (% male): 70 (83) Diagnosis type: ASD Mean age at baseline (years): 4.2 IQ (mean standard score): < 70 Language (mean standard score): - Adaptive behaviour (mean standard score): < 70
Notes	Conflict of interest: not reported Funding: not reported

Overall prognosis of preschool autism spectrum disorder diagnoses (Review)

Bopp 2006 (Continued)

Note(s): none

Brian 2016

Study characteristics

Methods	<p>Design: prospective Setting: cohort study Sample: clinical Location (country): Canada Length of follow-up (years): 6.4 Diagnostic tool (multidisciplinary or not): ADOS and DSM-IV (N)</p>
Population	<p>Sample size (% male): 18 (72) Diagnosis type: ASD Mean age at baseline (years): 3.2 IQ (mean standard score): > 70 Language (mean standard score): > 70 Adaptive behaviour (mean standard score): -</p>
Notes	<p>Conflict of interest: not reported Funding: Canadian Institutes of Health Research; Autism Speaks; Autism Speaks Canada; NeuroDe- vNet; and the Simons Foundation Note(s): none</p>

Chu 2017

Study characteristics

Methods	<p>Design: prospective Setting: cohort study Sample: clinical Location (country): Taiwan Length of follow-up (years): 1.5 Diagnostic tool (multidisciplinary or not): DSM-IV-TR and ADOS Social and Communication subset (N)</p>
Population	<p>Sample size (% male): 35 (89) Diagnosis type: ASD Mean age at baseline (years): 2.5 IQ (mean standard score): < 70 Language (mean standard score): - Adaptive behaviour (mean standard score): -</p>
Notes	<p>Conflict of interest: not reported Funding: National Science Council in Taiwan (NSC-96-2413-H-004-021-MY3 and MOST 102-2410- H-004 -044 -MY3) Note(s): none</p>

Demb 1989

Study characteristics

Overall prognosis of preschool autism spectrum disorder diagnoses (Review)

Demb 1989 (Continued)

Methods	Design: retrospective, with whole cohort considered Setting: cohort study Sample: clinical, from broad community base Location (country): USA Length of follow-up (years): 5.0 Diagnostic tool (multidisciplinary or not): DSM-III and DSM-III-R (N)
Population	Sample size (% male): 12 (75) Diagnosis type: ASD Mean age at baseline (years): 4.5 IQ (mean standard score): < 70 Language (mean standard score): - Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: not reported Funding: not reported Note(s): none

DeWaay 2010
Study characteristics

Methods	Design: prospective Setting: intervention trial, with one treatment arm Sample: clinical Location (country): USA Length of follow-up (years): 1.5 Diagnostic tool (multidisciplinary or not): GARS/GARS2 (U)
Population	Sample size (% male): 13 (77) Diagnosis type: ASD Mean age at baseline (years): 4.3 IQ (mean standard score): - Language (mean standard score): - Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: not reported Funding: not reported Note(s): none

Dietz 2007
Study characteristics

Methods	Design: prospective Setting: cohort study Sample: population based Location (country): the Netherlands Length of follow-up (years): 1.5 Diagnostic tool (multidisciplinary or not): DSM-IV (N)
Population	Sample size (% male): 14 (66). Note, data were only provided for the full sample (n = 39), which includes children without ASD. Diagnosis type: ASD

Dietz 2007 (Continued)

Mean age at baseline (years): 2.1
IQ (mean standard score): < 70
Language (mean standard score): < 70
Adaptive behaviour (mean standard score): -

Notes **Conflict of interest:** not reported
Funding: quote: "This study was supported by grants 940-38-045 and 940-38-014 (Chronic Disease Program), by grant 28.3000-2 of the Praeventiefonds-ZONMW, by the Netherlands Organisation for Scientific Research (NWO), by a grant from the Dutch Ministry of Health, Welfare and Culture, and by grants from Cure Autism Now, and the Korczak Foundation."
Note(s): children were assessed at two time points for ASD using the ADI-R, but only the stability results from IQ tests at two time points are presented. Unable to obtain required data from the study authors

Eaves 2004

Study characteristics

Methods **Design:** prospective
Setting: cohort study
Sample: population based
Location (country): Canada
Length of follow-up (years): 2.3
Diagnostic tool (multidisciplinary or not): CARS, DSM-IV, MDT (Y)

Population **Sample size (% male):** 43 (80)
Diagnosis type: ASD
Mean age at baseline (years): 2.8
IQ (mean standard score): < 70
Language (mean standard score): -
Adaptive behaviour (mean standard score): < 70

Notes **Conflict of interest:** not reported
Funding: Vancouver Foundation, British Columbia Medical Services Association
Note(s): none

Elmose 2014

Study characteristics

Methods **Design:** prospective
Setting: cohort study
Sample: unclear how sample drawn
Location (country): Denmark
Length of follow-up (years): 8.3
Diagnostic tool (multidisciplinary or not): ADOS & ICD-10 (Y)

Population **Sample size (% male):** 23 (78)
Diagnosis type: ASD
Mean age at baseline (years): 3.1
IQ (mean standard score): -
Language (mean standard score): -
Adaptive behaviour (mean standard score): -

Notes **Conflict of interest:** the authors declared they have no COI.

Elmose 2014 (Continued)

Funding: not reported
Note(s): none

Flanagan 2010

Study characteristics

Methods	<p>Design: prospective Setting: control group (wait list of a community intervention) Sample: clinical, from a broad community base Location (country): Canada Length of follow-up (years): 1.4 Diagnostic tool (multidisciplinary or not): CARS (N)</p>
Population	<p>Sample size (% male): 67 (82) Diagnosis type: ASD Mean age at baseline (years): 3.6 IQ (mean standard score): - Language (mean standard score): - Adaptive behaviour (mean standard score): < 70</p>
Notes	<p>Conflict of interest: not reported Funding: Canadian Institutes of Health Research (Canada Graduate Scholarship) and the Canadian Institutes of Health Research/National Alliance for Autism Research (Interdisciplinary Training Program in Autism Spectrum Disorders) Note(s): none</p>

Freeman 2004

Study characteristics

Methods	<p>Design: retrospective Setting: cohort study Sample: clinical, from a broad community base Location (country): Canada Length of follow-up (years): 2.2 Diagnostic tool (multidisciplinary or not): DSM IV (N)</p>
Population	<p>Sample size (% male): 59 (81) Diagnosis type: ASD Mean age at baseline (years): 4 IQ (mean standard score): < 70 Language (mean standard score): - Adaptive behaviour (mean standard score): -</p>
Notes	<p>Conflict of interest: not reported Funding: not reported (thesis) Note(s): none</p>

Gabriels 2007

Study characteristics

Overall prognosis of preschool autism spectrum disorder diagnoses (Review)

Gabriels 2007 *(Continued)*

Methods	Design: prospective Setting: cohort study Sample: clinical Location (country): USA Length of follow-up (years): 5.3 Diagnostic tool (multidisciplinary or not): DSM-IV and ADOS (N)
Population	Sample size (% male): 17 (71) Diagnosis type: ASD Mean age at baseline (years): 5.7 IQ (mean standard score): > 70 Language (mean standard score): - Adaptive behaviour (mean standard score): > 70
Notes	Conflict of interest: not reported Funding: not reported Note(s): none

Gillberg 1990
Study characteristics

Methods	Design: prospective Setting: cohort study Sample: clinical, from a broad community base Location (country): Sweden Length of follow-up (years): 4.0 Diagnostic tool (multidisciplinary or not): DSM-III-R (N)
Population	Sample size (% male): 25 (68) Diagnosis type: ASD Mean age at baseline (years): 1.1 IQ (mean standard score): < 70 Language (mean standard score): - Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: not reported Funding: Child Neuropsychiatry Centre, Goteborg, Sweden Note(s): none

Giserman-Kiss 2020
Study characteristics

Methods	Design: prospective Setting: cohort study where all received intervention Sample: clinical Location (country): USA Length of follow-up (years): 1.98 Diagnostic tool (multidisciplinary or not): ADOS (N)
Population	Sample size (% male): 60 (87) Diagnosis type: ASD Mean age at baseline (years): 2.31

Overall prognosis of preschool autism spectrum disorder diagnoses (Review)

Giserman-Kiss 2020 (Continued)

IQ (mean standard score): < 70
Language (mean standard score): < 70
Adaptive behaviour (mean standard score): -

Notes **Conflict of interest:** the authors declared they have no COI.
Funding: HRSA (R40MC26195), NIMH (R01MH104400), Autism Speaks, the UMB Office of Graduate Studies, and the UMB Graduate Student Assembly.
Note(s): none

Gonzalez 1993
Study characteristics

Methods **Design:** prospective
Setting: cohort study
Sample: clinical
Location (country): USA
Length of follow-up (years): 1.0
Diagnostic tool (multidisciplinary or not): DSM-III, DSM-III-R, DSM-IV & ICD-10 (N)

Population **Sample size (% male):** 30 (73)
Diagnosis type: ASD
Mean age at baseline (years): 4.5
IQ (mean standard score): < 70
Language (mean standard score): -
Adaptive behaviour (mean standard score): -

Notes **Conflict of interest:** not reported
Funding: supported, in part, by USPHS Grants MH-18915 (Drs Gonzalez, Shay, and Campbell), MH-32212 (Dr Campbell), and P01 MH-47200 (DSM-IV Autism/Pervasive Developmental Disorder Field Trial - American Psychiatric Association) from the National Institute of Mental Health; the Stallone Fund for Autism Research; the Hirschell E and Deanna E Levine Foundation; and the Marion O and Maximilian E Hoffman Foundation, Inc
Note(s): none

Haglund 2020
Study characteristics

Methods **Design:** prospective
Setting: intensive intervention in a community setting
Sample: clinical from broad community base
Location (country): Sweden
Length of follow-up (years): 3.2
Diagnostic tool (multidisciplinary or not): ADOS (N)

Population **Sample size (% male):** 27 (81.5)
Diagnosis type: ASD
Mean age at baseline (years): 4.9
IQ (mean standard score): both
Language (mean standard score): -
Adaptive behaviour (mean standard score): -

Notes **Conflict of interest:** the author(s) declared no potential COI with respect to the research, authorship, and/or publication of the article.

Haglund 2020 *(Continued)*

Funding: the author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by ALF Foundation, Region Skåne, and the Lindhaga Foundation.

Note(s): none

Hinnebusch 2017

Study characteristics

Methods	<p>Design: prospective Setting: cohort study Sample: unclear how sample was drawn Location (country): USA Length of follow-up (years): 2.2 Diagnostic tool (multidisciplinary or not): DSM-IV, ADOS, CARS (N)</p>
Population	<p>Sample size (% male): 219 (81) Diagnosis type: ASD Mean age at baseline (years): 2.13 IQ (mean standard score): both Language (mean standard score): < 70 Adaptive behaviour (mean standard score): -</p>
Notes	<p>Conflict of interest: Deborah Fein is part owner of M-CHAT-R, LLC, which receives royalties from companies that incorporate the M-CHAT-R into commercial products and charge for its use. Data reported in the current paper are from the freely available paper versions of the M-CHAT and M-CHAT-R. Alexander Hinnebusch and Lauren Miller declared that they had no COI. Funding: this study was funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Grant number R01HD039961) and the Maternal and Child Health Bureau (Grant number R40MC00270). Note(s): none</p>

Kim 2016

Study characteristics

Methods	<p>Design: prospective Setting: cohort study Sample: clinical Location (country): USA Length of follow-up (years): 1.3 Diagnostic tool (multidisciplinary or not): ADOS (Y)</p>
Population	<p>Sample size (% male): 100 (84) Diagnosis type: ASD Mean age at baseline (years): 1.8 IQ (mean standard score): > 70 Language (mean standard score): - Adaptive behaviour (mean standard score): both</p>
Notes	<p>Conflict of interest: the authors declared they have no COI. Funding: IMH grant #P50MH081756-0 awarded to Fred Volkmar, Ami Klin, Rhea Paul, and KC, NIMH grant #1R03MH086732 awarded to SM, and the Associates of Child Study Center Note(s): none</p>

Klintwall 2015

Study characteristics

Methods	Design: prospective Setting: cohort study Sample: clinical Location (country): USA Length of follow-up (years): 1.4 Diagnostic tool (multidisciplinary or not): ADOS, ADOS-T (U)
Population	Sample size (% male): 70 (89) Diagnosis type: ASD Mean age at baseline (years): 1.8 IQ (mean standard score): > 70 Language (mean standard score): < 70 Adaptive behaviour (mean standard score): > 70
Notes	Conflict of interest: not reported Funding: NIMH P50 MH081756 (Project 2, PI: KC), NIMH R01 MH087554 (PI: KC), and NICHD P01 HD003008 (Project 1, PI: KC) Note(s): none

Lombardo 2015

Study characteristics

Methods	Design: prospective Setting: cohort study Sample: clinical, drawn from broad community base (well baby 1 year check and community referral) Location (country): USA Length of follow-up (years): 1.1 (ASD poor language outcome) and 1 (ASD good language outcome) Diagnostic tool (multidisciplinary or not): ADOS (N)
Population	Sample size (% male): 24 (79) ASD poor and 36 (78) ASD good Diagnosis type: ASD Mean age at baseline (years): 2.0 (ASD poor) and 1.78 (ASD good) IQ (mean standard score): both Language (mean standard score): both Adaptive behaviour (mean standard score): both
Notes	Conflict of interest: not reported Funding: this work was supported by NIMH Autism Center of Excellence grant P50-MH081755 (EC), NIMH R01-MH080134 (KP), NFAR grant (KP), NIMH R01-MH036840 (EC), and fellowships from Jesus College, Cambridge and the British Academy (MVL). Note(s): contacted authors to request the required data. Authors responded quote: "we didn't compute a calibrated ADOS severity score, so unfortunately cannot provide that data". ADOS data had not been transformed to be able to be used. Authors reported that all children in the study retained their diagnosis at follow-up because that was a requirement for inclusion, so it is unclear how many from baseline may have no longer retained their diagnosis

Malhi 2011

Study characteristics

Overall prognosis of preschool autism spectrum disorder diagnoses (Review)

Copyright © 2022 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Malhi 2011 (Continued)

Methods	Design: prospective Setting: cohort study Sample: clinical Location (country): India Length of follow-up (years): 1.7 Diagnostic tool (multidisciplinary or not): CARS (Y)
Population	Sample size (% male): 77 (83) Diagnosis type: ASD Mean age at baseline (years): 2.5 IQ (mean standard score): < 70 Language (mean standard score): - Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: the authors declared they have no COI. Funding: none (self-funded - Department of Pediatrics, Post Graduate Institute of Medical Education and Research, Sector 12, Chandigarh) Note(s): none

Martin-Borreguero 2021
Study characteristics

Methods	Design: retrospective with whole cohort considered Setting: cohort study Sample: clinical Location (country): Spain Length of follow-up (years): 2 Diagnostic tool (multidisciplinary or not): DSM-5, CARS and ADOS but CARS repeated several times (Y)
Population	Sample size (% male): 52 (82-85) Diagnosis type: ASD Mean age at baseline (years): Between 2-6 years. Categorical data provided with no overall mean. IQ (mean standard score): - Language (mean standard score): - Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: the authors declared they have no COI. Funding: Spanish Society of Paediatrics and SPAOYEX Note(s): none

Moore 2003
Study characteristics

Methods	Design: retrospective Setting: cohort study Sample: clinical, from a broad community base Location (country): UK Length of follow-up (years): 1.6 Diagnostic tool (multidisciplinary or not): ADOS (N)
Population	Sample size (% male): 19 (80)

Overall prognosis of preschool autism spectrum disorder diagnoses (Review)

Moore 2003 (Continued)

Diagnosis type: ASD
Mean age at baseline (years): 2.8
IQ (mean standard score): > 70
Language (mean standard score): < 70
Adaptive behaviour (mean standard score): -

Notes **Conflict of interest:** not reported
 Funding: not reported
 Note(s): none

Moss 2008
Study characteristics

Methods **Design:** prospective
 Setting: intervention trial, with treatment and control arms
 Sample: clinical, from a broad community base
 Location (country): UK
 Length of follow-up (years): 7.0
 Diagnostic tool (multidisciplinary or not): ADI-R (N)

Population **Sample size (% male):** 35 (91)
 Diagnosis type: ASD
 Mean age at baseline (years): 3.5
 IQ (mean standard score): < 70
 Language (mean standard score): < 70
 Adaptive behaviour (mean standard score): < 70

Notes **Conflict of interest:** not reported. This was a publication of a PhD thesis. There was no information included in the paper about COI.
 Funding: Action Research
 Note(s): none

Naigles 2016
Study characteristics

Methods **Design:** prospective
 Setting: cohort study
 Sample: clinical, drawn from broad community base
 Location (country): USA
 Length of follow-up (years): 1.4
 Diagnostic tool (multidisciplinary or not): ADOS (N)

Population **Sample size (% male):** 15 (100)
 Diagnosis type: ASD
 Mean age at baseline (years): 2.6
 IQ (mean standard score): -
 Language (mean standard score): > 70
 Adaptive behaviour (mean standard score): -

Notes **Conflict of interest:** the authors declared they have no COI.
 Funding: National Institute on Deafness and Other Communication Disorders grant to L Naigles (Grant number: R01 DC007428)

Naigles 2016 (Continued)

Note(s): unable to extract change in diagnosis over time based on data provided in the paper. Only change in ADOS scores at V1 and V5 provided. Unable to obtain required data from study authors

Neuhaus 2016
Study characteristics

Methods	Design: prospective Setting: cohort study Sample: clinical, drawn from broad community base Location (country): USA Length of follow-up (years): 10.6 Diagnostic tool (multidisciplinary or not): ADOS-G and ADI-R (N)
Population	Sample size (% male): 26 (88) Diagnosis type: ASD Mean age at baseline (years): 3.7 IQ (mean standard score): > 70 Language (mean standard score): < 70 Adaptive behaviour (mean standard score): < 70
Notes	Conflict of interest: not reported Funding: support for this project was provided by NICHD and NIDCD PO1HD34565, and an Autism Speaks Meixner Translational Postdoctoral Fellowship (Neuhaus) Note(s): completed ADOS and ADI-R at multiple time points but did not present scores or diagnostic status at each time point. Unable to obtain required data from study authors

Ozonoff 2015
Study characteristics

Methods	Design: prospective Setting: cohort study Sample: clinical Location (country): USA Length of follow-up (years): 1.0 Diagnostic tool (multidisciplinary or not): ADI-R, DSM-IV, best clinical estimate (N)
Population	Sample size (% male): 79 (ND) Diagnosis type: ASD Mean age at baseline (years): 2.0 IQ (mean standard score): - Language (mean standard score): - Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: the authors declared they have no COI. Funding: National Institutes of Health Grants, Canadian Institutes of Health Research, Autism Speaks Canada Note(s): none

Paul 2008
Study characteristics

Methods	Design: prospective Setting: cohort study Sample: clinical, from a broad community base Location (country): USA Length of follow-up (years): 1.1 Diagnostic tool (multidisciplinary or not): ADOS (Y)
Population	Sample size (% male): 37 (ND) Diagnosis type: ASD Mean age at baseline (years): 1.82 IQ (mean standard score): > 70 Language (mean standard score): < 70 Adaptive behaviour (mean standard score): > 70
Notes	Conflict of interest: not reported Funding: National Institute of Mental Health (NIMH)P01-03008, National Institute on Deafness and Other Communication Disorders (NIDCD)U54 MH66494, The National Institute of Environmental Health Sciences (NIEHS), The National Institute of Child Health and Human Development (NICHD), The National Institute of Neurological Disorders and Stroke (NINDS), NIDCDK24 HD045576, The National Alliance for Autism Research Note(s): none

Qian 2018
Study characteristics

Methods	Design: retrospective Setting: cohort study Sample: clinical Location (country): China Length of follow-up (years): 2 Diagnostic tool (multidisciplinary or not): CARS, ADI-R (N)
Population	Sample size (% male): 37 (86) Diagnosis type: ASD Mean age at baseline (years): 2.57 IQ (mean standard score): < 70 Language (mean standard score): - Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: not reported Funding: Science Foundation of China (81771478), and the Jiangsu Provincial Key Research and Development Program (BE2016616) and the Major National Research and Development Program of China (2016YFC1306205) Note(s): none

Rivard 2019
Study characteristics

Methods	Design: prospective Setting: all had early behaviour intervention, no control arm
---------	--

Rivard 2019 (Continued)

	Sample: clinical Location (country): Canada Length of follow-up (years): 1 Diagnostic tool (multidisciplinary or not): GARS and CARS but only GARS used at outcome (N)
Population	Sample size (% male): 32 (66) Diagnosis type: ASD Mean age at baseline (years): 3.9 IQ (mean standard score): < 70 Language (mean standard score): - Adaptive behaviour (mean standard score): <70
Notes	Conflict of interest: the author(s) declared the following potential COI with respect to the research, authorship, and/or publication of this article: Amélie Terroux declared to be an employee of the Centre de réadaptation en déficience intellectuelle et troubles envahissant du développement de la Montérégie-Est. Marjorie Morin and Céline Mercier were also under contract for the same agency for the duration of the study. Funding: the author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by grants by the Montérégie Health Agency and the Québec Ministry of Health and Social Services to Céline Mercier and Mélina Rivard. Note(s): none

Robain 2020
Study characteristics

Methods	Design: prospective Setting: intervention study (RCT with treatment and control arm) Sample: clinical from broad community base Location (country): Switzerland Length of follow-up (years): 1 Diagnostic tool (multidisciplinary or not): ADOS (not described)
Population	Sample size (% male): 60 (100) Diagnosis type: ASD Mean age at baseline (years): 3 IQ (mean standard score): > 70 Language (mean standard score): - Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: the authors declared they have no COI. Funding: this study was supported by the National Center of Competence in Research "Synapsy," financed by the Swiss National Science Foundation (SNF, Grant Number: 51AU40_125759), by a SNF Grant to MS (#163859), and by the "Fondation Pôle Autisme" (https://www.pole-autisme.ch). Martina Franchini was also supported by an individual fellowship from the SNF (#P2GEP1_171686) Note(s): none

Santocchi 2012
Study characteristics

Methods	Design: not described Setting: cohort study Sample: not described Location (country): Italy
---------	--

Santocchi 2012 (Continued)

Length of follow-up (years): 1.75
Diagnostic tool (multidisciplinary or not): ADOS-G and CARS (not described)

Population	Sample size (% male): 98 (-) Diagnosis type: ASD Mean age at baseline (years): 3.25 IQ (mean standard score): - Language (mean standard score): - Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: not reported Funding: not described Note(s): conference abstract. We attempted to obtain the full text through multiple sources and contacted the study authors but did not receive a response so we have extracted as much data as possible from the abstract and have not been able to complete risk of bias assessment.

Sheinkopf 1998
Study characteristics

Methods	Design: prospective Setting: intervention trial, with control arm Sample: clinical, from a broad community base Location (country): USA Length of follow-up (years): 1.5 Diagnostic tool (multidisciplinary or not): DSM III (Y)
Population	Sample size (% male): 11 (ND) Diagnosis type: ASD Mean age at baseline (years): 2.94 IQ (mean standard score): < 70 Language (mean standard score): - Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: not reported Funding: not reported Note(s): none

Smith 2019
Study characteristics

Methods	Design: prospective Setting: intervention arm with one treatment arm Sample: clinical Location (country): Norway Length of follow-up (years): 12.41 Diagnostic tool (multidisciplinary or not): ADI-R (N)
Population	Sample size (% male): 19 (84) Diagnosis type: ASD Mean age at baseline (years): 2.92 IQ (mean standard score): < 70 Language (mean standard score): -

Smith 2019 (Continued)

Adaptive behaviour (mean standard score): <70

Notes

Conflict of interest: the authors declared they have no COI.

Funding: quote: "The author(s) received no financial support for the research, authorship, and/or publication of this article."

Note(s): none

Soke 2011
Study characteristics

Methods

Design: prospective
Setting: cohort study
Sample: clinical
Location (country): USA
Length of follow-up (years): 2.1
Diagnostic tool (multidisciplinary or not): ADI-R (Y)

Population

Sample size (% male): 28 (79)
Diagnosis type: AD
Mean age at baseline (years): 2.8
IQ (mean standard score): < 70
Language (mean standard score): -
Adaptive behaviour (mean standard score): -

Notes

Conflict of interest: not reported (Masters thesis)

Funding: Grant # U19HD35468 from the National Institute for Child Health and Human Development (NICHD)

Note(s): none

Solomon 2014
Study characteristics

Methods

Design: prospective
Setting: RCT, with control arm study
Sample: population based
Location (country): USA
Length of follow-up (years): 1.0
Diagnostic tool (multidisciplinary or not): ADOS (U)

Population

Sample size (% male): 55 (84)
Diagnosis type: ASD
Mean age at baseline (years): 4.2
IQ (mean standard score): < 70
Language (mean standard score): -
Adaptive behaviour (mean standard score): -

Notes

Conflict of interest: Small Business Innovation Research (SBIR) grants, sponsored by the NIMH, are administered through business entities to support research in technological innovation and dissemination. Therefore, all SBIR principle investigators if they are directly involved in the grant have a financial conflict of interest. R Solomon, the principle investigator of the study, was involved in the design of the study, wrote the first draft of the manuscript and was involved in the decision to submit the article. To limit his bias, R Solomon was assiduously excluded from evaluation of outcomes, data collection, or data analysis, all of which were done independently at MichiganState University under the auspices of

Solomon 2014 (Continued)

LA Van Egeren, a senior level researcher and director of the Community Evaluation and Research Collaborative. The only involvement with data occurred when the data collected at Easter Seals sites were identified at the “central office” in Ann Arbor and sent on to Michigan State University for analysis. R Solomon received no other funds outside of the grant (such as honoraria, consultant fees, etc.) before or during the grant. G Mahoney received a fee for consulting as an original part of the grant protocol. The remaining authors declared no COI.

Funding: National Institute of Mental Health (NIMH) and Small Business Innovation Research (SBIR) grant (grant# 2 R44 MH078431-02A)

Note(s): none

Solomon 2016
Study characteristics

Methods	Design: prospective Setting: cohort study Sample: clinical, from a broad community base Location (country): USA Length of follow-up (years): 2.8 Diagnostic tool (multidisciplinary or not): ADOS (N)
Population	Sample size (% male): 102 (80) Diagnosis type: ASD Mean age at baseline (years): 2.9 IQ (mean standard score): > 70 Language (mean standard score): both Adaptive behaviour (mean standard score): both
Notes	<p>Conflict of interest: Drs Solomon, Iosif, Reinhardt, Libero, Nordahl, Ozon-off, and Rogers reported no biomedical financial interests or conflicts of interest. Dr Amaral is on the Scientific Advisory Board of Stemina Biomaker Discovery and Axial Biotherapeutics.</p> <p>Funding: Dr Solomon was supported by National Institutes of Health (NIH) grants R01MH106518 and R01MH103284. Dr Nordahl was supported by R01MH104438. Dr Amaral was supported by R01MH103371. Dr Reinhardt was supported by 5T32MH073124. Ana-Maria Iosif, PhD provided statistical support as part of the MIND Institute Intellectual and Developmental Disabilities Research Center (U54 HD079125).</p> <p>Note(s): none</p>

Spjut Jansson 2016
Study characteristics

Methods	Design: prospective Setting: cohort study Sample: clinical Location (country): Sweden Length of follow-up (years): 2.0 Diagnostic tool (multidisciplinary or not): ADOS, DISCO, ADI-R (Y)
Population	Sample size (% male): 71 (79) Diagnosis type: ASD Mean age at baseline (years): 3.0 IQ (mean standard score): both Language (mean standard score): -

Spjut Jansson 2016 (Continued)

Adaptive behaviour (mean standard score): both

Notes

Conflict of interest: the authors declared they have no COI.

Funding: Health & Medical Care Committee of the Regional Executive Board, Region Västra Götaland (BSJ)

Note(s): none

Sullivan 2010
Study characteristics

Methods

Design: prospective

Setting: All participants enrolled in a community intervention (IBI).

Sample: clinical, from a broad community base

Location (country): Canada

Length of follow-up (years): 2.2

Diagnostic tool (multidisciplinary or not): CARS (N)

Population

Sample size (% male): 75 (83)

Diagnosis type: ASD

Mean age at baseline (years): 3.9

IQ (mean standard score): < 70

Language (mean standard score): -

Adaptive behaviour (mean standard score): < 70

Notes

Conflict of interest: not reported

Funding: Provincial Centre of Excellence for Child and Youth Mental Health at CHEO, Social Sciences and Humanities Research Council of Canada, and the Canadian Institutes of Health Research/National Alliance for Autism Research (Interdisciplinary Training Program in Autism Spectrum Disorders)

Note(s): none

Szatmari 2021
Study characteristics

Methods

Design: prospective

Setting: cohort study

Sample: clinical

Location (country): Canada

Length of follow-up (years): 7.36

Diagnostic tool (multidisciplinary or not): clinical consensus, ADOS and ADI-R but ADOS scores reported twice (N)

Population

Sample size (% male): 272 (86)

Diagnosis type: ASD

Mean age at baseline (years): 3.39

IQ (mean standard score): both

Language (mean standard score): both

Adaptive behaviour (mean standard score): both

Notes

Conflict of interest: Dr Szatmari reported receiving grants from Canadian Institutes of Health Research (CIHR) during the conduct of this study. Dr Cost reported receiving grants from the CIHR during the conduct of the study. Dr Bennett reported receiving grants from CIHR and grants from Hamilton Health Sciences Foundation during the conduct of the study; and grants from Hamilton Health Sciences and Brain Canada outside the submitted work. Dr Smith reported receiving grants from the Centre for Ad-

Szatmari 2021 (Continued)

diction and Mental Health/CIHR as 1 of 5 study site principle investigators during the conduct of the study. Dr Zwaigenbaum reported receiving personal fees from Roche as a data monitoring board member outside the submitted work. No other disclosures were reported.

Funding: this study was supported by the Canadian Institutes of Health Research, Kids Brain Health Network, Autism Speaks, the Government of British Columbia, Alberta Innovates Health Solutions, and the Sinneave Family Foundation. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Note(s): none

Takeda 2007
Study characteristics

Methods	Design: prospective Setting: cohort study Sample: clinical, from a broad community base Location (country): Japan Length of follow-up (years): 2.9 Diagnostic tool (multidisciplinary or not): ICD-10, CARS (N)
Population	Sample size (% male): 126 (81) Diagnosis type: ASD Mean age at baseline (years): 2.62 IQ (mean standard score): < 70 Language (mean standard score): - Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: not reported Funding: not reported Note(s): none

Thomas 2009
Study characteristics

Methods	Design: retrospective Setting: cohort study (community intervention received over 5 years) Sample: clinical, from a broad community base Location (country): USA Length of follow-up (years): 5.0 Diagnostic tool (multidisciplinary or not): CARS (U)
Population	Sample size (% male): 69 (79) Diagnosis type: ASD Mean age at baseline (years): 4.4 IQ (mean standard score): - Language (mean standard score): both Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: not reported (thesis) Funding: not reported (thesis) Note(s): none

Venker 2014

Study characteristics

Methods	Design: prospective Setting: cohort study Sample: clinical, from a broad community base Location (country): USA Length of follow-up (years): 5.9 Diagnostic tool (multidisciplinary or not): ADOS, DSM-IV (Y)
Population	Sample size (% male): 129 (87) Diagnosis type: ASD Mean age at baseline (years): 2.8 IQ (mean standard score): > 70 Language (mean standard score): both Adaptive behaviour (mean standard score): > 70
Notes	Conflict of interest: not reported Funding: NIH R01DC007223-05 (Ellis Weismer, PI; Gernsbacher, co-PI); T32DC005359-10 (Ellis Weismer, PI); P30HD003352-46 (Seltzer, PI) Note(s): none

Wu 2016

Study characteristics

Methods	Design: retrospective Setting: cohort study Sample: clinical Location (country): USA Length of follow-up (years): 1.4 Diagnostic tool (multidisciplinary or not): DSM-IV-TR, file record review (N)
Population	Sample size (% male): 8564 (83) Diagnosis type: ASD Mean age at baseline (years): 3.7 IQ (mean standard score): - Language (mean standard score): - Adaptive behaviour (mean standard score): -
Notes	Conflict of interest: the authors declared they have no COI. Funding: the sponsor for the data collection of the Autism and Developmental Disabilities Monitoring (ADDM) Network is the US Centers for Disease Control and Prevention (CDC). Note(s): none

Zappella 1990

Study characteristics

Methods	Design: retrospective, with whole cohort considered Setting: intervention trial, with one treatment arm Sample: clinical, from a broad community base
---------	--

Zappella 1990 *(Continued)*

Location (country): Italy
Length of follow-up (years): 1.8
Diagnostic tool (multidisciplinary or not): DSM-III (N)

Population **Sample size (% male):** 15 (87)
Diagnosis type: AD
Mean age at baseline (years): 4.5
IQ (mean standard score): > 70
Language (mean standard score): -
Adaptive behaviour (mean standard score): both

Notes **Conflict of interest:** not reported
Funding: not reported
Note(s): none

Zappella 2010

Study characteristics

Methods **Design:** retrospective
Setting: cohort study
Sample: clinical
Location (country): Italy
Length of follow-up (years): 2.7
Diagnostic tool (multidisciplinary or not): DSM-IV-TR (U)

Population **Sample size (% male):** 534 (84)
Diagnosis type: ASD
Mean age at baseline (years): 5.0
IQ (mean standard score): -
Language (mean standard score): -
Adaptive behaviour (mean standard score): -

Notes **Conflict of interest:** not reported
Funding: not reported
Note(s): none

Zwaigenbaum 2015

Study characteristics

Methods **Design:** prospective
Setting: cohort study
Sample: clinical, from a broad community base
Location (country): Canada
Length of follow-up (years): 1.5
Diagnostic tool (multidisciplinary or not): DSM-IV-TR (N)

Population **Sample size (% male):** 23 (69)
Diagnosis type: ASD
Mean age at baseline (years): 1.5
IQ (mean standard score): -
Language (mean standard score): -
Adaptive behaviour (mean standard score): > 70

Zwaigenbaum 2015 (Continued)

Notes

Conflict of interest: Dr Zwaigenbaum was supported by the Stollery Children's Hospital Foundation Chair in Autism Research. Drs Bryson and Smith were supported by the Jack and Joan Craig Chair in Autism Research, Dr Szatmari was supported by the Chedoke Health Chair in Child Psychiatry, and Dr Vaillancourt was supported by a Canada Research Chair in Children's Mental Health and Violence Protection.

Funding: Canadian Institutes of Health Research (grant numbers 62924 and 102665), Autism Speaks Canada and NeuroDevNet

Note(s): none

Mean scores (IQ, adaptive behaviour or language) for the cohort is < 70 or more than 70% are less than 70. If cohort evenly spread this is signified 'both'.

AD: autistic disorder; **ADOS:** Autism Diagnostic Observation Schedule; **ADI:** Autism Diagnostic Interview; **ASD:** autism spectrum disorder; **CARS:** Childhood Autism Rating Scale; **COI:** conflict of interest; **DSM:** Diagnostic Statistical Manual of Mental Disorders; **ICD:** International Classification of Diseases; **IQ:** intelligence quotient; **N:** no, not multidisciplinary; **NIH:** National Institutes of Health; **PDD-NOS:** pervasive developmental disorder; **PI:** principal investigator; **RCT:** randomised controlled trial; **U:** unclear; **Y:** yes, multidisciplinary; **-:** not reported by study

Characteristics of excluded studies [ordered by study ID]

Study	Reason for exclusion
Bacon 2018	Study only followed up those children who had consistent diagnosis at two time points and excluded those who moved off spectrum. Unable to obtain information on numbers that moved off spectrum from study authors.
Bal 2019	Did not use same diagnostic methods/tools at baseline and follow-up
Berry 2009	Did not use same diagnostic methods/tools at baseline and follow-up
Canal-Bedia 2016	Did not use same diagnostic methods/tools at baseline and follow-up
Charman 2005	Did not use same diagnostic methods/tools at baseline and follow-up
Chawarska 2007	Did not use same diagnostic methods/tools at baseline and follow-up
Clark 2017	Did not use same diagnostic methods/tools at baseline and follow-up
De Giacomo 2009	Did not use same diagnostic methods/tools at baseline and follow-up
Guthrie 2013	Did not use same diagnostic methods/tools at baseline and follow-up.
Hedvall 2014	Did not use same diagnostic methods/tools at baseline and follow-up
Jónsdóttir 2007	Did not use same diagnostic methods/tools at baseline and follow-up
Kadam 2021	Did not use same diagnostic methods/tools at baseline and follow-up
Kantzer 2018	Did not use same diagnostic methods/tools at baseline and follow-up
Ozonoff 2018	No baseline diagnostic assessment
Shumway 2012	Did not use same diagnostic methods/tools at baseline and follow-up
Stone 1999	Did not use same diagnostic methods/tools at baseline and follow-up

Study	Reason for exclusion
Sutera 2010	Did not use same diagnostic methods/tools at baseline and follow-up
Thurm 2015	Did not use same diagnostic methods/tools at baseline and follow-up
Tunc 2021	Sample only included those diagnosed at baseline and follow-up therefore unable to determine how many may have lost diagnosis at follow-up.
Van Daalen 2009	Did not use same diagnostic methods/tools at baseline and follow-up
Venter 1992	Did not use same diagnostic methods/tools at baseline and follow-up

Characteristics of studies awaiting classification *[ordered by study ID]*

[Anglim 2012](#)

Notes	Conference abstract. May meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	--

[Boi 2017](#)

Notes	Conference abstract. May meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	--

[Brown 1997](#)

Notes	May meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	---

[Chang 2017](#)

Notes	Conference abstract. Appears to meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	---

[Chang 2017a](#)

Notes	Appears to meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	--

Da Silva 2003

Notes	May meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	---

Eapen 2019

Notes	Conference abstract. Appears to meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	---

Faroghizadeh 2021

Notes	May meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	---

Gabis 2011

Notes	Conference abstract. May meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data. Appears to have overlapping participants with Millikovsky-Ayalon 2012 which is also awaiting classification.
-------	---

Ghamari Kivi 2012

Notes	Conference abstract. May meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	--

Jimenez-Martinez 2018

Notes	Conference abstract. May meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	--

Melville 1987

Notes	Dissertation. We requested a copy of the paper from numerous libraries, but given the date of the thesis, which is in hard copy (no digital copy), we were unable to access the full thesis.
-------	--

Millikovsky-Ayalon 2012

Notes	Study is in Hebrew. We contacted the authors for further information (an email was written to the authors in Hebrew) but we did not receive any response.
-------	---

Mohanta 2019

Notes	Conference abstract. May meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	--

Mosconi 2009

Notes	Insufficient data to accurately classify as included or excluded. Contacted study authors for further data but did not receive a response
-------	---

Muratori 2002

Notes	Study is in Italian. We had this article translated from Italian but require further information from the study authors to determine whether the study is eligible for inclusion. We emailed the authors and are awaiting a response.
-------	---

Ozyurt 2020

Notes	May meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	---

Perucchini 2005

Notes	Study is in Italian. Article was translated in Italian but further information was required from the authors to determine whether the study was eligible for inclusion. We emailed the authors and are awaiting a response. Participants may overlap with Muratori 2002
-------	---

Selvakumar 2018

Notes	Study meets eligibility criteria, however it may include the same participants as another included study by the same authorship group (Malhi 2011). We contacted the study authors several times to confirm whether the participants were the same/overlapping but did not receive a response.
-------	--

Takesada 1992

Notes	May meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	---

Zhang 2019

Notes	May meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	---

Zirakashvili 2018

Notes	Conference abstract. May meet eligibility criteria. We attempted to obtain the full text through multiple sources and contacted the study authors but were unable to confirm eligibility for inclusion or obtain data.
-------	--

ADDITIONAL TABLES
Table 1. Changes to the classification systems over time

Year published	Classification system	Subgroups (as specified in the classification system)
1975	International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)	Autistic disorder
1980	Diagnostic and Statistical Manual of Mental Disorders, Third Edition (DSM-III)	PDD: infantile autism, childhood onset PDD and atypical PDD
1987	Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised (DSM-III-R)	PDD: autistic disorder, PDD-not otherwise specified (PDD-NOS)
1994 to 2000	Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)	Asperger's disorder, autistic disorder, PDD-NOS
1996	International Classification of Diseases, Tenth Revision (ICD-10)	Childhood autism, Asperger's syndrome, atypical autism, pervasive developmental disorder (PDD) - unspecified
2000 to 2013	Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR)	Asperger's disorder, autistic disorder, PDD-NOS
2013 to current	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)	Autism spectrum disorder
2018	International Classification of Diseases, Eleventh Revision (ICD-11)	Autism spectrum disorder

PDD-NOS: pervasive developmental disorder

Table 2. Studies that had multiple publications

Primary publication	Additional publications from the same study
Anderson 2009^a	Anderson 2007, Bedford 2016, Gotham 2012, Gotham 2011, Hus 2011, Lord 1995, Lord 2004, Lord 2012 , Luyster 2007, Pickles 2014, Richler 2010, Thurm 2007
Baghdadli 2012	Baghdadli 2018, Baghdadli 2008, Baghdadli 2007, Darrou 2010, Pry 2011, Pry 2012

Overall prognosis of preschool autism spectrum disorder diagnoses (Review)

Copyright © 2022 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Table 2. Studies that had multiple publications *(Continued)*

Bopp 2006	Bopp 2009; Smith 2007
Flanagan 2010	Flanagan 2012
Giserman-Kiss 2020	Giserman-Kiss 2018
Moss 2008	Magiati 2007, Magiati 2011a, Magiati 2011b
Qian 2018	Ke 2017, Li 2019
Rivard 2019	Mello 2018
Solomon 2014	Mahoney 2016
Solomon 2016	Solomon 2018, Waizbard-Bartov 2021
Szatmari 2021	Baribeau 2020, Baribeau 2021, Courchesne 2021, Bennett 2014, Bennett 2015, Georgiades 2014, Georgiades 2021, Szatmari 2015
Venker 2014	Ellis-Weismer 2015, Davidson 2017, Ray-Subramanian 2012, Venker 2016

^aMet inclusion criteria but unable to extract data for synthesis as children without a diagnosis of autism spectrum disorder were also included in the cohort. Authors were contacted but we were unable to obtain required data.

Table 3. Characteristics of individual studies included in synthesis (n = 34)

Study	Diagnosis type	N at baseline (% male)	IQ (mean standard score) ^a	Adaptive behavior (mean standard score) ^a	Language (mean standard score) ^a	Age at baseline (years)	Follow-up duration (years)	Diagnostic tool used at baseline (multidisciplinary or not)	Proportion who met diagnostic criteria at follow-up
Baghdadli 2012	ASD	152 (82)	< 70	< 70	NR	4.90	3.00	ICD-10 & CARS (Y)	1.0
Benedetto 2021	ASD	147 (80)	NR	NR	NR	2.3	1	DSM-5 and ADOS (Y)	0.73
Brian 2016	ASD	18 (72)	> 70	NR	> 70	3.15	6.36	DSM-IV-TR & ADOS (N)	0.94
Demb 1989	ASD	12 (75)	< 70	NR	NR	4.50	5.00	DSM-III & DSM-III R (N)	0.83
Eaves 2004	ASD	43 (80)	< 70	< 70	NR	2.75	2.25	DSM-IV, CARS, MDT (Y)	0.93
Elmose 2014	ASD	23 (78)	NR	NR	NR	3.10	8.30	ICD-10, ADOS (Y)	1.00
Flanagan 2010	ASD	67 (82)	NR	< 70	NR	3.59	1.38	CARS (N)	0.81
Freeman 2004	ASD	59 (81)	< 70	NR	NR	4.00	2.2	DSM IV, CARS (N)	0.97
Gillberg 1990	ASD	25 (68)	< 70	NR	NR	1.13	4.04	DSM-III-R (N)	0.92
Giserman-Kiss 2020	ASD	60 (87)	<70	NR	<70	2.31	1.98	ADOS	0.883
Gonzalez 1993	ASD	30 (73)	< 70	NR	NR	4.50	1.00	DSM-III, DSM-III-R, DSM-IV and ICD 10 (N)	0.97
Hinnebusch 2017	ASD	219 (81)	Both	NR	< 70	2.13	2.16	DSM-IV, ADOS, CARS (N)	0.83
Kim 2016	ASD	100 (84)	> 70	Both	NR	1.80	1.30	ADOS (Y)	0.93
Klintwall 2015	ASD	70 (89)	> 70	> 70	< 70	1.83	1.36	ADOS G, ADOS T (U)	0.93
Malhi 2011	ASD	77 (83)	< 70	NR	NR	2.48	1.65	CARS (Y)	0.95
Moore 2003	ASD	19 (80)	> 70	NR	< 70	2.83	1.59	ADI-R (Y)	1.00
Moss 2008	ASD	35 (91)	< 70	< 70	< 70	3.5	7.00	ADI-R (N)	0.80

Table 3. Characteristics of individual studies included in synthesis (n = 34) (Continued)

Ozonoff 2015	ASD	79 (NR)	NR	NR	NR	2	1	ADI-R, DSM IV, best clinical estimate (N)	0.82
Paul 2008	ASD	37 (NR)	> 70	> 70	< 70	1.82	1.09	ADOS (Y)	1.00
Qian 2018	ASD	37 (86)	<70	NR	NR	2.57	2	DSM IV TR; CARS ADI-R (N)	1.00
Robain 2020	ASD	60 (100)	>70	NR	NR	3	1	DSM 5 ADOS (N)	1.00
Santocchi 2012	ASD	98 (NR)	NR	NR	NR	3.25	1.75	ADOS, CARS	0.86
Sheinkopf 1998	ASD	11 (NR)	< 70	NR	NR	2.94	1.51	DSM-III (Y)	1.00
Soke 2011	AD	28 (79)	< 70	NR	NR	2.75	2.08	ADI-R (Y)	0.89
Solomon 2014	ASD	55 (84)	< 70	NR	NR	4.21	1.00	ADOS (U)	0.78
Solomon 2016	ASD	102 (80)	> 70	Both	Both	2.86	2.76	ADOS (N)	0.95
Spjut Jansson 2016	ASD	71 (79)	Both	Both	NR	3.03	2.00	ADOS, DISCO, ADI-R (Y)	0.93
Sullivan 2010	ASD	75 (83)	< 70	< 70	NR	3.94	2.18	CARS (N)	0.53
Takeda 2007	ASD	126 (81)	< 70	NR	NR	2.62	2.90	ICD-10, CARS (N)	1.00
Venker 2014	ASD	129 (87)	> 70	> 70	Both	2.80	5.85	DSM-IV, ADOS (Y)	1.00
Wu 2016	ASD	8564 (83)	NR	NR	NR	3.67	1.43	DSM-IV-TR file record review (N)	0.91
Zappella 1990	AD	15 (87)	> 70	Both	NR	4.50	1.83	DSM-III (N)	0.60
Zappella 2010	ASD	534 (84)	NR	NR	NR	5.00	2.67	DSM-IV-TR (U)	0.93
Zwaigenbaum 2015	ASD	23 (69)	NR	> 70	NR	1.50	1.50	DSM-IV-TR (N)	0.83

^aMean score (IQ, adaptive behaviour or language) for the cohort is < 70 or more than 70% are less than 70. If cohort evenly spread this is signified 'both'.

AD: autistic disorder; **ADI:** Autism Diagnostic Interview; **ADOS:** Autism Diagnostic Observation Schedule; **ASD:** autism spectrum disorder; **CARS:** Childhood Autism Rating Scale; **DISCO:** Diagnostic Interview for Social and Communication Disorders; **DSM:** Diagnostic Statistical Manual of Mental Disorders; **ICD:** International Classification of Diseases; **IQ:** intelligence quotient; **N:** no; **NR:** not reported; **PDD-NOS:** pervasive developmental disorder- not otherwise specified; **U:** unclear; **Y:** yes.

Table 4. Prognostic factor analyses (eight comparisons), with effect sizes and confidence intervals

Domain		Relative effect (95% CIs)	No. of participants (studies)	I ²
Age at baseline	0 to 2 years	0.94 (0.88 to 0.98)	251 (5 studies)	52.64%, P = 0.08
	2 to 3 years	0.92 (0.88 to 0.95)	9989 (22 studies)	90.17%, P < 0.01
	4 to 5 years	0.91 (0.76 to 0.99)	152 (5 studies)	90.48%, P < 0.01
	5 to 6 years	0.93 (0.90 to 0.95)	534 (1 study)	-
Age at follow-up	< 4 years	0.89 (0.79 to 0.96)	443 (6 studies)	86.80%, P < 0.01
	4 to 6 years	0.92 (0.88 to 0.95)	9794 (21 studies)	87.88%, P < 0.01
	7 to 12 years	0.96 (0.89 to 1.00)	868 (7 studies)	88.18%, P < 0.01
Duration of follow-up	1 to 2 years	0.91 (0.88 to 0.94)	10,745 (27 studies)	87.86%, P < 0.01
	2 to 5 years	0.99 (0.92 to 1.00)	293 (4 studies)	78.16%, P < 0.01
	6 to 17 years	0.92 (0.77 to 1.00)	67 (3 studies)	-
Decade of publication	1980 to 1989	0.83 (0.55 to 0.95)	12 (1 studies)	-
	1990 to 1999	0.91 (0.74 to 1.00)	82 (4 studies)	73.16% P = 0.01
	2000 to 2009	0.98 (0.93 to 1.00)	479 (7 studies)	80.57% P < 0.01
	2010 to 2019	0.90 (0.87 to 0.93)	10,273 (19 studies)	86.84% P < 0.01
	2020 to 2029	0.90 (0.68 to 1.00)	259 (3 studies)	-
Intelligence ^a	< 70	0.93 (0.85 to 0.98)	793 (15 studies)	90.88%, P < 0.01
	> 70	0.97 (0.92 to 1.00)	502 (9 studies)	77.54%, P < 0.01
	Both < 70 and > 70	0.86 (0.81, 0.89)	289 (2 studies)	-
Language ^a	< 70	0.92 (0.84 to 0.98)	382 (6 studies)	79.65%, P < 0.01
	> 70	0.94 (0.74 to 0.99)	18 (1 study)	-
	Both	0.98 (0.96 to 1.00)	205 (2 studies)	-
Adaptive behaviour ^a	< 70	0.85 (0.60 to 0.99)	300 (5 studies)	96.33%, P < 0.01
	> 70	0.97 (0.8 to 1.00)	233 (4 studies)	83.28%, P < 0.01
	Both	0.91 (0.82, 0.97)	283 (4 studies)	73.88%, P = 0.01
Multidisciplinary assessment	Yes	0.97 (0.91 to 1.00)	767 (13 studies)	87.97%, P < 0.01
	No	0.88 (0.83 to 0.93)	9468 (16 studies)	89.46%, P < 0.01

^aMean score (IQ, adaptive behaviour or language) for the cohort is < 70 or more than 70% are less than 70. If cohort was evenly spread this is signified 'both'.

CI: confidence interval; **I²:** a statistic that describes the percentage of variation across studies; **No.:** number.

APPENDICES

Appendix 1. Search strategies

Ovid MEDLINE

12 October 2017 (7283 records)

5 July 2021 (3916 records)

- 1 child development disorders, pervasive/
- 2 asperger syndrome/
- 3 autism spectrum disorder/
- 4 autistic disorder/
- 5 autis\$.tw.
- 6 asperger\$.tw.
- 7 pervasive development\$ disorder\$.tw.
- 8 (child\$ adj3 pervasiv\$).tw.
- 9 (PDD adj3 (specified or unspecified)).tw.
- 10 PDD-NOS.tw.
- 11 or/1-10
- 12 prognosis/
- 13 prognos\$.tw,kf.
- 14 prevalence/
- 15 prevalenc\$.tw,kf.
- 16 follow up studies/
- 17 (follow\$ up\$ or followup\$).tw.
- 18 ((diagnos\$ or temporal\$) adj3 (change\$ or stable or unstable or reliab\$ or stabili#e\$ or stability or instability or re-evaluat\$)).tw,kf.
- 19 ((developmental or diagnos\$) adj1 (outcome\$ or trajector\$)).tw.
- 20 (diagnos\$ adj1 (baseline or base-line or early or earlier or first or improve\$ or improving or initial\$ or original\$ or previous\$)).ab.
- 21 (diagnos\$ adj1 (final or second or later or subsequent\$)).ab.
- 22 (outcome\$ adj2 (change\$ or improve\$ or severe or severity or trajector\$ or worse or worst or worsen\$)).ab.
- 23 (symptom\$ adj2 (change\$ or improve\$ or improving or reduc\$ or severe or severity or trajector\$ or worse or worst or worsen\$)).ab.
- 24 ((measure\$ or score\$ or rating\$) adj2 (change\$ or improve\$ or improving or severe or severity or worse or worst or worsen\$)).ab.
- 25 predict\$.ab,kf.
- 26 or/12-25
- 27 11 and 26
- 28 ((autis\$ or asperger\$ or pervasive) and (improve\$ or improving or stability or stable)).ti.
- 29 27 or 28 [Final line of 2017 search]
- 30 (201710* or 2018* or 2019* or 2020* or 2021*).dt,ez,da.
- 31 29 and 30
- 32 remove duplicates from 31 [Final line of 2021 search]

Ovid Epub Ahead of Print (via Ovid MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations, Daily and Versions(R))

12 October 2017 (412 records)

5 July 2021 (413 records)

- 1 autis\$.tw.
- 2 asperger\$.tw.
- 3 pervasive development\$ disorder\$.tw,kf.
- 4 (child\$ adj3 pervasiv\$).tw,kf.
- 5 (PDD adj3 (specified or unspecified)).tw,kf.
- 6 PDD-NOS.tw,kf.
- 7 or/1-6
- 8 prognos\$.tw,kf.
- 9 prevalenc\$.tw,kf.)
- 10 (follow\$ up\$ or followup\$).tw,kf.

Overall prognosis of preschool autism spectrum disorder diagnoses (Review)

Copyright © 2022 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

11 ((diagnos\$ or temporal\$) adj3 (change\$ or stable or unstable or reliab\$ or stabili#e\$ or stability or instability or re-evaluat\$)).tw,kf.
 12 ((developmental or diagnos\$) adj1 (outcome\$ or trajector\$)).tw,kf.
 13 (diagnos\$ adj1 (baseline or base-line or early or earlier or first or initial\$ or original\$ or previous\$)).ab.
 14 (diagnos\$ adj1 (final or second or later or subsequent\$)).ab.
 15 (outcome\$ adj2 (change\$ or improve\$ or improving or severe or severity\$ or trajector\$ or worse or worst or worsen\$)).ab.
 16 (symptom\$ adj2 (change\$ or improve\$ or improving or reduc\$ or severe or severity\$ or trajector\$ or worse or worst or worsen\$)).ab.
 17 ((measure\$ or score\$ or rating\$) adj2 (change\$ or improve\$ or improving or severe or severity\$ or worse or worst or worsen\$)).ab.
 18 predict\$.ab,kf.
 19 or/8-18
 20 7 and 19
 21 ((autis\$ or asperger\$ or pervasive) and (improve\$ or improving or stability or stable)).ti.
 22 20 or 21
 23 limit 22 to publisher

Ovid MEDLINE In-Process, In-Data-Review & Other Non-Indexed Citations (via Ovid MEDLINE(R)) and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations, Daily and Versions(R)

2 October 2017 (1296 records)

5 July 2021 (1457 records)

1 autis\$.tw.
 2 asperger\$.tw.
 3 pervasive development\$ disorder\$.tw,kf.
 4 (child\$ adj3 pervasiv\$).tw,kf.
 5 (PDD adj3 (specified or unspecified)).tw,kf.
 6 PDD-NOS.tw,kf.
 7 or/1-6
 8 prognos\$.tw,kf.
 9 prevalenc\$.tw,kf.
 10 (follow\$ up\$ or followup\$).tw,kf.
 11 ((diagnos\$ or temporal\$) adj3 (change\$ or stable or unstable or reliab\$ or stabili#e\$ or stability or instability or re-evaluat\$)).tw,kf.
 12 ((developmental or diagnos\$) adj1 (outcome\$ or trajector\$)).tw,kf.
 13 (diagnos\$ adj1 (baseline or base-line or early or earlier or first or initial\$ or original\$ or previous\$)).ab.
 14 (diagnos\$ adj1 (final or second or later or subsequent\$)).ab.
 15 (outcome\$ adj2 (change\$ or improve\$ or improving or severe or severity or trajector\$ or worse or worst or worsen\$)).ab.
 16 (symptom\$ adj2 (change\$ or improve\$ or improving or reduc\$ or severe or severity or trajector\$ or worse or worst or worsen\$)).ab.
 17 ((measure\$ or score\$ or rating\$) adj2 (change\$ or improve\$ or improving or severe or severity\$ or worse or worst or worsen\$)).ab.
 18 predict\$.ab,kf.
 19 or/8-18
 20 7 and 19
 21 ((autis\$ or asperger\$ or pervasive) and (improve\$ or improving or stability or stable)).ti.
 22 20 or 21
 23 limit 22 to ("in data review" or in process or "pubmed not medline") [Annotation: Final line 2017]
 24 (201710* or 2018* or 2019* or 2020* or 2021*).dt,ed,ez.
 25 23 and 24 [Annotation: Final line 2021]

Embase Ovid

12 October 2017 (8263 records)

5 July 2021 (4452 records)

1 *autism/ or *asperger syndrome/ or *childhood disintegrative disorder/ or *"pervasive developmental disorder not otherwise specified"/
 2 autis\$.tw.
 3 asperger\$.tw.
 4 pervasive development\$ disorder\$.tw.
 5 (pervasive adj3 child\$).tw.
 6 PDD-NOS\$.tw.
 7 or/1-6
 8 *prognosis/
 9 prognos\$.tw,kw.
 10 *follow up/
 11 (follow\$ up\$ or followup\$).tw.
 12 ((diagnos\$ or temporal\$) adj3 (change\$ or improve\$ or stable or unstable or reliab\$ or stabili#e\$ or stability or instability or re-evaluat\$)).tw.

Overall prognosis of preschool autism spectrum disorder diagnoses (Review)

Copyright © 2022 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

13 ((developmental or diagnos\$) adj1 (outcome\$ or trajector\$)).tw.
 14 (diagnos\$ adj1 (baseline or base-line or early or earlier or first or initial\$ or original\$ or previous\$)).ab.
 15 (diagnos\$ adj1 (final or second or later or subsequent\$)).ab.
 16 (outcome\$ adj2 (change\$ or improve\$ or improving or severe or severity or trajector\$ or worse or worst or worsen\$)).ab.
 17 (symptom\$ adj2 (baseline or change\$ or improve\$ or improving or reduc\$ or severe or severity or trajector\$ or worse or worst or worsen\$)).ab.
 18 ((measure\$ or score\$ or level\$) adj2 (baseline or change\$ or improve\$ or improving or severe or severity or worse or worst or worsen\$)).ab.
 19 (predict\$ adj3 (baseline or change\$ or course or reduc\$ or severe or severity or trajector\$ or worse or worst or worsen\$)).ab.
 20 predict\$.kw.
 21 or/8-20
 22 7 and 21
 23 ((autis\$ or asperger\$ or pervasive) and (improve\$ or improving or stability or stable)).ti.
 24 22 or 23 [Annotation: Final line 2017]
 25 limit 24 to yr="2017 -Current"
 26 remove duplicates from 25 [Annotation: Final line 2021]

CINAHL Plus EBSCOhost

12 October 2017 (3260 records)
 6 July 2021 (2070 records)

S1 (MM "Child Development Disorders, Pervasive")
 S2 (MM "Autistic Disorder")
 S3 (MM "Asperger Syndrome")
 S4 (MM "Pervasive Developmental Disorder-Not Otherwise Specified")
 S5 TI(autis*) or AB (autis*)
 S6 TI(asperger*) or AB (asperger*)
 S7 TI (pervasive development* disorder*) OR AB (pervasive development* disorder*)
 S8 TI (child* N3 pervasiv*) OR AB (child* N3 pervasiv*)
 S9 TI ((PDD N3 (specified or unspecified))) OR AB ((PDD N3 (specified or unspecified)))
 S10 TI PDD-NOS OR AB PDD-NOS
 S11 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10
 S12 (MM "Prognosis")
 S13 TI prognos* OR AB prognos*
 S14 (MH "Prospective Studies+")
 S15 TI ("follow up" or follow-up) OR AB("follow up" or follow-up)
 S16 TI (((developmental or diagnos*) N1 (outcome* or trajector*))) OR AB (((developmental or diagnos*) N1 (outcome* or trajector*)))
 S17 TI (((diagnos* or temporal*) N3 (change* or stable or unstable or reliab* or stabili* or stability or instability or re-evaluat*))) OR AB (((diagnos* or temporal*) N3 (change* or stable or unstable or reliab* or stabili* or instability or re-evaluat*)))
 S18 AB (diagnos* N1 (baseline or base-line or early or earlier or first or initial* or original* or previous*))
 S19 AB (diagnos* N1 (final or second or later or subsequent*))
 S20 AB (outcome* N2 (change* or improve* or improving or severe or severity or trajector* or worse or worst or worsen*))
 S21 AB (symptom* N2 (change* or improve* or improving or reduc* or severe or severity or trajector* or worse or worst or worsen*))
 S22 AB ((measure* or score* or rating*) N2 (change* or improve* or improving or severe or severity or worse or worst or worsen*))
 S23 AB predict*
 S24 S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23
 S25 S11 AND S24
 S26 TI ((autis* or asperger* or pervasive) and (improve* or improving or stability or stable))
 S27 S25 OR S26 [Annotation: Final line 2017]
 S28 EM 20171001-
 S29 S27 AND S28 [Annotation: Final line 2021]

APA PsycINFO OVID

12 October 2017 (7476 records)
 5 July 2021 (3286 records)

1 autism spectrum disorders/
 2 autis\$.tw.
 3 asperger\$.tw.
 4 pervasive development\$ disorder\$.tw.
 5 (child\$ adj3 pervasiv\$).tw.
 6 (PDD adj3 (specified or unspecified)).tw.

Overall prognosis of preschool autism spectrum disorder diagnoses (Review)

Copyright © 2022 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

7 PDD-NOS.tw.
 8 or/1-7
 9 prognosis/
 10 prognos\$.tw.
 11 disease course/
 12 "severity (disorders)"/
 13 (follow\$ up\$ or followup\$).tw.
 14 ((diagnos\$ or temporal\$) adj3 (change\$ or stable or unstable or reliab\$ or stabili#e\$ or stability or instability or re-evaluat\$)).tw.
 15 ((developmental or diagnos\$) adj1 (outcome\$ or trajector\$)).tw.
 16 (diagnos\$ adj1 (baseline or base-line or early or earlier or first or initial\$ or original\$ or previous\$)).ab.
 17 (outcome\$ adj2 (change\$ or improve\$ or improving or sever\$ or trajector\$ or worse or worst or worsen\$)).ab.
 18 (symptom\$ adj2 (change\$ or improve\$ or improving or reduc\$ or sever\$ or trajector\$ or worse or worst or worsen\$)).ab.
 19 ((measure\$ or score\$ or rating\$) adj2 (change\$ or improve\$ or improving or sever\$ or worse or worst or worsen\$)).ab.
 20 predict\$.ab.
 21 or/9-20
 22 8 and 21
 23 ((autis\$ or asperger\$ or pervasive) and (improve\$ or improving or stability or stable)).ti.
 24 22 or 23 [Annotation: Final line 2017]
 25 limit 24 to up=20171001-20210628
 26 remove duplicates from 25 [Annotation: Final line 2021]

Conference Proceedings Citation Index-Science (CPCI-S) and Conference Proceedings Citation Index-Social Sciences & Humanities (CPCI-SSH) (searched via Web of Science Clarivate)

12 October 2017 CPCI-S (328 records) CPCI-SSH (47 records)
 6 July 2021 CPCI-S (273 records) CPCI-SSH (4 records)
 #17 #12 OR #11
 Indexes=CPCI-SSH Timespan=2017-2021 [Annotation: Final line 2021]
 # 16 #12 OR #11
 Indexes=CPCI-S Timespan=2017-2021 [Annotation: Final line 2021]
 #15 #12 OR #11
 Indexes=CPCI-SSH Timespan=All years [Annotation: Final line 2017]
 # 14 #12 OR #11
 Indexes=CPCI-S Timespan=All years [Annotation: Final line 2017]
 #12 TI=((autis* or asperger* or pervasive) and (improve* or improving or stability or stable))
 Indexes=CPCI-S, CPCI-SSH Timespan=All years
 #11 #10 AND #1
 Indexes=CPCI-S, CPCI-SSH Timespan=All years
 #10 #9 OR #8 OR #7 OR #6 OR #5 OR #4 OR #3 OR #2
 Indexes=CPCI-S, CPCI-SSH Timespan=All years
 #9 TS=((measure* or score* or rating*) Near/1 (change* or improve* or improving or severe* or severity or worse or worst or worsen*))
 Indexes=CPCI-S, CPCI-SSH Timespan=All years
 #8 TS=(symptom* Near/1 (change* or improve* or improving or severe* or severity or reduc* or trajector* or worse or worst or worsen*))
 Indexes=CPCI-S, CPCI-SSH Timespan=All years
 #7 TS=(outcome* Near/1 (change* or improve* or improving or severe* or severity or trajector* or worse or worst or worsen*))
 Indexes=CPCI-S, CPCI-SSH Timespan=All years
 #6 TS=(diagnos* Near/1 (final or second or later or subsequent*))
 Indexes=CPCI-S, CPCI-SSH Timespan=All years
 #5 TS=(diagnos* Near/1 (baseline or base-line or early or earlier or first or initial* or original* or previous*))
 Indexes=CPCI-S, CPCI-SSH Timespan=All years
 #4 TS=((diagnos* or temporal*) Near/3 (change* or stable or unstable or reliab* or stabili*e* or stability or instability or re-evaluat*))
 Indexes=CPCI-S, CPCI-SSH Timespan=All years
 #3 TS=((developmental or diagnos*) Near/1 (outcome* or trajector*))
 Indexes=CPCI-S, CPCI-SSH Timespan=All years
 #2 TS=(prognosis* or followup or "follow-up" or "follow* up" or predict*)
 Indexes=CPCI-S, CPCI-SSH Timespan=All years
 #1 TS=(autis* or asperger* or PDD OR "PDD-NOS" OR PERVASIVE NEAR/1 (DISORDER* OR CHILD*))
 Indexes=CPCI-S, CPCI-SSH Timespan=All years

Cochrane Database of Systematic Reviews (in the Cochrane Library)

12 October 2017 (4 records)
 6 July 2021 (3 records)

#1 MeSH descriptor: [Child Development Disorders, Pervasive] explode all trees
 #2 (autis* or asperger*):ti
 #3 (PDD N3 (specified or unspecified)):ti or "pervasive development* disorder*":ti or PDD next NOS:ti
 #4 {or #1-#3}
 #5 (prognos* or diagnos* or predict*):ti
 #6 #4 and #5, in Cochrane Reviews, Cochrane Protocols
 #7 #4 and #5 with Cochrane Library publication date Between Oct 2017 and Jul 2021, in Cochrane Reviews, Cochrane Protocols

Database of Reviews of Abstracts of Effectiveness (DARE) in the Cochrane Library

12 October 2017 (2 records)

#1 MeSH descriptor: [Child Development Disorders, Pervasive] explode all trees
 #2 (autis* or asperger*):ti
 #3 (PDD N3(specified or unspecified)) or "pervasive development* disorder*" or PDD next NOS:ti
 #4 {or #1-#3}
 #5 (prognos* or diagnos* or predict*):ti
 #6 #4 and #5

Epistemonikos (www.epistemonikos.org)

12 October 2017 (21 records)

6 July 2021. Limited to records added between 1 October 2017 and 6 July 2021 (4 records)

title:((autis* OR asperger* OR PDD* OR "pervasive disorder" OR "pervasive development") AND (predict* OR prognos* OR stabil* OR stable OR trajectory)) OR (title:((autis* OR asperger* OR PDD* OR "pervasive disorder" OR "pervasive development") AND ("diagnostic stability" OR "stable diagnosis")) OR abstract:((autis* OR asperger* OR PDD* OR "pervasive disorder" OR "pervasive development") AND ("diagnostic stability" OR "stable diagnosis"))))
 Systematic review filter applied.

SciELO (*Scientific Electronic Library Online*)

17 November 2021 (493 records)

(autis* or pervasiv* or asperger*) AND (prognos* or diagnos* or predict*)

NIHR (*National Institute for Health Research*)

17 November 2021 (30 records)

autis*; pervasive; asperger*

Clarivate Web of Science forward citations search

17 November 2021 (990 records)

We completed forward citation searches for all included studies.

Appendix 2. Data collection spreadsheet

Column heading	Definition
Study number	-
Author	First author (surname and first initial)
Country of publication	-
Year of publication	-
Description of study	Study description, prospective cohort, retrospective cohort, assessment of outcome, controlled, with/without intervention, aim of the study

(Continued)

Study population/group	Clinic versus population versus clinical drawn from a broad population base
Sampling frame	Description of where sample was collected from
Study sample	Description of baseline study sample
Inclusion/exclusion criteria	Participants that were eligible for study are described
Adequacy of participation	Adequacy of participation in the study by all who were eligible
Size of population/group	Number (N) at baseline, denominator for proportion analyses; proportion (%) male
Diagnostic criteria	DSM; ICD; or Kanner and edition number
Diagnostic tool/measure at baseline and follow-up	ADI-R; ADOS; CARS; GARS; 3di; or DISCO
Consistency of tool	Same diagnostic tool for all; same method and setting of outcome for all participants; whether valid reliable tool; completeness of outcome measure
Timing of diagnosis	Prior to study, at baseline, etc.
Multidisciplinary assessment	Diagnosis was completed by two or more professionals
Diagnosis	AD; ASD; AD + PDD-NOS; as defined by diagnostic criteria
Age at baseline in years	-
Age at follow-up in years	-
Period of follow-up in years	Length of follow-up for the study
Cognitive ability/IQ	Outcome; measure used
Language ability	Outcome; measure used
Adaptive behaviour ability	Outcome; measure used
Study approach and outcomes	When outcomes were measured
Numerator for primary outcome	Number diagnosed with ASD at follow-up
Denominator for primary outcome	Number assessed for ASD at follow-up
Proportion continuing to meet diagnostic criteria	Numerator divided by denominator
Autistic symptoms - core	Outcome: social communication/repetitive, restricted behaviours, and interests; measure used
Autistic symptoms - other	Outcome: what symptoms or measure used
Study attrition	Number of participants lost to follow-up; participants that did not complete all parts of follow-up or tools; reasons for loss to follow-up; whether reasons have been linked to outcome

(Continued)

Interventions	Type and amount of interventions
Groups	Control group versus intervention group
Notes	-

Footnotes

AD: autistic disorder; **ADI-R:** Autism Diagnostic Interview - Revised; **ADOS:** Autism Diagnostic Observation Schedule; **ASD:** autism spectrum disorder; **CARS:** Childhood Autism Rating Scale; **DISCO:** Diagnostic Interview for Social and Communication Disorders; **DSM:** Diagnostic and Statistical Manual of Mental Disorders; **GARS:** Gilliam Autism Rating Scale; **ICD:** International Classification of Diseases; **IQ:** intelligence quotient; **PDD-NOS:** pervasive developmental disorder-not otherwise specified; **3di:** developmental, dimensional and diagnostic interview.

Appendix 3. Description of risk of bias criteria and the criteria for assigning judgements

1. Study participation: the study sample adequately represents population of interest

Criteria	Unclear	High	Moderate	Low
Sample (described)	-	Clinical (not community based)	Clinical but drawn from broad community base	Population based
Description of sampling frame	-	Not described	Some description but not adequate or complete	Well described
Description of baseline study sample	-	Not described	Some description but not adequate or complete	Well described
Description of inclusion or exclusion criteria	-	Not described	Some description but not adequate or complete	Well described
Adequacy of participation in study by all eligible	-	No	-	Yes

2. Study attrition: the study data available (those not lost to follow-up) adequately represent the study sample

Criteria	Unclear	High	Moderate	Low
Recruitment	-	Retrospective	Retrospective with whole cohort considered	Prospective
LFU (%)	-	< 80% remain	≥ 80% remain	≥ 85% remain
Description of attempts to collect information on those LFU	-	No	Some information provided but not adequate	Yes
Reasons for LFU provided?	-	No	Some information provided but not adequate	Yes
Reasons for LFU linked to outcome?	-	No	Some information provided but not adequate	Yes

(Continued)

Adequate description of LFU participants?	-	No	Some information provided but not adequate	Yes
Analysis: important differences between LFU and non-LFU in study?	-	Important differences	-	No important differences

3. Outcome measurement: the outcomes of interest are measured in a similar way for all participants

Criteria	Unclear	High	Moderate	Low
Blinding	-	Not blinded	Blinding inadequate	Blinding adequate
Clear definition of outcome provided?	-	No	-	Yes
Same outcome tool for all?	-	Not same tool for all	-	Same for all
Valid and reliable tool?	-	Not valid, reliable tool used	Valid or reliable tool, but parent rating	Standardised, reliable, valid tool used
Method and setting of outcome measurement same for all participants?	-	No	-	Yes
Completeness of outcome measure	-	Not all tools completed (> 90% missing)	Not all tools completed but not > 90% missing	All tools completed

Footnotes

LFU: Loss to follow-up.

Appendix 4. GRADE assessment for judging the overall quality of the evidence for prognosis

In grading the quality of the evidence we considered observational studies starting as high quality.

	Domain	Description
Rate down if:	Risk of bias	The overall quality is driven by the study with lowest quality (if only low risk of bias studies are used, then the quality is rated as high ; individual studies are rated down one or two levels for serious or critical risk of bias.
	Inconsistency	<p>Unexplained heterogeneity or variability in results (point estimates) across studies with differences in estimates exceeding decisional thresholds.</p> <p>Large I^2 value (significant heterogeneity) and visual inspection of the forest plot (effect sizes on either side of the lines of no effect and with confidence intervals showing little to no overlap) usually prompt concerns around heterogeneity</p>

(Continued)

	Indirectness	The study sample or the outcomes in the study, or both, do not accurately reflect the population of interest or the measured outcome does not capture what is believed to be important
	Imprecision	This is based primarily the position of the confidence interval relative to a clinical decision threshold
	Publication bias	Forrest plot or statistical testing suggesting that small negative studies are underrepresented
Rate up if:	Large effect	Moderate or large effect reported by most studies or in pooled findings in the meta-analysis
	Dose-response gradient	Gradient exists between studies for factors measured at different doses or an increase or decrease in events over time, which follows a well-defined pattern (e.g. linear)

Footnotes

Table modified from [Guyatt 2011](#), [Hayden 2014](#) and [Iorio 2015](#).

Appendix 5. Levels of quality

Quality level	Definition
High	We are very confident that the true prognosis (probability of future events) lies close to that of the estimate
Moderate	We are moderately confident that the true prognosis (probability of future events) is likely to be close to the estimate, but there is a possibility that it is substantially different
Low	Our confidence in the estimate is limited: the true prognosis (probability of future events) may be substantially different from the estimate
Very low	We have very little confidence in the estimate: the true prognosis (probability of future events) is likely to be substantially different from the estimate

Footnotes

This table has been reproduced from [Iorio 2015](#), with permission from the first author.

Appendix 6. Methods for future updates and unused methods

Unit of analysis issues

In future updates of this review, we may be required to complete some data manipulation if continuous scores rather than dichotomised categories are presented for diagnostic groups.

Dealing with missing data

In future reviews, if indicated, we will assess the sensitivity of any primary analyses to missing data using the strategy described in the *Cochrane Handbook for Systematic Reviews of Interventions* ([Deeks 2022](#)). That is, we will perform sensitivity analyses to assess how sensitive the results are by excluding studies if they present data requiring transformations with uncertain assumptions or where they contain a large amount of missing data. We were unable to complete this analysis for primary analyses as the only data we analysed were where all cases were followed up. Therefore, there were no missing data in primary analyses.

Overall prognosis of preschool autism spectrum disorder diagnoses (Review)

Copyright © 2022 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Assessment of reporting biases

The very small number of studies that were rated at low risk of bias precluded sensitivity analyses for this review. If future updates have more studies with low risk of bias ratings we will conduct sensitivity analyses to assess the impact of risk bias on outcome.

Data synthesis

If, in future updates of the review included studies are found to be more homogenous than expected, we will analyse the data using a fixed-effect model.

Prognostic factor analyses

We planned to complete analyses of prognostic factors on studies that used the same or different versions or editions of the diagnostic tool at baseline and follow-up, or on autism spectrum disorder subgroups (i.e. autistic disorder versus pervasive developmental disorder - not otherwise specified). However, due to the small number of studies that had presented data for these areas we were not able to complete these analyses. We may be able to complete this analysis in future updates.

Sensitivity analysis

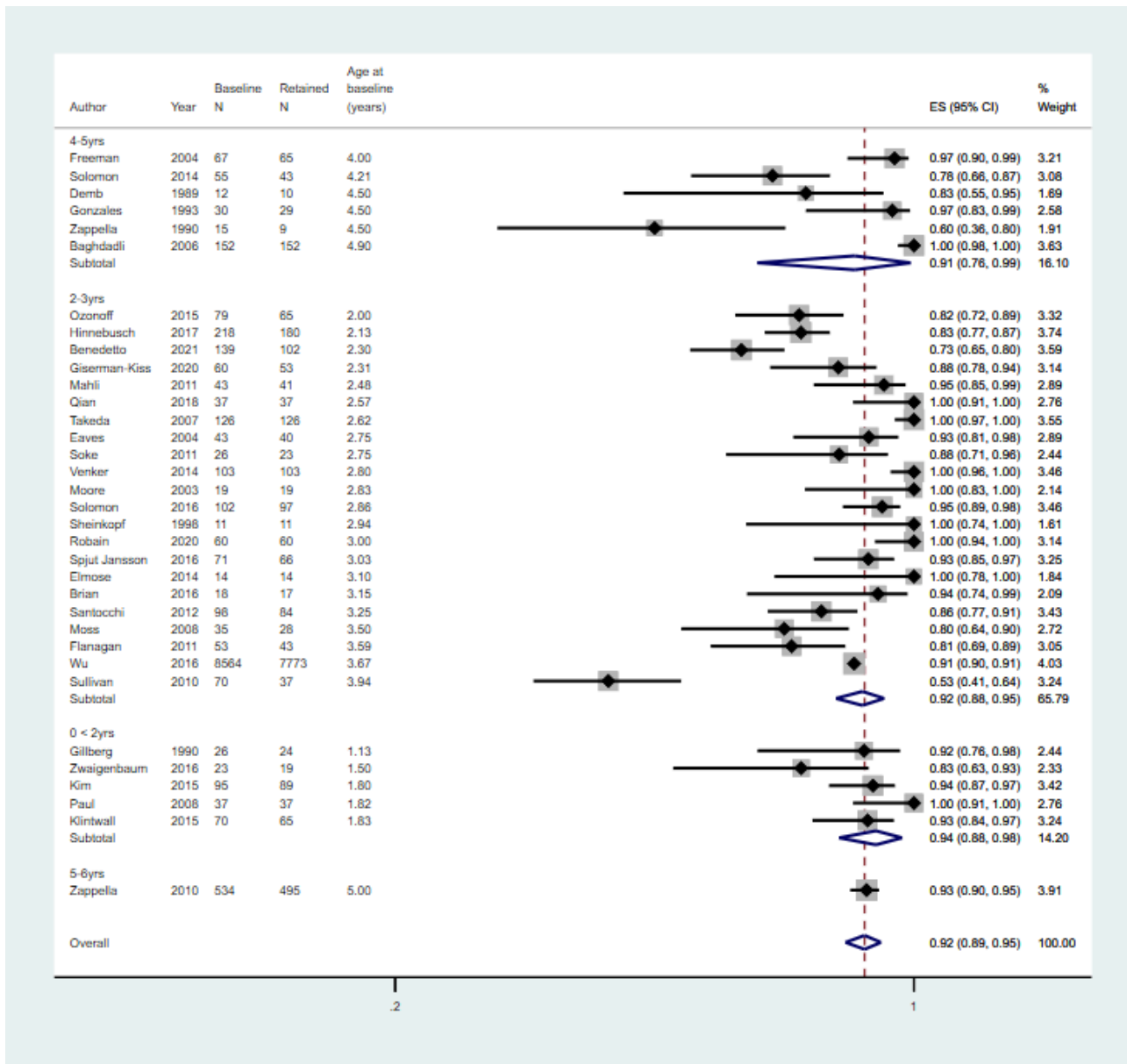
In future updates of this review, if there are additional studies with low risk of bias ratings, we will use sensitivity analyses to assess the impact of our decisions made during the review (e.g. inclusion of studies in the review and risk of bias of studies, taking into account recruitment, blinding and outcome measurement factors). This will be achieved by repeating the analyses using an alternative method or assumption, in order to explore the influence of our risk of bias assessments; for example, by the exclusion of lower-quality studies (those at high or unclear risk of bias due to study participation, participant attrition or outcome measurement).

Appendix 7. Forest plots

a) Forest plot of diagnostic stability by age at baseline

[Figure 6](#)

Figure 6. Age at baseline: < 2 years; 2 to 3 years; 4 to 6 years; 7 to 12; years; 13 to 17 years Footnote
CI: confidence interval; ES: effect size; N: number in sample

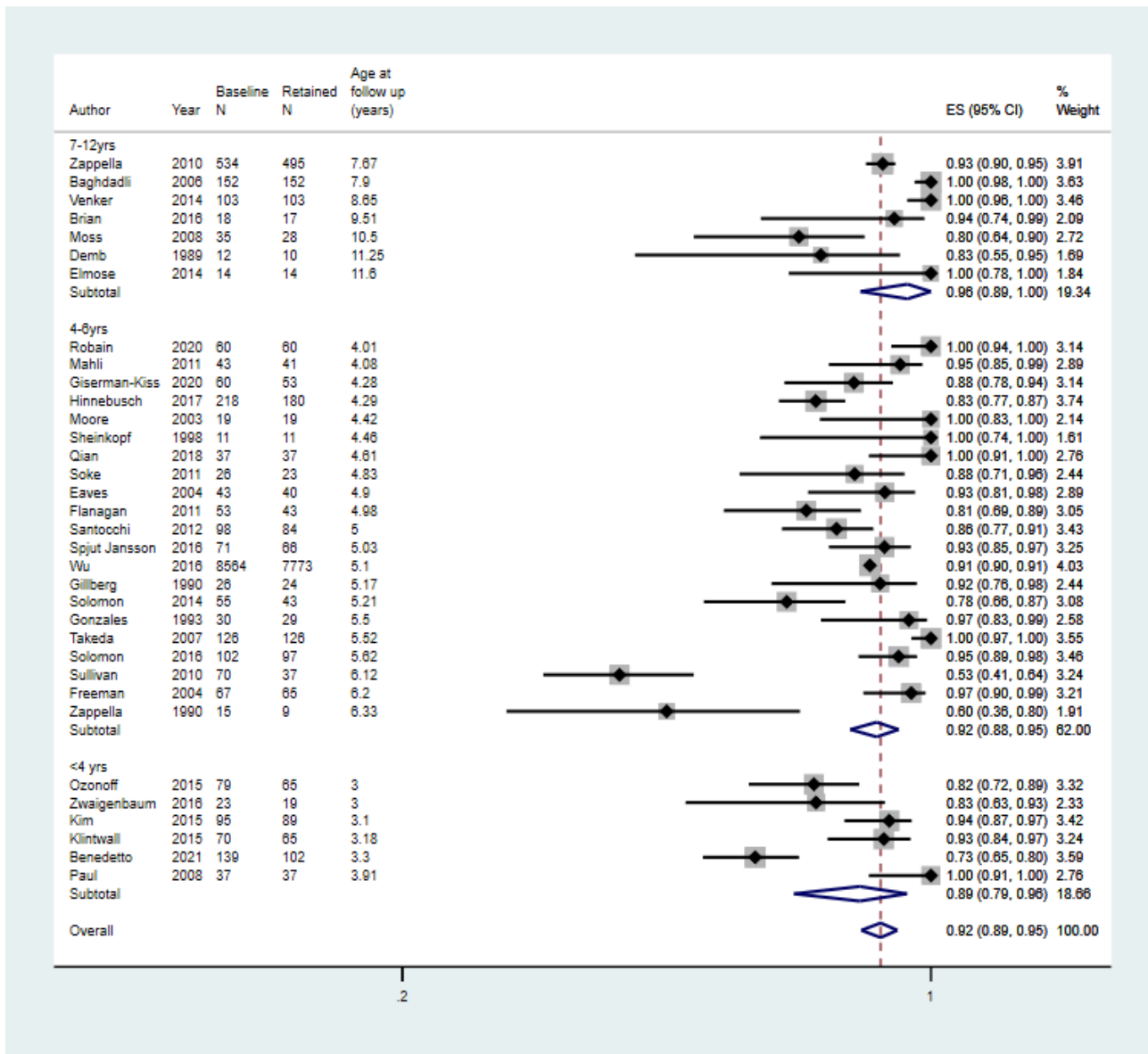


There was no significant association between the proportion of children who continued to meet diagnostic criteria for autism spectrum disorder at follow-up and age at baseline.

b) Forest plot of diagnostic stability by age at follow-up

Figure 7

Figure 7. Age at follow up: 2 to 3 years; 4 to 6 years; 7 to 12 years; 13 to 18 years *Footnote*
CI: confidence interval; ES: effect size; N: number in sample



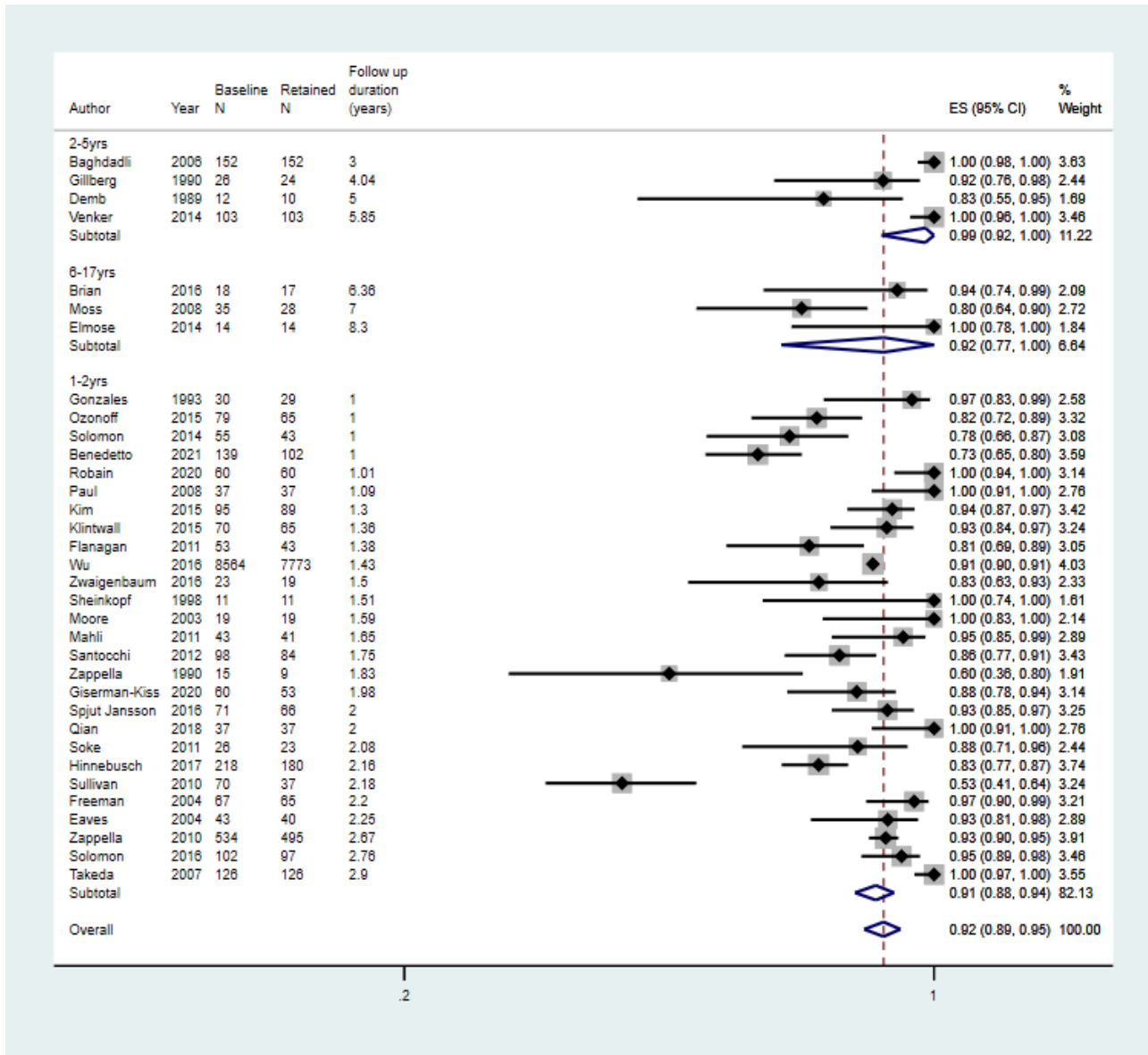
There was no significant association between the proportion of children who continued to meet diagnostic criteria for autism spectrum disorder at follow-up and age at follow-up.

c) Forest plot of diagnostic stability by duration of follow-up

Figure 8

Figure 8. Duration of follow-up: short-term (up to 2 years), medium-term (2 to 5 years), and long-term (6 to 17 years) follow-up *Footnote*

CI: confidence interval; ES: effect size; N: number in sample



There was no significant association between the proportion of children who continued to meet diagnostic criteria for autism spectrum disorder at follow-up and duration of follow-up.

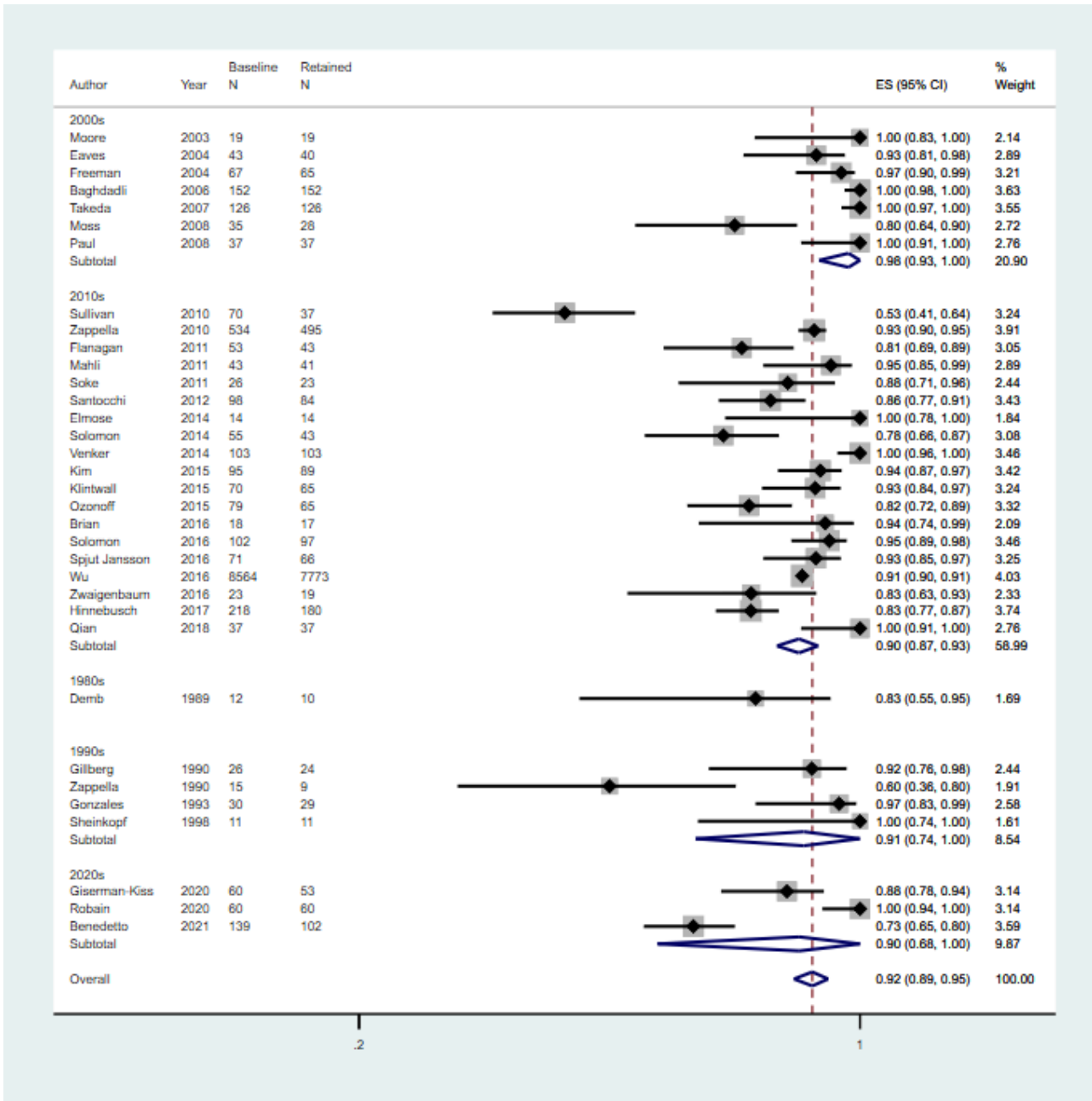
d) Forest plot of diagnostic stability by decade of publication

Figure 9

Figure 9. Decade of publication: 1960 to 1969; 1970 to 1979; 1980 to 1989; 1990 to 1999; 2000 to 2009; 2010 to 2019

Footnote

CI: confidence interval; ES: effect size; N: number in sample

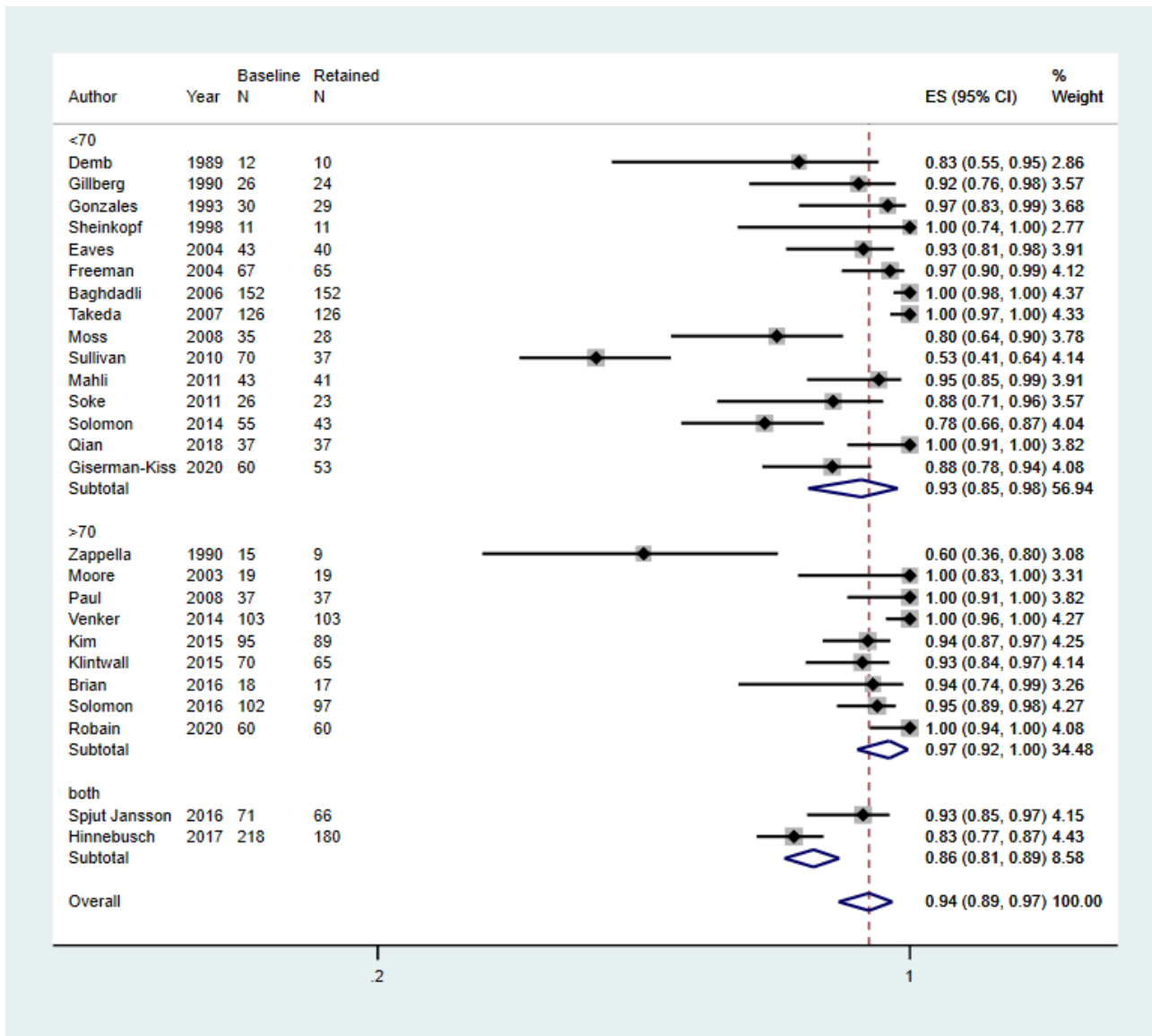


There was no significant association between the proportion of children who continued to meet diagnostic criteria for autism spectrum disorder at follow-up and decade of publication.

e) Forest plot of diagnostic stability by mean intelligence quotient

Figure 10

Figure 10. Intelligence: mean IQ 70; mean IQ > 70; or more than 70% of the cohort has IQ 70 Footnote
CI: confidence interval; ES: effect size; IQ: intelligence quotient; N: number in sample

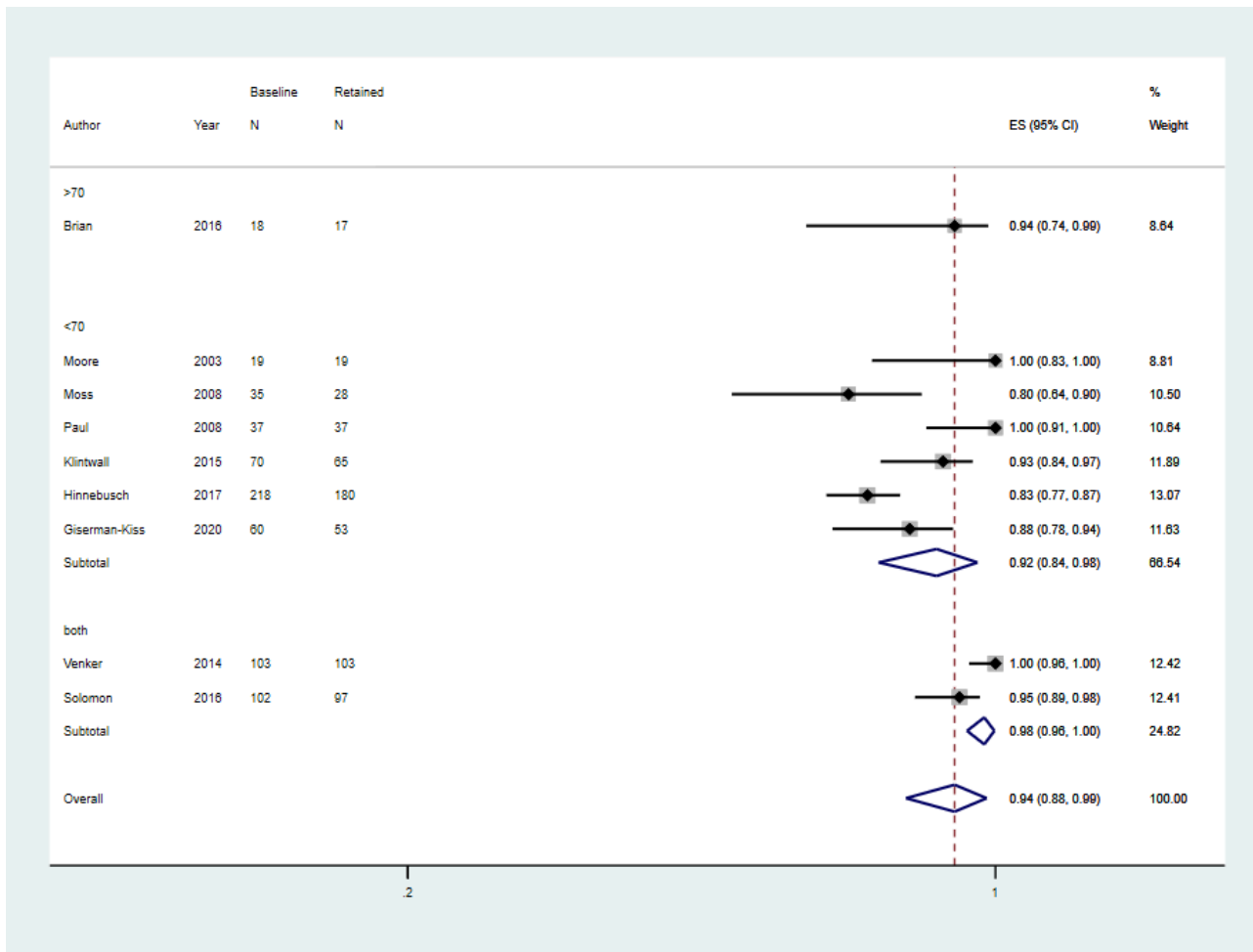


There was no significant association between the proportion of children who continued to meet diagnostic criteria for autism spectrum disorder at follow-up and mean intelligence quotient.

f) Forest plot of diagnostic stability by language ability

Figure 11

Figure 11. Language: > 70% verbal; > 70% non-verbal (i.e. use < 15 words); mean standardised language score < 70; mean standardised language score 70; or > 70% of the cohort has mean language score < 70 *Footnote*
CI: confidence interval; ES: effect size; N: number in sample



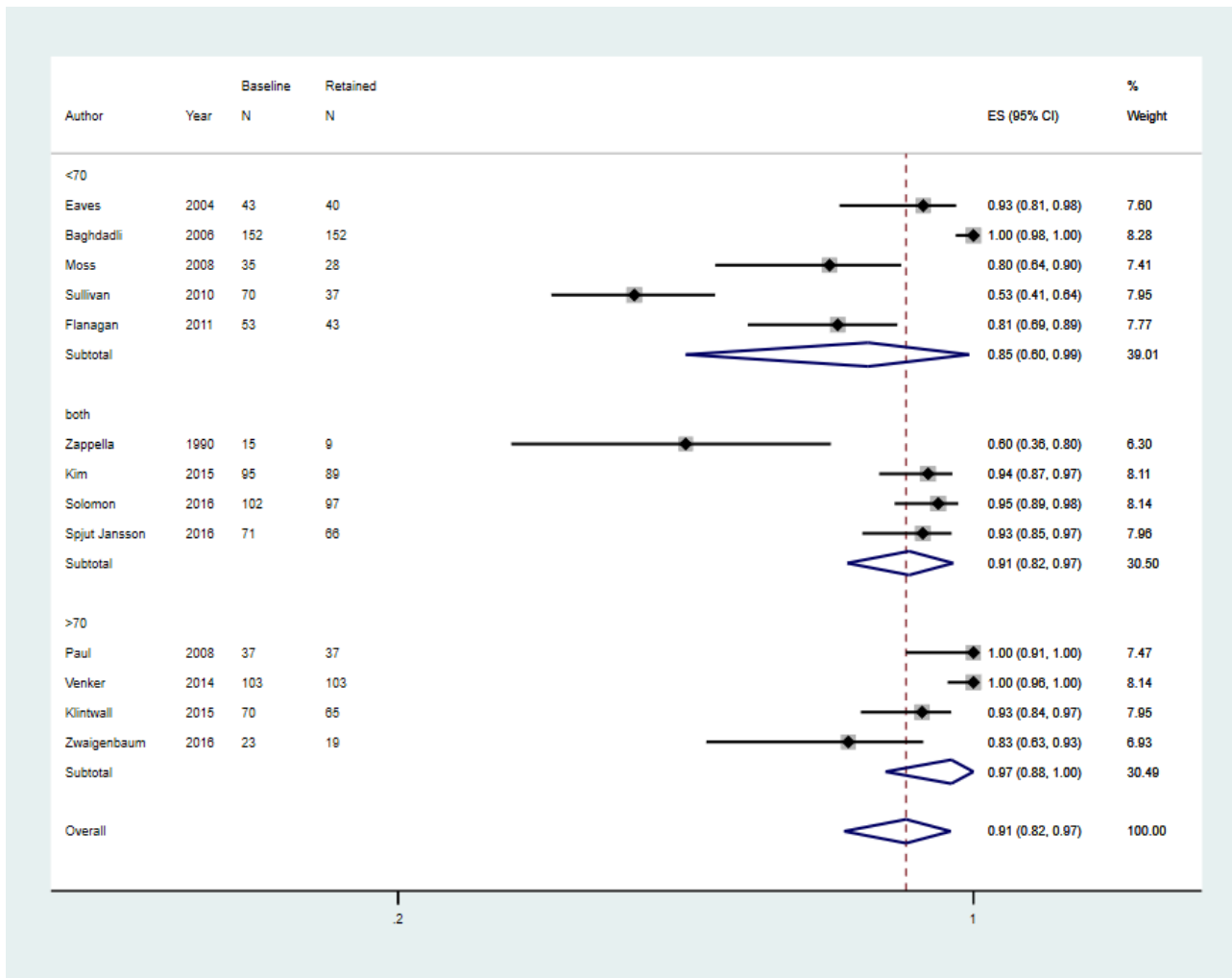
There was no significant association between the proportion of children continued to meet diagnostic criteria for autism spectrum disorder at follow-up and language ability.

g) Forest plot of diagnostic stability by adaptive behaviour

Figure 12

Figure 12. Adaptive behaviour: mean standard score 70; mean standard score > 70; or > 70% of the cohort has mean standard score 70 *Footnote*

CI: confidence interval; ES: effect size; N: number in sample

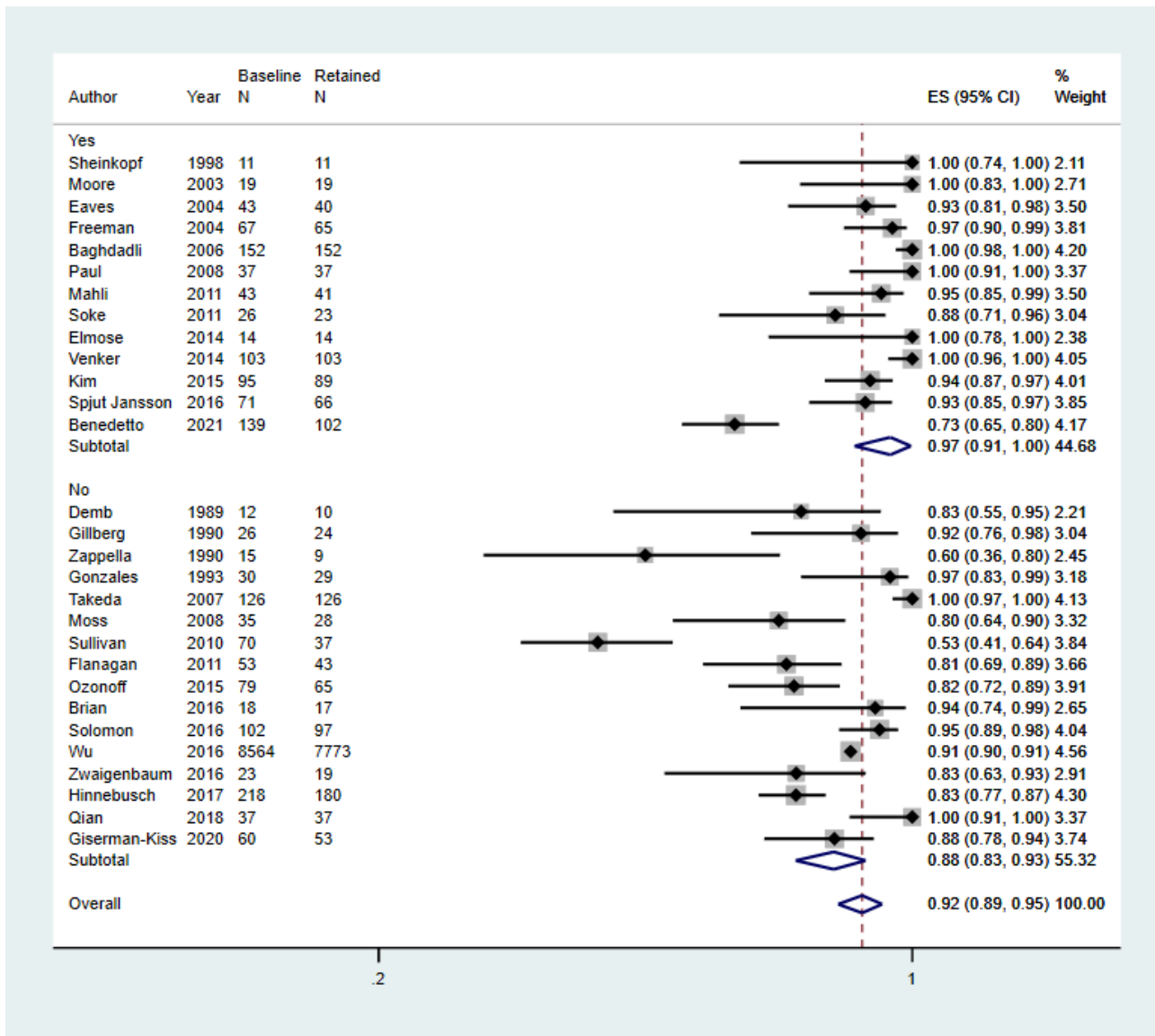


There was no significant association between the proportion of children who continued to meet diagnostic criteria for autism spectrum disorder at follow-up and adaptive behaviour ability.

h) Forest plot of diagnostic stability by multidisciplinary diagnosis

Figure 13

Figure 13. Multidisciplinary team used for diagnosis, Yes or No Footnote
CI: confidence interval; ES: effect size; N: number in sample



There was no significant association between the proportion of children who continued to meet diagnostic criteria for autism spectrum disorder at follow-up and whether the diagnosis involved a multidisciplinary team or not.

Appendix 8. Key characteristics of included studies

Variable		Not included in meta-analysis (n = 8)		Included in meta-analysis (n = 34)	
		n	%	n	%
Year published	Older (< 2013)	3	38	18	53
	Recent (2013-2021)	5	62	16	47

(Continued)

Tools used to diagnose autism spectrum disorder	One tool	8	100	21	62
	Two tools	0	0	10	29
	Three tools +	0	0	3	9
Multidisciplinary approach		0 ^a	0	13 ^b	45
Autism sub-group	Autism spectrum disorder	6	100	32	94
	Autistic disorder	0	0	2	6
	Childhood autism	0	0	0	
IQ	< 70	4 ^c	80	15 ^d	58
	> 70	0	0	9	35
	Mixed	1	20	2	7
Male		433	80	9139	82
Sample size mean (range)		67 (13-272)		329 ^e (11-8564)	
Age at baseline in years mean (range)		3.81 (2.5-4.9)		3.04 (1.13-5)	
Length of follow-up in years mean (range)		4.24 (1-7.36)		2.53 (1-8.3)	
Risk of bias (rated low)	Sample (clinical, clinical from broad base, population)	0	0	4	12
	Description of sampling frame	2	25	9	26
	Description of baseline study sample	4	50	15	44
	Description of inclusion or exclusion criteria	3	38	14	41
	Adequacy of participation in study by all eligible	3	38	15	44
	Recruitment (prospective)	7	88	22	65
	Loss to follow-up (LFU; low= >85% retained)	5	63	13	38
	Description of attempts to collect info on those LFU	0	0	3	9
	Reasons for LFU provided	0	0	3	9
	Reasons for LFU linked to outcome	1	13	1	3

(Continued)

Description of LFU participants	0	0	1	3
Analysis: important differences LFU vs non-LFU in study	2	25	6	18
Blinding	1	13	5	15
Clear definition of diagnosis	6	75	33	97
Same diagnosis outcome tool for all	8	100	34	100
Valid and reliable tool	8	100	34	100
Method and setting of outcome measurements same for all participants	4	50	21	62
Completeness of outcome measure	7	88	33	97

IQ: intelligence quotient; n: number.

Footnotes

LFU: Loss to follow-up.

 $a_n = 3$
 $b_n = 29$
 $c_n = 3$
 $d_n = 26$

 eIf we remove the outlier study with $n = 8564$, mean $n = 79$.

Appendix 9. Justifications for risk of bias assessments across 18 criteria

Study ID: Baghdadli 2012		
Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	High	Participation in study by all eligible not adequate
Recruitment	Low	Prospective
Loss to follow-up (LFU)	High	51% of sample lost to follow-up

(Continued)

Description of attempts to collect information on those LFU	Moderate	Some information provided but inadequate
Reasons for LFU provided	Low	Yes
Reasons for LFU linked to outcome	Moderate	Some information provided but inadequate
Description of LFU participants	High	Not described
Analysis: important differences LFU vs non-LFU in study	Low	No
Blinding	Moderate	Inadequately blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	High	No
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Benedetto 2021

Domain	Risk of bias level	Support for judgement
Sample (described)	Moderate	Clinical sample from a broad community base
Description of sampling frame	High	Not described
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	Low	Adequate participation by all eligible
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	13% of sample lost to follow-up
Description of attempts to collect information on those LFU	High	Not described
Reasons for LFU provided	High	No
Reasons for LFU linked to outcome	High	No

(Continued)

Description of LFU participants	High	Not described
Analysis: important differences LFU vs non-LFU in study	Unclear	Not described
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Bopp 2006

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Low	Well described
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	High	Not described
Adequacy of participation in study by all eligible	Low	Adequate participation by all eligible
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	None of sample lost to follow-up
Description of attempts to collect info on those LFU	Not applicable as no LFU	
Reasons for LFU provided	Not applicable as no LFU	
Reasons for LFU linked to outcome	Not applicable as no LFU	
Description of LFU participants	Not applicable as no LFU	
Analysis: important differences LFU vs non-LFU in study	Not applicable as no LFU	
Blinding	Unclear	Not described
Clear definition of diagnosis provided at follow-up	Low	Yes

(Continued)

Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	High	No
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Brian 2016

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Low	Well described
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	High	Not described
Adequacy of participation in study by all eligible	High	Not described
Recruitment	Low	Prospective
Loss to follow-up (LFU)	High	29% of sample lost to follow-up
Description of attempts to collect information on those LFU	High	Not described
Reasons for LFU provided	High	No
Reasons for LFU linked to outcome	High	No
Description of LFU participants	High	Not described
Analysis: important differences LFU vs non-LFU in study	Low	No
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes

(Continued)

Completeness of outcome measure	Low	Diagnostic tools completed by all study participants
---------------------------------	-----	--

Study ID: Chu 2017

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Low	Well described
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	High	Participation in study by all eligible not adequate
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	13% of sample lost to follow-up
Description of attempts to collect info on those LFU	High	Not described
Reasons for LFU provided	High	Not described
Reasons for LFU linked to outcome	High	Not described
Description of LFU participants	High	Not described
Analysis: important differences LFU vs non-LFU in study	Unclear	Not described
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Demb 1989

(Continued)

Domain	Risk of bias level	Support for judgement
Sample (described)	Moderate	Clinical sample from a broad community base
Description of sampling frame	Moderate	Some description
Description of baseline study sample	High	Not described
Description of inclusion or exclusion criteria	High	Not described
Adequacy of participation in study by all eligible	High	Participation in study by all eligible not adequate
Recruitment	Low	Prospective
Loss to follow-up (LFU)	High	33.3% of sample lost to follow-up
Description of attempts to collect information on those LFU	Moderate	Some information but information inadequate
Reasons for LFU provided	Moderate	Some information but information inadequate
Reasons for LFU linked to outcome	High	No
Description of LFU participants	High	No
Analysis: important differences LFU vs non-LFU in study	Unclear	Not described
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	High	No, 11 were done in person and one via phone
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: DeWaay 2012

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Moderate	Some description

(Continued)

Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	Moderate	Some description
Adequacy of participation in study by all eligible	Low	Adequate participation by all eligible
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	None of sample lost to follow-up but participants selected retrospectively from a prospective cohort
Description of attempts to collect info on those LFU	Unclear	Not described
Reasons for LFU provided	Unclear	Not described
Reasons for LFU linked to outcome	Unclear	Not described
Description of LFU participants	Unclear	Not described
Analysis: important differences LFU vs non-LFU in study	Unclear	Not described
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	High	No
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	High	No
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Eaves 2004

Domain	Risk of bias level	Support for judgement
Sample (described)	Low	Population-based sample
Description of sampling frame	Low	Well described
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	Moderate	Some description
Adequacy of participation in study by all eligible	Low	Adequate participation by all eligible
Recruitment	Low	Prospective

(Continued)

Loss to follow-up (LFU)	Low	None of sample lost to follow-up
Description of attempts to collect information on those LFU	Not applicable as no LFU	
Reasons for LFU provided	Not applicable as no LFU	
Reasons for LFU linked to outcome	Not applicable as no LFU	
Description of LFU participants	Not applicable as no LFU	
Analysis: important differences LFU vs non-LFU in study	Not applicable as no LFU	
Blinding	Unclear	Not discussed
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Elmore 2014

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	High	Not described
Adequacy of participation in study by all eligible	Low	Adequate participation by all eligible
Recruitment	Moderate	Retrospective, with the whole cohort considered
Loss to follow-up (LFU)	Not applicable as the study is retrospective	
Description of attempts to collect information on those LFU	Low	Yes
Reasons for LFU provided	Moderate	Some information but information is inadequate
Reasons for LFU linked to outcome	High	Not described

(Continued)

Description of LFU participants	High	Not described
Analysis: important differences LFU vs non-LFU in study	High	Not described
Blinding	Low	Blinding adequate
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Flanagan 2011

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	High	Participation in study by all eligible not adequate
Recruitment	High	Retrospective
Loss to follow-up (LFU)	Not applicable as the study is retrospective	
Description of attempts to collect information on those LFU	Not applicable as the study is retrospective	
Reasons for LFU provided	Not applicable as the study is retrospective	
Reasons for LFU linked to outcome	Not applicable as the study is retrospective	
Description of LFU participants	Not applicable as the study is retrospective	
Analysis: important differences LFU vs non-LFU in study	Not applicable as the study is retrospective	
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes

(Continued)

Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Freeman 2004

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Low	Well described
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	High	Not described
Adequacy of participation in study by all eligible	Low	Adequate participation by all eligible
Recruitment	High	Retrospective
Loss to follow-up (LFU)	Not applicable as the study is retrospective	
Description of attempts to collect information on those LFU	Not applicable as the study is retrospective	
Reasons for LFU provided	Not applicable as the study is retrospective	
Reasons for LFU linked to outcome	Not applicable as the study is retrospective	
Description of LFU participants	Not applicable as the study is retrospective	
Analysis: important differences LFU vs non-LFU in study	Not applicable as the study is retrospective	
Blinding	Moderate	Blinding inadequate
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Unclear	Not described
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Gillberg 1990

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	High	Not described
Adequacy of participation in study by all eligible	Low	Adequate participation by all eligible
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	None of sample lost to follow-up
Description of attempts to collect information on those LFU	Not applicable as no LFU	
Reasons for LFU provided	Not applicable as no LFU	
Reasons for LFU linked to outcome	Not applicable as no LFU	
Description of LFU participants	Not applicable as no LFU	
Analysis: important differences LFU vs non-LFU in study	Not applicable as no LFU	
Blinding	High	Not
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Moderate	Valid or reliable but parent-rated tool
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Giserman-Kiss

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Low	Well described

(Continued)

Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	Moderate	Some description
Adequacy of participation in study by all eligible	Low	Adequate participation in study by all eligible
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	None of sample lost to follow-up
Description of attempts to collect information on those LFU	Not applicable as no LFU	Not described
Reasons for LFU provided	Not applicable as no LFU	Not described
Reasons for LFU linked to outcome	Not applicable as no LFU	Not described
Description of LFU participants	Not applicable as no LFU	Not described
Analysis: important differences LFU vs non-LFU in study	Not applicable as no LFU	Not described
Blinding	Unclear	Not described
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Unclear	Not described
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Gonzalez 1993

Domain	Risk of bias level	Support for judgement
Sample (described)	Low	Population-based sample
Description of sampling frame	Low	Well described
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	High	Not described

(Continued)

Adequacy of participation in study by all eligible	High	Participation in study by all eligible not adequate
Recruitment	Moderate	Retrospective with the whole cohort considered
Loss to follow-up (LFU)	Not applicable as the study is retrospective	
Description of attempts to collect information on those LFU	Not applicable as the study is retrospective	
Reasons for LFU provided	Not applicable as the study is retrospective	
Reasons for LFU linked to outcome	Not applicable as the study is retrospective	
Description of LFU participants	Not applicable as the study is retrospective	
Analysis: important differences LFU vs non-LFU in study	Not applicable as the study is retrospective	
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Haglund 2020

Domain	Risk of bias level	Support for judgement
Sample (described)	Moderate	Clinical sample from a broad community base
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	Moderate	Some description
Adequacy of participation in study by all eligible	High	Participation in study by all eligible not adequate
Recruitment	Low	Prospective
Loss to follow-up (LFU)	High	2% of sample lost to follow-up

(Continued)

Description of attempts to collect information on those LFU	High	Not described
Reasons for LFU provided	High	No
Reasons for LFU linked to outcome	High	No
Description of LFU participants	High	Not described
Analysis: important differences LFU vs non-LFU in study	Unclear	Not described
Blinding	Low	Blinding adequate
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Hinnebusch 2017

Domain	Risk of bias level	Support for judgement
Sample (described)	Low	Population-based sample
Description of sampling frame	Low	Well described
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	High	Participation in study by all eligible not adequate
Recruitment	Low	Prospective
Loss to follow-up (LFU)	High	44% of sample lost to follow-up
Description of attempts to collect information on those LFU	Low	Yes
Reasons for LFU provided	High	No
Reasons for LFU linked to outcome	Unclear	Not described
Description of LFU participants	Low	Yes

(Continued)

Analysis: important differences LFU vs non-LFU in study	Low	Some differences but these would not impact outcome
Blinding	Unclear	Not described
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	High	More than 90% of diagnostic tools not completed

Study ID: Kim 2015

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Low	Not described
Description of inclusion or exclusion criteria	Low	Not described
Adequacy of participation in study by all eligible	Low	Adequate participation by all eligible
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	None of sample lost to follow-up
Description of attempts to collect information on those LFU	Not applicable as no LFU	
Reasons for LFU provided	Not applicable as no LFU	
Reasons for LFU linked to outcome	Not applicable as no LFU	
Description of LFU participants	Not applicable as no LFU	
Analysis: important differences LFU vs non-LFU in study	Not applicable as no LFU	
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes

(Continued)

Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Klintwall 2015

Domain	Risk of bias level	Support for judgement
Sample (described)	Unclear	Not described
Description of sampling frame	Moderate	Some description
Description of baseline study sample	High	Not described
Description of inclusion or exclusion criteria	Moderate	Some
Adequacy of participation in study by all eligible	High	Participation in study by all eligible not adequate
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	None of sample lost to follow-up
Description of attempts to collect information on those LFU	Not applicable as no LFU	
Reasons for LFU provided	Not applicable as no LFU	
Reasons for LFU linked to outcome	Not applicable as no LFU	
Description of LFU participants	Not applicable as no LFU	
Analysis: important differences LFU vs non-LFU in study	Not applicable as no LFU	
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Mahli 2011

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	Unclear	Not described
Recruitment	Low	Prospective
Loss to follow-up (LFU)	High	54.2% of sample lost to follow-up
Description of attempts to collect information on those LFU	High	Not described
Reasons for LFU provided	High	No
Reasons for LFU linked to outcome	Unclear	Not described
Description of LFU participants	Moderate	Some information but inadequate
Analysis: important differences LFU vs non-LFU in study	Low	Yes
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Moore 2003

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Moderate	Some description

(Continued)

Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	High	Not described
Adequacy of participation in study by all eligible	Unclear	Not described
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	% of sample lost to follow-up
Description of attempts to collect information on those LFU	Not applicable as no LFU	
Reasons for LFU provided	Not applicable as no LFU	
Reasons for LFU linked to outcome	Not applicable as no LFU	
Description of LFU participants	Not applicable as no LFU	
Analysis: important differences LFU vs non-LFU in study	Not applicable as no LFU	
Blinding	Low	Blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Moss 2008

Domain	Risk of bias level	Support for judgement
Sample (described)	Moderate	Clinical sample from a broad community base
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	Low	Participation in study by all eligible was adequate
Recruitment	Low	Prospective

(Continued)

Loss to follow-up (LFU)	Moderate	20% of sample lost to follow-up
Description of attempts to collect information on those LFU	Low	Well described
Reasons for LFU provided	Low	Yes
Reasons for LFU linked to outcome	Low	Yes
Description of LFU participants	High	Not described
Analysis: important differences LFU vs non-LFU in study	Low	Yes
Blinding	Moderate	Blinding inadequate
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Ozonoff 2015

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	Low	Participation in study by all eligible not adequate
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	None of sample lost to follow-up
Description of attempts to collect information on those LFU	High	Not described
Reasons for LFU provided	High	No
Reasons for LFU linked to outcome	High	No
Description of LFU participants	High	Not described

(Continued)

Analysis: important differences LFU vs non-LFU in study	High	Not described
Blinding	Low	Blinding adequate
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	High	No
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Paul 2008

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	Unclear	Not described
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	None of sample lost to follow-up
Description of attempts to collect information on those LFU	Not applicable as no LFU	
Reasons for LFU provided	Not applicable as no LFU	
Reasons for LFU linked to outcome	Not applicable as no LFU	
Description of LFU participants	Not applicable as no LFU	
Analysis: important differences LFU vs non-LFU in study	Not applicable as no LFU	
Blinding	Moderate	Blinding inadequate
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes

(Continued)

Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Qian 2018

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	High	Not described
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	Moderate	Some described
Adequacy of participation in study by all eligible	High	Participation in study by all eligible not adequate
Recruitment	High	Retrospective
Loss to follow-up (LFU)	Not applicable as the study is retrospective	
Description of attempts to collect information on those LFU	Not applicable as the study is retrospective	
Reasons for LFU provided	Not applicable as the study is retrospective	
Reasons for LFU linked to outcome	Not applicable as the study is retrospective	
Description of LFU participants	Not applicable as the study is retrospective	
Analysis: important differences LFU vs non-LFU in study	Not applicable as the study is retrospective	
Blinding	Unclear	Not described
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Unclear	Not described
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Rivard 2019

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	High	Not described
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	Moderate	Some description
Adequacy of participation in study by all eligible	High	Inadequate participation in study by all eligible
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	None of sample lost to follow-up but selected participants retrospectively from a prospective cohort
Description of attempts to collect information on those LFU	Unclear	Not described
Reasons for LFU provided	Unclear	Not described
Reasons for LFU linked to outcome	Unclear	Not described
Description of LFU participants	Unclear	Not described
Analysis: important differences LFU vs non-LFU in study	Unclear	Not described
Blinding	High	Not described
Clear definition of diagnosis provided at follow-up	High	No
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	High	No
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Robain 2020

Domain	Risk of bias level	Support for judgement
--------	--------------------	-----------------------

(Continued)

Sample (described)	Moderate	Clinical sample from a broad community base
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	Moderate	Some description
Adequacy of participation in study by all eligible	Unclear	Not described
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	None of sample lost to follow-up
Description of attempts to collect information on those LFU	Not applicable as no LFU	
Reasons for LFU provided	Not applicable as no LFU	
Reasons for LFU linked to outcome	Not applicable as no LFU	
Description of LFU participants	Not applicable as no LFU	
Analysis: important differences LFU vs non-LFU in study	Not applicable as no LFU	
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	High	No
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Sheinkopf 1998

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	High	Not described
Description of baseline study sample	High	Not described
Description of inclusion or exclusion criteria	Moderate	Some description

(Continued)

Adequacy of participation in study by all eligible	Unclear	Not described
Recruitment	Moderate	Retrospective, with the whole cohort considered
Loss to follow-up (LFU)	Not applicable as the study is retrospective	
Description of attempts to collect information on those LFU	Not applicable as the study is retrospective	
Reasons for LFU provided	Not applicable as the study is retrospective	
Reasons for LFU linked to outcome	Not applicable as the study is retrospective	
Description of LFU participants	Not applicable as the study is retrospective	
Analysis: important differences LFU vs non-LFU in study	Not applicable as the study is retrospective	
Blinding	Unclear	Not described
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Unclear	Not described
Completeness of outcome measure	Moderate	Not all tools completed but not > 90% missing

Study ID: Smith 2019

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	High	Participation in study by all eligible not adequate
Recruitment	Low	Prospective
Loss to follow-up (LFU)	High	48% of sample lost to follow-up
Description of attempts to collect information on those LFU	High	Not described

(Continued)

Reasons for LFU provided	Low	Yes
Reasons for LFU linked to outcome	Moderate	Some information provided but inadequate
Description of LFU participants	Moderate	Some information provided but inadequate
Analysis: important differences LFU vs non-LFU in study	Low	No important differences reported
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	High	No
Completeness of outcome measure	Moderate	Not all tools completed but not >90% missing

Study ID: Soke 2011

Domain	Risk of bias level	Support for judgement
Sample (described)	Moderate	Clinical sample from a broad community base
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Low	Not described
Description of inclusion or exclusion criteria	Low	Not described
Adequacy of participation in study by all eligible	Low	Participation in study by all eligible not adequate
Recruitment	Low	Prospective
Loss to follow-up (LFU)	High	22% of sample lost to follow-up
Description of attempts to collect information on those LFU	High	Not described
Reasons for LFU provided	High	No
Reasons for LFU linked to outcome	High	No
Description of LFU participants	High	Not described

(Continued)

Analysis: important differences LFU vs non-LFU in study	Unclear	Not described
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Unclear	Not described
Completeness of outcome measure	Low	All tools complete

Study ID: Solomon 2014

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Low	Well described
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	High	Participation in study by all eligible not adequate
Recruitment	Low	Prospective
Loss to follow-up (LFU)	High	25% of sample lost to follow-up
Description of attempts to collect information on those LFU	High	Not described
Reasons for LFU provided	High	No
Reasons for LFU linked to outcome	High	No
Description of LFU participants	High	Not described
Analysis: important differences LFU vs non-LFU in study	Unclear	Not described
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes

(Continued)

Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Unclear	Not described
Completeness of outcome measure	Low	All tools complete

Study ID: Solomon 2016

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	High	Participation in study by all eligible not adequate
Recruitment	Low	Prospective
Loss to follow-up (LFU)	High	47% of sample lost to follow-up
Description of attempts to collect information on those LFU	High	Not described
Reasons for LFU provided	Low	Yes
Reasons for LFU linked to outcome	Moderate	Some information but inadequate
Description of LFU participants	Moderate	Some information but inadequate
Analysis: important differences LFU vs non-LFU in study	Low	No
Blinding	Low	Blinding adequate
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	High	No
Completeness of outcome measure	Low	All tools complete

Study ID: Spjut Jansson 2016

Domain	Risk of bias level	Support for judgement
Sample (described)	Low	Population based sample
Description of sampling frame	Low	Well described
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	Moderate	Some description
Adequacy of participation in study by all eligible	High	Participation in study by all eligible not adequate
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	None of sample lost to follow-up
Description of attempts to collect information on those LFU	Not applicable as no LFU	
Reasons for LFU provided	Not applicable as no LFU	
Reasons for LFU linked to outcome	Not applicable as no LFU	
Description of LFU participants	Not applicable as no LFU	
Analysis: important differences LFU vs non-LFU in study	Not applicable as no LFU	
Blinding	Moderate	Blinding inadequate
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	High	No
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	All tools complete

Study ID: Sullivan 2010

Domain	Risk of bias level	Support for judgement
Sample (described)	Moderate	Clinical sample from a broad community base
Description of sampling frame	High	Not described

(Continued)

Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	Moderate	Some description
Adequacy of participation in study by all eligible	Unclear	Not described
Recruitment	Moderate	Retrospective with the whole cohort considered
Loss to follow-up (LFU)	Not applicable as the study is retrospective	
Description of attempts to collect information on those LFU	Not applicable as the study is retrospective	
Reasons for LFU provided	Not applicable as the study is retrospective	
Reasons for LFU linked to outcome	Not applicable as the study is retrospective	
Description of LFU participants	Not applicable as the study is retrospective	
Analysis: important differences LFU vs non-LFU in study	Not applicable as the study is retrospective	
Blinding	Unclear	Not described
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Unclear	Not described
Completeness of outcome measure	Low	All tools complete

Study ID: Szatmari 2021

Domain	Risk of bias level	Support for judgement
Sample (described)	Moderate	Clinical sample from a brand community base
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	High	Not described
Recruitment	Low	Prospective
Loss to follow-up (LFU)	High	28% of sample lost to follow-up

(Continued)

Description of attempts to collect information on those LFU	High	Not described
Reasons for LFU provided	High	No
Reasons for LFU linked to outcome	Low	Yes
Description of LFU participants	High	Not described
Analysis: important differences LFU vs non-LFU in study	High	Yes, important differences reported
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	High	No
Completeness of outcome measure	Low	Diagnostic tools completed by all study participants

Study ID: Takeda 2007

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	High	Not described
Adequacy of participation in study by all eligible	Low	Adequate participation in study by all eligible
Recruitment	High	Retrospective
Loss to follow-up (LFU)	Not applicable as the study is retrospective	
Description of attempts to collect information on those LFU	Not applicable as the study is retrospective	
Reasons for LFU provided	Not applicable as the study is retrospective	
Reasons for LFU linked to outcome	Not applicable as the study is retrospective	
Description of LFU participants	Not applicable as the study is retrospective	

(Continued)

Analysis: important differences LFU vs non-LFU in study	Not applicable as the study is retrospective	
Blinding	High	Blinding inadequate
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	All tools complete

Study ID: Thomas 2009

Domain	Risk of bias level	Support for judgement
Sample (described)	Unclear	Not described
Description of sampling frame	High	Not described
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	High	Not described
Adequacy of participation in study by all eligible	Unclear	Not described
Recruitment	High	Retrospective
Loss to follow-up (LFU)	Not applicable as the study is retrospective	
Description of attempts to collect info on those LFU	Not applicable as the study is retrospective	
Reasons for LFU provided	Not applicable as the study is retrospective	
Reasons for LFU linked to outcome	Not applicable as the study is retrospective	
Description of LFU participants	Not applicable as the study is retrospective	
Analysis: important differences LFU vs non-LFU in study	Not applicable as the study is retrospective	
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used

(Continued)

Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	All tools complete

Study ID: Venker 2014

Domain	Risk of bias level	Support for judgement
Sample (described)	Moderate	Clinical sample from a broad community base
Description of sampling frame	Low	Well described
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	High	Not described
Adequacy of participation in study by all eligible	Low	Adequate participation in study by all eligible
Recruitment	Low	Prospective
Loss to follow-up (LFU)	High	20.2% of sample lost to follow-up
Description of attempts to collect information on those LFU	High	Not described
Reasons for LFU provided	High	No
Reasons for LFU linked to outcome	High	No
Description of LFU participants	High	Not described
Analysis: important differences LFU vs non-LFU in study	High	Yes
Blinding	Unclear	Not described
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Unclear	Not described
Completeness of outcome measure	Low	All tools complete

Study ID: Wu 2016

Domain	Risk of bias level	Support for judgement
Sample (described)	Moderate	Clinical sample from a broad community base
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	Unclear	Not described
Recruitment	High	Retrospective
Loss to follow-up (LFU)	Not applicable as the study is retrospective	
Description of attempts to collect information on those LFU	Not applicable as the study is retrospective	
Reasons for LFU provided	Not applicable as the study is retrospective	
Reasons for LFU linked to outcome	Not applicable as the study is retrospective	
Description of LFU participants	Not applicable as the study is retrospective	
Analysis: important differences LFU vs non-LFU in study	Not applicable as the study is retrospective	
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Moderate	Valid or reliable but parent rated
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	All tools complete

Study ID: Zappella 1990

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	High	Not described
Description of baseline study sample	Moderate	Some description

(Continued)

Description of inclusion or exclusion criteria	Moderate	Some description
Adequacy of participation in study by all eligible	Low	Adequate participation in study by all eligible
Recruitment	High	Retrospective
Loss to follow-up (LFU)	Not applicable as the study is retrospective	
Description of attempts to collect information on those LFU	Not applicable as the study is retrospective	
Reasons for LFU provided	Not applicable as the study is retrospective	
Reasons for LFU linked to outcome	Not applicable as the study is retrospective	
Description of LFU participants	Not applicable as the study is retrospective	
Analysis: important differences LFU vs non-LFU in study	Not applicable as the study is retrospective	
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Unclear	Not described
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	All tools complete

Study ID: Zappella 2010

Domain	Risk of bias level	Support for judgement
Sample (described)	High	Clinical sample
Description of sampling frame	High	Not described
Description of baseline study sample	Moderate	Some description
Description of inclusion or exclusion criteria	Moderate	Some description
Adequacy of participation in study by all eligible	Low	Adequate participation in study by all eligible
Recruitment	High	Retrospective
Loss to follow-up (LFU)	Not applicable as the study is retrospective	
Description of attempts to collect information on those LFU	Not applicable as the study is retrospective	

(Continued)

Reasons for LFU provided	Not applicable as the study is retrospective	
Reasons for LFU linked to outcome	Not applicable as the study is retrospective	
Description of LFU participants	Not applicable as the study is retrospective	
Analysis: important differences LFU vs non-LFU in study	Not applicable as the study is retrospective	
Blinding	High	Not blinded
Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Unclear	Not described
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	All tools complete

Study ID: Zwaigenbaum 2016

Domain	Risk of bias level	Support for judgement
Sample (described)	Moderate	Clinical sample from a broad community base
Description of sampling frame	Moderate	Some description
Description of baseline study sample	Low	Well described
Description of inclusion or exclusion criteria	Low	Well described
Adequacy of participation in study by all eligible	Low	Adequate participation in study by all eligible
Recruitment	Low	Prospective
Loss to follow-up (LFU)	Low	None of sample lost to follow-up
Description of attempts to collect information on those LFU	Not applicable as no LFU	
Reasons for LFU provided	Not applicable as no LFU	
Reasons for LFU linked to outcome	Not applicable as no LFU	
Description of LFU participants	Not applicable as no LFU	
Analysis: important differences LFU vs non-LFU in study	Not applicable as no LFU	
Blinding	Low	Adequately blinded

(Continued)

Clear definition of diagnosis provided at follow-up	Low	Yes
Same diagnosis outcome tool for all	Low	Yes
Valid and reliable tool	Low	Standardised, reliable valid tool used
Method and setting of outcome measurements same for all participants	Low	Yes
Completeness of outcome measure	Low	All tools complete

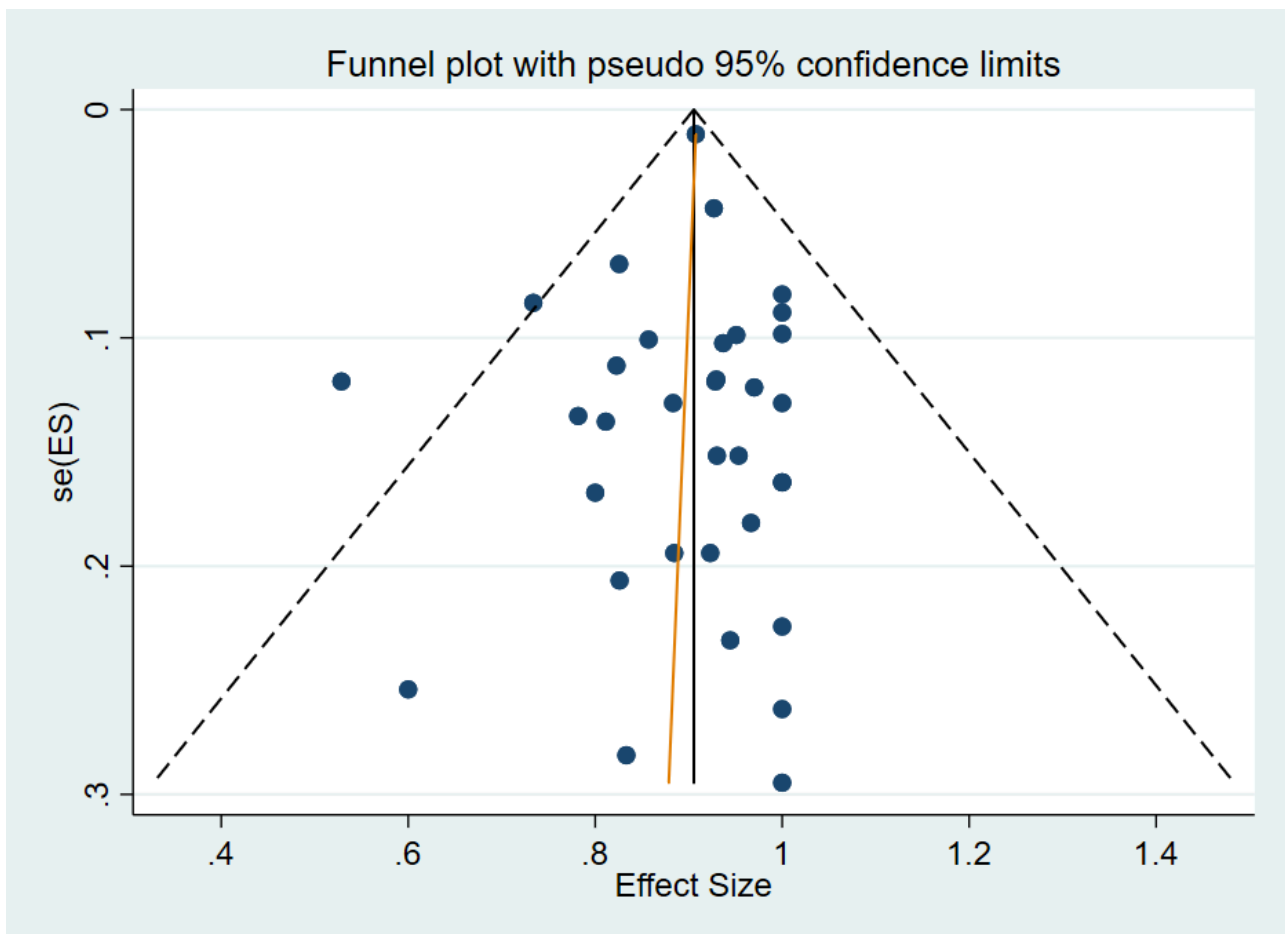
Footnotes

LFU: loss to follow up; vs: versus

Appendix 10. Funnel Plot

Figure 14

Figure 14. Funnel Plot of included studies



Funnel Plot of included studies

HISTORY

Protocol first published: Issue 8, 2017

Date	Event	Description
18 July 2021	Amended	Updating author's (AI) declaration of interest. See Declarations of interest .

CONTRIBUTIONS OF AUTHORS

KW, SW and AB conceived the review.

KW, AB, SW designed the review.

AB and RH co-ordinated the review.

AB, NA-U, SW, AI, KW developed the protocol.

AB, RH, SW and KW screened search results at title, abstract level

AB, RH, SW, KW, and AM screened search results at full text level.

AB, RH, RB (acknowledgement), AM, SW and TM extracted data and rated risk of bias. KW resolved any conflicts.

AB, TM and RH analysed the data.

TM conducted the meta-analyses, sensitivity and prognostic factor analyses.

AB, RH TM, KW and AI interpreted the data.

AB and TM graded the quality of the evidence using GRADE. KW and AI were consulted for their expertise when completing the GRADE assessments.

AB and RH wrote the review report, with KW and TM providing methodological, clinical and general advice.

All authors reviewed the final version.

AB is the guarantor for the review.

DECLARATIONS OF INTEREST

AB works as a Speech Pathologist at Monash Children's Hospital, Clayton, Australia. She is also an Associate Editor for Cochrane Developmental, Psychosocial and Learning Problems (DPLP). AB was not involved in the editorial process for this review.

RH: is employed by Western Health as a Paediatric Registrar and works at multiple teaching hospitals in Victoria Australia, caring for children who have been diagnosed with autism spectrum disorder, including conducting diagnostic assessments for autism spectrum disorder as part of a multidisciplinary team.

TM works as a Psychologist in private practice in Victoria Australia.

AM is a Developmental and Community Paediatrician for the Sydney Local Health District and Developmental Paediatrics, Bondi, Sydney, Australia.

SW is the Director of Paediatrics at Sydney Local Health District, Sydney, Australia.

AI is a Hematology Consultant with Hamilton Health Sciences and Professor and Chair of the Department of Research Methods, Evidence, and Impact, McMaster University, Canada. AI is an Editor for Cochrane Cystic Fibrosis and Genetic Disorders and Cochrane Prognosis Methods.

KW is a Developmental Paediatrician for Monash Children's Hospital, Clayton, Australia, an Editor for Cochrane DPLP and a Convenor for Cochrane Child and Cochrane Prognosis. She was not involved in the editorial process for this review. KW declares a grant from the National Health and Medical Research Council for an autism prognosis study about predictors of autism outcome that will also publish diagnostic

stability outcomes, and could be included in an update of this systematic review; paid to La Trobe University. The funder had no role in the design of this study or the methods, and will have no role in the data analysis and reporting.

SOURCES OF SUPPORT

Internal sources

- Monash University, Australia

Salary support provided for AB, TM and KW

- Murdoch Children's Research Institute (MCRI), Australia

MCRI provided some salary support for AB

- Sydney Children's Hospital Network, Australia

Salary support provided for SW

- National Health and Medical Research Council (NHMRC), Australia

Susan Woolfenden is supported by an NHMRC Career Development Grant (GNT1158954)

- University of New South Wales, Australia

Salary support provided for SW

External sources

- None, Australia

We do not have any external sources of support to declare for this review.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Criteria for considering studies for this review

Types of outcome measures

We added an inclusion criterion to clarify the same diagnostic methods and tools needed to be used at both baseline and follow-up. This was to ensure minimal impact of type of tool on diagnostic status.

We removed the wording "For studies that presented data using continuous measures (e.g. a score on a diagnostic scale), we analysed the data by computing a dichotomous variable" because a mean score on a tool did not allow us to determine change in diagnostic status. The primary outcome was proportion diagnosed with autism spectrum disorder, so it was not appropriate to use continuous measures for this outcome.

Data collection and analysis

We trialled the planned data collection forms and adapted these to ensure the most relevant information was collected (see [Appendix 2](#) for adapted data collection form).

We added additional text on the criteria used to assign ratings and summaries. We added the following to the methods so the process for judging risk of bias was fully transparent.

"These items were then summarised into three domains (study participation, study attrition and outcome measurement) by combining the individual item ratings to provide a risk of bias rating for each summary domain. For study participation, we prioritised ratings for the 'participation in the study by all eligible' and 'study recruitment' criteria. Poor participation and retrospective studies were marked at high risk of bias. Those studies with good participation and prospective recruitment were marked at low risk of bias, which was then graded down to moderate risk of bias if they had one other high risk of bias criterion or three moderate risk of bias criteria across the remainder of the study participation domain criteria. In the study attrition domain, 'loss to follow-up' criteria were prioritised for determining the domain rating. Those with no loss to follow-up received a rating of low risk of bias for the study attrition domain rating, except in the case of retrospective studies, where the loss to follow-up is determined by the selection of participants retrospectively, based on data availability. For the outcome measurement domain rating, we prioritised blinding of the study. If the study was unblinded, it was given a rating of high risk of bias for the domain rating. If the blinding was unclear and the remainder of the criteria were at low risk of bias, then the study was rated at moderate risk of bias; however, if blinding was unclear and there was at least one other criteria in the domain rated at moderate or high, then that study was rated at high risk of bias for the outcome measure domain. Lastly, we provided one overall risk of bias rating for each study. We rated studies to have an overall low risk of bias if all three summary domains were rated at low or moderate risk of bias. Those rated to be at overall high risk of bias were those where one or more summary domains were rated at high risk of bias."

We changed the wording of the term 'subgroup' (written in the protocol) to 'prognostic factor' throughout the manuscript. Prognostic factors modify overall prognosis rather than subgroups, so this term is preferred.

Unit of analysis issues

We stated in the protocol that "if studies had reported data for subgroups (e.g. autistic or autism spectrum disorder; male or female) we will calculate a composite mean score, if this is meaningful. We will do this by conducting a fixed-effect meta-analysis of within-study groups, following the methods described by [Borenstein 2009](#)". Continuous mean scores provided for a whole group do not allow us to categorise the number of participants within the group who did and did not meet autism criteria on that measure using any type of analysis including a fixed-effects meta-analysis. Upon further consideration, we have removed the sentence on conducting a fixed-effect meta-analysis from the review, as we do not plan to use these methods in the future.

Dealing with missing data

We added the following text to the methods section of the review and followed these methods for the review.

"We only included studies when baseline and follow-up data were provided, detailing the number of children diagnosed with autism spectrum disorder, and where the method of diagnosis was explicitly provided. For studies where we could not extract data on the primary outcome, we compared the characteristics of studies included and excluded from the meta-analysis and reported any differences in study samples."

Reporting bias

It was not appropriate to complete funnel plots due to the nature of the data (ceiling effect of 100% made the analyses invalid). Therefore, we removed the following sections from the methods.

"If we are able to pool 10 or more studies, we will examine publication bias and other small study effects, using a funnel plot in Review Manager (RevMan), version 5 (Review Manager 2014). We will check for asymmetry at a 10% level. We will attempt to obtain the results of unpublished studies by contacting study authors. Where this is not possible, and the missing studies are thought to introduce significant bias, we will explore the impact of including such studies in the overall assessment of results using sensitivity analyses."

Data synthesis

We used Stata ([StataCorp 2019](#)) rather than Revman Web ([RevMan Web 2020](#)), to construct forest plots because Revman does not enable forest plots of proportion data.

We changed the minimum number of studies required for meta-analysis from three to two since a meta-analysis is possible with two studies. We wrote the following: "We conducted meta-analyses since data were available from two or more sufficiently homogeneous studies".

We modified the wording of the GRADE assessment table ([Appendix 4](#)).

We decided not to collect information in a 'GRADE Evidence Profile' table ([Schünemann 2013](#)). We removed the 'GRADE Evidence Profile' table that had been included in the study protocol ([Brignell 2017](#)), as it was not required.

INDEX TERMS

Medical Subject Headings (MeSH)

*Autism Spectrum Disorder [diagnosis]; Prognosis; Prospective Studies; Retrospective Studies; Schools

MeSH check words

Adult; Child; Child, Preschool; Female; Humans; Infant; Male; Young Adult