Autism & Developmental Language Impairments

*Review article*

Autism & Developmental Language Impairments
Volume 3: 1–19

DOI: 10.1177/2396941518767610
journals.sagepub.com/home/dli

SAGE

# A systematic review and meta-analysis of the prognosis of language outcomes for individuals with autism spectrum disorder

**Amanda Brignell**
Speech and Language, Murdoch Children's Research Institute, Parkville, Australia

**Angela T Morgan**
Speech and Language, Murdoch Children's Research Institute, Parkville, Australia;
Department of Paediatrics and Department of Audiology and Speech Pathology, University of Melbourne, Parkville, Australia

**Susan Woolfenden**
Community Child Health, Sydney Children's Hospital Network, Sydney Children's Hospital, Randwick, Australia;
University of New South Wales, Randwick, Australia

**Felicity Klopper**
School of Psychology, Deakin University, Geelong, Australia

**Tamara May**
Developmental Disability and Rehabilitation Research, Murdoch Children's Research Institute, Parkville, Australia;
School of Psychology, Deakin University, Geelong, Australia

**Vanessa Sarkozy**
Tumbatin, Sydney Children's Hospital Network, Randwick, Australia;
University of New South Wales, Randwick, Australia

**Katrina Williams**
Developmental Disability and Rehabilitation Research, Murdoch Children's Research Institute, Parkville, Australia;
Department of Paediatrics, University of Melbourne, Parkville, Australia; Neurodevelopment and Disability, Royal Children's Hospital, Parkville, Australia

## Abstract

**Background:** Language difficulties are common in autism spectrum disorder, yet little is known about the prognosis of language in children with autism spectrum disorder. The aim of this study was to systematically review studies reporting language outcomes in individuals with autism spectrum disorder.

**Method:** A comprehensive search strategy with a well-established sensitive prognosis filter for Medline, adapted for five other databases, was used. Included studies observed individuals diagnosed with autism spectrum disorder for ≥12 months and had ≥30 participants. Risk of bias was assessed.

**Results:** Fifty-four studies (N = 5064) met inclusion criteria. Language outcomes were standardised assessments (n = 35), notation of presence/absence of verbal language (n = 11) or both (n = 8). Age at baseline ranged from 17 months to 26 years, duration of follow-up from 1 to 38 years. Most publications (92%) were rated medium to high risk of bias. In all but one study individuals had below-average scores at baseline and follow-up. However, in most (n = 24/25; 96%) studies reporting standard scores, individuals (aged ≤ 11 years at follow-up) progressed at a comparable rate to

**Corresponding author:**
Amanda Brignell, Speech and Language, Murdoch Children's Research Institute, Level 5, Royal Children's Hospital, Parkville, Melbourne 3052, Australia.
Email: Amanda.brignell@mcri.edu.au

age-expected norms or demonstrated some 'catch up' over time. Meta-analyses found mean standard scores increased over time in three language domains (composite receptive language, composite expressive language and adaptive language). Nineteen to thirty percent of children aged five years and under gained verbal language. For children aged over five years 5–32% gained verbal language over the course of study. Age, baseline language scores, IQ and length of follow-up did not moderate between study differences in composite language or adaptive language growth or the acquisition of verbal language.

**Conclusion:** Despite variability in study methods, findings were consistent, with the majority of studies reporting children under 11 years on average progressed at a comparable rate to age-expected norms or with some 'catchup' over time.

**Implications:** This review provides synthesised information for families and clinicians on language development over time and on language outcomes for individuals with autism spectrum disorder. Such information can be useful for prognostic counselling and may assist planning around future resources and support needs. This review also makes recommendations regarding methodology for future studies so that prognosis can become more fine-tuned at an individual level.

**Keywords**

Autism spectrum disorder, systematic review, language, speech, prognosis, outcomes, longitudinal, follow-up

## Introduction

To meet criteria for a diagnosis of autism spectrum disorder (ASD), the current Diagnostic Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013) requires the presence of social communication difficulties and repetitive and restricted interests and behaviours. While receptive and expressive language difficulties (i.e. difficulties with semantics, syntax and morphology) are not requisite for an ASD diagnosis, they are a common comorbidity and DSM-5 now requires specification about whether such a language impairment is present.

Language development is often the first issue raised by parents of children later diagnosed with autism (De Giacomo & Fombonne, 1998; Herlihy, Knoch, Vibert, & Fein, 2015) and around 63% of children diagnosed with ASD present with co-occurring language disorders (Levy et al., 2010). Pragmatic language impairments are considered universal to ASD (Tager-Flusberg & Joseph, 2003); however, there is substantial heterogeneity in the extent of structural language difficulties that co-occur with ASD (Ellis Weismer & Kover, 2015; Tager-Flusberg, Paul, & Lord, 2005; Tek, Mesite, Fein, & Naigles, 2014). Some individuals have well preserved (or even superior) structural language abilities on formal testing, with sophisticated vocabulary and sentence structure (Boucher, 2012; Tager-Flusberg & Caronna, 2007), yet studies of individuals ascertained through population-based or clinical samples have found between 25% and 30% of children with ASD are minimally verbal or nonverbal (Anderson et al., 2007; Norrelgen et al., 2014). These estimates may vary according to age, the definition of minimally verbal or nonverbal used and the way the sample was ascertained.

A range of language domains may be variably impacted in individuals with ASD including phonology, semantics, morphology and syntax (Gernsbacher, Morson, & Grace, 2016; Tager-Flusberg, 2006). There has also been debate around whether there is a subgroup of individuals with ASD who have language phenotypes that resemble children with specific language impairment (Rapin, Dunn, Allen, Stevens, & Fein, 2009; Tager-Flusberg, 2015) and whether these two conditions have shared or separate aetiological pathways (Kjelgaard & Tager-Flusberg, 2001; Whitehouse, Barry, & Bishop, 2008). Further to the variability noted in language skills in this group, there is variation in the types of difficulties seen. Stereotypical verbal behaviours are commonly described including repetitive language, idiosyncratic phrases, difficulties with pronouns and echolalia (American Psychiatric Association, 2013; Tager-Flusberg & Caronna, 2007).

Children with ASD with higher verbal ability are reported to have similar rates of development to children who are typically developing on a range of structural language measures (Tek, Mesite, Fein, & Naigles, 2013). However, children with ASD with lower verbal ability typically make slower progress and have flatter language trajectories (Tek, Mesite, Fein, & Naigles, 2013). One study of language growth in children with ASD found several factors (i.e. cognitive ability, maternal education and response to joint attention) accurately classified over 80% of children into higher or lower language outcome groups (Ellis Weismer & Kover, 2015). Specifically, ASD severity was a significant predictor of receptive and expressive language growth during the preschool period, while nonverbal intelligence quotient (IQ) predicted growth in

expressive language only (Ellis Weismer & Kover, 2015). Another study found the age at which early language milestones were acquired for children with ASD and nonverbal IQ $\geq$ 70 was predictive of later structural language outcomes (based on a sentence repetition task) and adaptive communication skills, but not social communication or other areas of adaptive behaviour (Kenworthy et al., 2012).

There is consensus that verbal skills play a critical role in predicting long-term outcomes for children with ASD in areas such as adaptive functioning, psychosocial adjustment and wellbeing (Gillespie-Lynch et al., 2012; Hofvander et al., 2009; Howlin, 2003; Howlin & Moss, 2012; Mawhood & Howlin, 2000; Szatmari, Bryson, Boyle, Streiner, & Duku, 2003). Language skills are also critical to school placements and academic performance as well as the ability to participate in successful social interactions (Thurm, Lord, Lee, & Newschaffer, 2007). Despite recognition of the importance of this domain to children with ASD, language outcomes remain poorly understood.

Parents of children with ASD commonly ask health professionals whether their child will talk, and if so, to what extent and how their language development will progress compared to their peers. A substantial number of studies have investigated language outcomes in children with ASD. Due to significant variation in methodologies across studies, however (e.g. different length of follow-up, ages and developmental profiles of the children, language domains measured and the tools used to measure language), it is difficult to interpret study findings in a clinically meaningful way. High-quality systematic review evidence is required about the likely language outcomes for children so clinicians can provide well-informed answers.

Two prior reviews of language outcomes have been published. One systematic review reported on adult language outcomes (Magiati, Tay, & Howlin, 2014). This review found a large amount of individual variability regarding change in language over time. Half the included studies reported improvements in raw, age-equivalent or standard scores from childhood to adulthood. One study in this review reported only 12% of individuals achieved normal or near fluency. IQ during childhood and early language ability appeared to be the strongest predictors of adult outcome in this review. One other summative review reported on the outcomes of nonverbal individuals with ASD (Pickett, Pullara, O'Grady, & Gordon, 2009). This review identified 64 published materials that had reported a total of 167 children with ASD who had developed speech at or after the age of five. In studies that reported clear ASD diagnoses and age of speech acquisition (n = 36 studies), most children (n = 44/78; 56%) who acquired speech after four years of age did so at five or six years; however, a small number of children (n = 8/78; 10%) acquired speech at 11–13 years. In this review, no cases were reported of children acquiring speech after 13 years of age.

To our knowledge, this is the first study to systematically review outcomes for a range of language domains in individuals with ASD across the lifespan and to assess risk of bias of included studies. The aim of this review was to provide a narrative description of language outcomes and where studies report similar language outcomes, to synthesise current evidence for language outcomes in children with ASD using meta-analysis. We also aimed to explore potential moderators of language outcomes.

## Method

Criteria for including studies in this review are specified in the following section.

### *Participants*

Children and adults of all ages with a diagnosis of ASD, autism, autistic disorder, childhood autism, pervasive developmental disorder (PDD), PDD-not otherwise specified (PDD-NOS), atypical autism, PDD-unspecified or Asperger's disorder/syndrome were included in the review.

Studies were included only if the diagnosis was made at the beginning of the study using a standardised diagnostic instrument including the Autism Diagnostic Interview-Revised (ADI-R; Le Couteur, Lord, & Rutter, 2003), Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000), Diagnostic Interview for Social and Communication Disorders (Wing, Leekam, Libby, Gould, & Larcombe, 2002), Childhood Autism Rating Scale (CARS; Schopler, Reichler, DeVellis, & Daly, 1980), Gillian Autism Rating Scale (Gilliam, 1995) and/or by using established diagnostic criteria of an accepted classification system at the time, such as DSM III–IV–IV, IV-TR, 5 (American Psychiatric Association, 1980, 2000, 2013) or International Statistical Classification of Diseases and Related Health Problems (ICD 9–10) (World Health Organization, 2010). A dual diagnosis (e.g. Asperger disorder and attention deficit hyperactivity disorder, or autism and Fragile X) did not prevent inclusion.

### *Types of studies*

Intervention and observation studies were eligible for inclusion if initially defined participants (diagnosed with ASD) were followed up for a period of 12 months or more. Retrospective and prospective cohorts were included. Studies had at least 30 participants to differentiate case series from a cohort

study. Randomised control trials presenting data separately for the comparison/control and intervention groups were only included if the comparison/control arm had more than 30 participants. Studies published in languages other than English were also included.

### Types of outcome measures

Studies were included if language outcomes were measured by standardised assessments or where the study reported on the presence/absence of verbal language (e.g. no words, use of single words, phrases). Studies were also required to have a baseline and follow-up measure of language. Standardised parent report and direct assessment of language tools were included, along with broader tools such as the Vineland Adaptive Behaviour Scales (VABS) (Sparrow, Cicchetti, & Balla, 2005), if expressive and/or receptive communication subdomains were assessed. The term 'language outcomes' is used to refer to all included tools. Studies of nonverbal language (e.g. augmentative/alternative communication such as use of signs or symbols to communicate) were excluded. We did not include studies that had used non-standardised measures or coded language measures that were only specific to one study because the measures were so diverse and the aim of this review was to compare findings across studies.

### Search strategy for identification of studies

Databases were searched using the search filter 'prognosis sensitive' devised for the Medline database by Wilczynski and Haynes (2004). The filter was adapted for other databases that did not systematically offer this same filter. PsycINFO, Embase, CINAHL, Cochrane Database of Systematic Reviews and the Database of Reviews of Effectiveness were used. Conference proceedings and thesis dissertation abstracts were searched. The reference lists of included articles were reviewed and experts in the field were contacted. Online Appendix A lists database specific search terms.

### Review of studies

Titles and abstracts of all references identified were screened by at least two authors assessing every title and abstract. Studies failing to meet inclusion criteria were excluded. The full text of potentially relevant articles was obtained and again assessed by at least two authors. Disagreement between the two authors was resolved by consensus or referred to a third author for arbitration. Articles not fulfilling inclusion criteria were excluded.

### Quality assessment

Risk of bias was assessed using a modified Quality in Prognosis Studies tool (Hayden, van der Windt, Cartwright, Côté, & Bombardier, 2013). The modification was required because the current study did not analyse confounders or prognostic factors and therefore these assessment categories did not apply. Three authors (AB and either AM or TM) assessed risk of bias in all included studies and any differences in ratings were resolved through consensus. Studies were assessed using the domains: study participation (e.g. adequate description of inclusion/exclusion criteria, adequate participation in study by all eligible), study attrition (e.g. loss to follow-up, retrospective or prospective study) and outcome measurement (e.g. blinding, use of a valid, reliable tool). Online Appendix B lists subcategories of each domain and specific criteria used to rate each item as unclear, low, medium or high. If information required for assessment was not available in studies published after 2000, authors were emailed for further information. This was not conducted for studies published before 2000 due to expected increased difficulty contacting authors of studies published more than 17 years ago.

### Data management

Data extraction was independently completed via a standardised template, by a minimum of two authors (AB and either FK or TM). Important clinical information likely to influence applicability and interpretation of findings (and necessary to allow assessment of homogeneity of studies) was extracted (Online Appendix C). Information was collected on study setting (e.g. population-based or clinical sample), number of participants, study population, diagnostic tool, IQ, age, follow-up period, language tools used and proportion lost to follow-up.

Tools used for measuring outcomes were categorised by five language domains: receptive and expressive vocabulary, composite receptive and expressive language (i.e. the tool measures multiple areas of language such as morphology, syntax and semantics), parent-rated adaptive language, parent-rated vocabulary and proportion verbal/nonverbal or using phrases. For the purposes of this review, if individuals used 'no or few words consistently on a daily basis' or were described as 'minimally verbal', we grouped them as 'nonverbal'. We acknowledge, as have other authors (Bal, Katz, Bishop, & Krasileva, 2016; Kasari, Brady, Lord, & Tager-Flusberg, 2013; Norrelgen et al., 2014; Rose, Trembath, Keen, & Paynter, 2016; Tager-Flusberg & Kasari, 2013; Tager-Flusberg et al., 2016), there is a difference between an individual being totally nonverbal compared to using some words; however, such level

of detail was not often provided by studies and it was beyond the scope of this review to subdivide groups further.

Studies were also grouped based on developmental or cognitive level measured by either a standardised IQ or developmental quotient (DQ). These were identified by whether the mean IQ or DQ for the cohort was $\leq 70$ or $>70$, or if $>70\%$, or $<30\%$, of the sample scored IQ/DQ $\leq 70$. Where a mental age was given for a developmental tool we converted this to a DQ by dividing mental age by chronological age. If only nonverbal subtests (e.g. visual perception) of cognitive assessments were reported, we used those to estimate DQ or IQ. For tools such as the Mullen Scales of Early Learning (MSEL; Mullen, 1995), a standard T score (e.g. for the nonverbal subtests) of 30 which is 2 standard deviations below the mean score of 50 was considered the 'cut point' or equivalent to $<70$ on an IQ test.

If a median/mean was not provided for duration of follow-up, baseline and/or follow-up age, a mean duration or age was imputed by taking the average of the lower and upper end points of the range given. Where available, we collected information on any intervention. Where participants had been described as receiving a range of interventions in the community, they were grouped as 'treatment in the community', otherwise the treatment is described as a specific intervention. Descriptive information was also collected on whether each study analysed predictors of language outcomes. However, analysis of predictors was not performed because the methods of this systematic review were not developed for that purpose. In cases where the same participants were included in more than one publication (also with use of the same outcome measures) data were taken from the publication with the largest sample size and/or where data on language outcomes could most easily be extracted. Publications using the same participants were provided with an overall study identifier label.

### *Statistical analysis*

Outcomes were presented for relevant studies in graphical format where possible. Studies were graphed when data describing mean/median standard scores or age-equivalent scores could be extracted or when the proportion of children who were verbal or used phrases at two time points was reported. This allowed comparison of trajectories from baseline to follow-up. Standard scores allow comparison to a 'typically developing' reference score. For most tools this is a score of 100. If development progresses at a rate expected for an individual's age, standard scores should remain roughly the same over time (e.g. 100 at Time 1 and 100 at Time 2). When we investigated whether there was a significant increase or decrease in mean scores over time we used confidence intervals (CIs). If CIs overlapped between Time 1 and Time 2 we considered this to be within the normal range of fluctuation. Non-overlapping CIs indicated significant improvement or decline in language relative to reference norms.

Age-equivalent scores allow comparison of an individual's language age (age-equivalent) to their chronological age. In typical development one would expect chronological age to roughly match language age at any time point. If available, or able to be calculated, CIs were added to scores. A number of studies reported raw scores. It was not possible to track change relative to age-expected levels for raw scores, instead they were interpreted with reference to direction of change (gain, plateau or loss of language skills) over time. Studies that did not report scores on the same tool at two time points were described in the text.

Due to variability in study findings, we were not able to present summary statistics for all language measures; however, if five or more studies reported on the same language outcome at two time points, a meta-analysis was conducted followed by meta-regression co-varying for age at baseline, language level at baseline, length of follow-up and IQ. A random effects meta-analysis was conducted using the DerSimonian and Laird method. This method takes the heterogeneity into account in the final calculation of the effect size and the CI around that effect size. When proportions were used in meta-analyses, Stata's Metaprop statistical program for binomial data was used (Nyaga, Arbyn, & Aerts, 2014). Meta-regressions were conducted to assess whether specific co-variates (e.g. age at baseline, IQ, length of follow up) explained the heterogeneity in language outcomes. We were not able to develop a valid scale for ASD severity that could be applied across studies (studies used a variety of tools to report ASD severity (ADOS, ADI-R, CARS, social responsiveness scale)). Therefore, ASD severity was not included in the meta-regression model.

## Results

### *Search results*

The search was conducted and updated, adapting for any differences between databases. The final search was completed at the end of Week 2 in January 2016. Figure 1 shows a flow diagram of the literature search, the number of studies that were excluded or met inclusion criteria.

The combined search yielded 19,410 studies. Review of 319 full text articles identified 92 publications that met inclusion criteria and measured language as an outcome (see Online Appendix C for the full list of

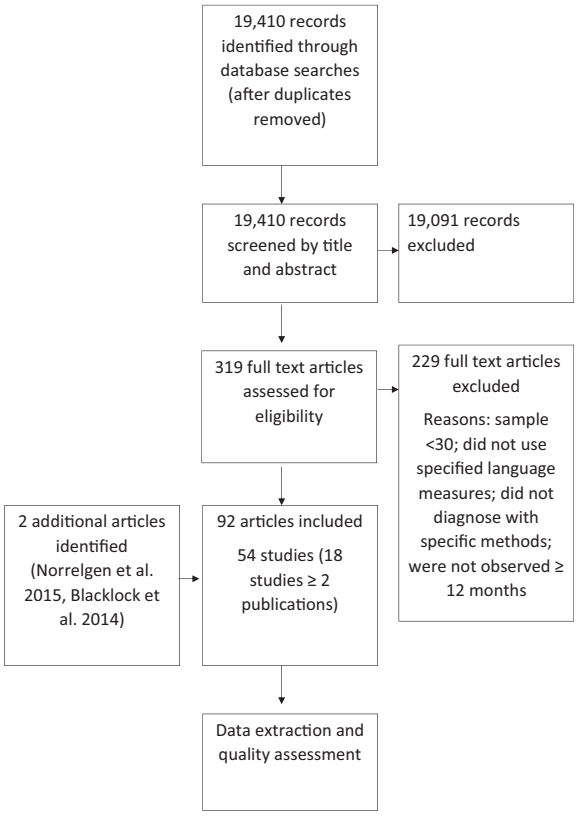


**Figure 1.** Results of the literature search and study selection.

publications and their characteristics). Eighteen studies had two or more publications and overlapping participants. A total of 54 studies (N = 5064) met inclusion criteria once duplicate publications were removed. Duplicate studies were grouped as one study and labelled with a study identifier (Online Appendix D). The reference list for included studies is provided in Online Appendix F.

## *Risk of bias assessment*

Risk of bias ratings are included in Table 1. We assessed all included publications (n = 92) of the included studies (n = 54) for risk of bias because factors linked to risk of bias varied for publications based on the same participants. For example, follow-up at one time point could be much higher than at another. Only one publication was at low risk of bias for study participation, study attrition and outcome measure risk of bias categories. A further eight were at low risk of bias for two of these three categories (Table 1). Thirty-three (36%) were at a high risk of bias for two or more categories. Sixty-seven publications (73%) collected information prospectively and the remainder were retrospective. Only nine publications derived their sample from a population source and the rest were derived from clinical samples. In all publications, individuals were diagnosed using DSM or ICD criteria or accepted standardised diagnostic instruments. Nine publications reported blinding the outcome assessor to the child's baseline status. The percentage of participants followed up ranged from 24% to 100% with 46 of the 67 prospective publications (70%) reporting less than 20% of participants lost to follow-up.

## *Language outcome measures*

Of the 54 included studies, standardised parent-completed tools were reported in 29 studies, standardised clinician-completed tools in 30 and the presence/absence of verbal language in 19 studies. Twenty-one of the aforementioned studies reported language outcomes using two or more measures at each time point. Language tools were grouped into five broad domains (Table 2).

## *Characteristics of studies*

The number of participants in a study ranged from 32 to 1433 with 40 studies (74%) having more than 50 participants. Study duration ranged from 1 to 38 years. Age at baseline ranged from 17 months to 26 years and at follow-up 35 months to 59 years. Nine studies (16%) followed children into adulthood. Eleven studies included children classified with autistic disorder or autism only, two did not specify the type of diagnosis and the remaining 40 (74%) included children from the autism spectrum. In 25 studies (46%), the majority of children had a cognitive impairment (as defined by 70% or more children having an IQ or DQ < 70 or mean IQ or DQ of 70 or less). Fifteen studies (28%) included children with mean IQ/DQ over 70 or > 70% of the sample had an IQ > 70 and 14 studies did not present data on IQ or it was not possible to extract the required information. Four studies (7%) were population-based. Ten studies (19%) involved administration of a specific intervention where outcomes had been followed up over time. The remaining studies (81%) were observational and included children who had received a broad range of interventions in the community (Online Appendix C).

## *Language outcome measures in each study*

The 54 studies used 14 different tools to assess language (Online Appendix D). The measures used by the most studies were the VABS (n = 25), MSEL (n = 14), notation of presence/absence of verbal language (n = 19) and the Peabody Picture Vocabulary Test (PPVT) (n = 6). The most commonly used standardised assessment of expressive and receptive language was the

**Table 1.** Risk of bias rating on included publications.

| Author | Study participation | Study attrition | Outcome measures |
|---|---|---|---|
| Howlin et al. (2004) | L | L | L |
| Bennett et al. (2014), Davidson & Ellis Weismer (2014), Knorring & Hägglöf (1993) | L | L | M |
| Anderson et al. (2009), Landa & Kalb (2012) | L | M | L |
| Berry (2010), Freeman, Ritvo, Needleman, & Yokota (1985) | L | H | L |
| Bennett et al. (2013), Eriksson et al. (2013) | L | M | M |
| Flanagan et al. (2015), Hedvall et al. (2015), Szatmari et al. (2015) | L | H | M |
| Siller, Hutman, & Sigman (2013) | M | L | L |
| Klintwall, Macari, Eikeseth, & Chawarska (2015) | M | L | M |
| Anderson et al. (2007), Ray-Subramanian & Ellis Weismer (2012) | M | M | L |
| Stone & Yoder (2003), Chawarska, Klin, Paul, Macari, & Volkmar (2009) | M | H | L |
| Anderson, Liang, & Lord (2014), Green & Carter (2014), Magiati et al. (2011), Oosterling et al. (2010), Pellicano (2010b, 2012), Zierhut (2002), Moss, Magiati, Charman, & Howlin (2008), Howlin et al. (2013), Howlin et al. (2014), Cederlund, Hagberg, Billstedt, Gillberg, & Gillberg (2008) | M | M | M |
| Smith, Flanagan, Garon, & Bryson (2015) | M | M | H |
| Bedford, Pickles, & Lord (2015), Ben-Itzchak, Watson, & Zachor (2014), Ellis Weismer & Kover (2015), Georgiades et al. (2014), Hellendoorn et al. (2015), Lombardo et al. (2015), Miniscalco, Rudling, Rastam, Gillberg, & Johnels (2014), Smith, Klorman, & Mruzek (2015), Thurm et al. (2015), Venker, Ray-Subramanian, Bolt, & Weismer (2014), Vivanti, Barbaro, Hudry, Dissanayake, & Prior (2013), Eaves & Ho (1996), Flanagan, Perry, & Freeman (2012), Kleinman et al. (2008), Steele, Joseph, & Tager-Flusberg (2003), Thomas (2009), Thurm et al. (2007) Fernell et al. (2011), Szatmari et al. (2009), Toth et al. (2006), Norrelgen et al. (2014) | M | H | M |
| Meyer (2002) | M | H | L |
| Blacklock, Perry, & Dunn Geier (2014), Perry et al. (2008) Perry, Blacklock, & Dunn Geier (2013), Perry et al. (2011), Pickles et al. (2014), Sullivan (2010) | M | H | H |
| Bennett et al. (2013) | H | M | L |
| Starr, Szatmari, Bryson, & Zwaigenbaum (2003) | H | L | M |
| Bennett et al. (2008), Bopp (2006), Bopp, Mirenda, & Zumbo (2009), Bopp & Mirenda (2011), Eaves & Ho (2004), Haebig, McDuffie, & Ellis Weismer (2013a), Haebig et al. (2013b) | H | H | L |
| Freeman et al. (1991), Jónsdóttir et al. (2007), Sigman & McGovern (2005), Wolf & Goldberg (1986) | H | M | M |
| Bal, Kim, Cheong, & Lord (2015), Pugliese et al. (2015), Woynaroski et al. (2015), Baghdadli et al. (2012), Bagley & McGeein (1989), Ben Itzchak & Zachor (2009), Ben Itzchak & Zachor (2011), Carbonnel-Chabas & Gepner (2009), Darrou et al. (2010), Mosconi, Steven Reznick, Mesibov, & Piven (2009), Munson, Faja, Meltzoff, Abbott, & Dawson (2008), Paul et al. (2008), Pry, Petersen, & Baghdadli (2011), Takeda, Koyama, Kanai, & Kurita (2005), Wodka et al. (2013) | H | H | M |
| Ballaban-Gil, Rapin, Tuchman, & Shinnar (1996) | H | M | H |
| Yoder, Watson, & Lambert (2015), Freeman & Perry (2010), Kobayashi et al. (1992), Mazurek, Kanne, & Miles (2012) | H | H | H |

Note: L: low risk of bias; M: medium risk of bias; H: high risk of bias.

**Table 2.** Language outcome domains and tools.

| Language outcome | Domain | Description | Tools |
|---|---|---|---|
| Standardised clinician-completed | Receptive and expressive syntax | Measures receptive or expressive language more broadly (i.e. range of language domains assessed) | Clinical Evaluation of Language Fundamentals (CELF)<br>Mullen Scales of Early Learning (MSEL)<br>Preschool Language Scales (PLS)<br>Reynell Developmental Language Scales (RDLS)<br>Sequenced Inventory of Communication Development (SICD)<br>Test of Auditory Comprehension of Language (TACL)<br>Test of Oral Language Development (TOLD) |
| | Receptive and expressive vocabulary | Measures receptive or expressive vocabulary only | Expressive One Word Vocabulary Test (EOWVT)<br>Expressive Vocabulary Test (EVT)<br>Peabody Picture Vocabulary Test (PPVT)<br>British Picture Vocabulary Scales (BPVS) |
| Standardised parent-completed | Adaptive language | Measures adaptive communication skills | Vineland Adaptive Behaviour Scales (VABS) |
| | Expressive and receptive vocabulary | Measures number of words understood or expressed | MacArthur Bates Communicative Inventories (CDI) |
| Presence/absence of verbal language | Verbal or phrase language | Measures whether individual had verbal language/were non-verbal or using phrases | ADI-R questions<br>ADOS module<br>Categorical descriptions/rating scales developed by the study, e.g. < 10 words used functionally on a daily basis or the use of phrases |

Preschool Language Scales (PLS) (n = 6). For a variety of reasons (e.g. study did not use same measure at each time point or could not be grouped with others) some studies could not be represented graphically and are presented in the text. If we were not able to obtain required information on language (e.g. only overall scores were provided for the VABS rather than specific subscales), we could not report it in the text or tables but recorded the study as having collected this information in Online Appendix C.

## *Clinician-administered tools*

*Composite receptive and expressive language.* Change in standard scores: Seven studies measured receptive (Ben Itzchak cohort; Berry-Kleinman cohort; Wisconsin cohort; Lombardo, 2015; Paul, 2008; T. Smith, 2015; Vivanti, 2015) and composite expressive language, respectively (Ben Itzchak cohort; Berry-Kleinman cohort; Wisconsin cohort, 2014; Lombardo, 2015; Paul, 2008; T. Smith, 2015; Vivanti, 2015). In all the aforementioned studies, mean language scores for children with ASD were below age-expected levels.

All seven studies that measured composite receptive language reported an increase in mean standard scores for children with ASD (i.e. more gain than expected relative to age-matched peers) (Figure 2(a)). Fifty-seven percent (n = 4/7) of these studies showed a statistically significant ($p < 0.05$) increase in scores over time (Paul, 2008; Berry-Kleinman cohort; Wisconsin cohort; Vivanti, 2015) indicating some language 'catch up' to reference norms. All seven composite expressive language studies reported an increase in mean standard scores with 43% (n = 3/7) reporting a statistically significant ($p < 0.05$) increase (Paul, 2008; Wisconsin cohort; Vivanti, 2015) (Figure 2(b)). Participants in Ben Itzchak cohort and I. Smith (2015) received intensive behavioural interventions. Participants in the

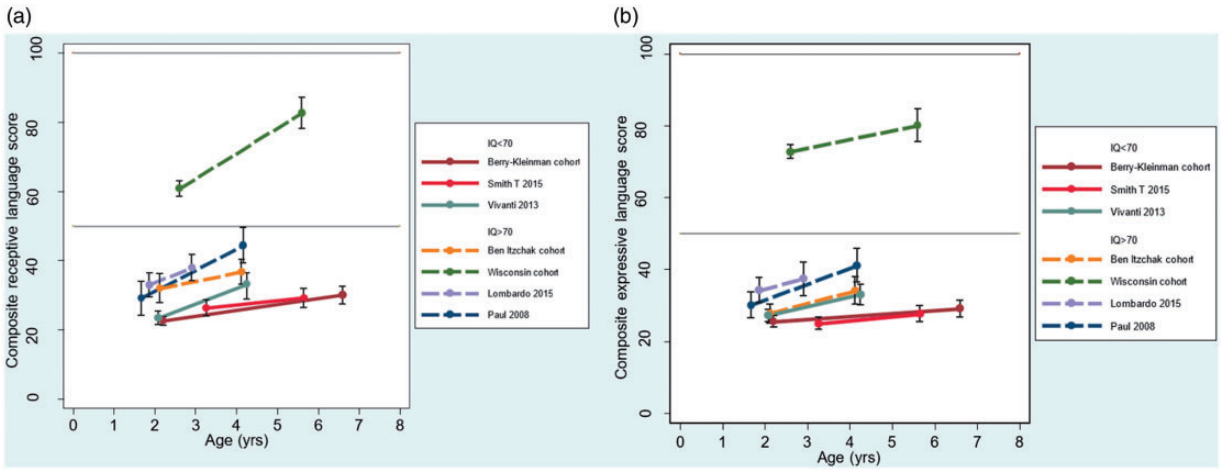


**Figure 2.** Standard scores in composite receptive (a) and expressive language (b) at baseline and follow-up with 95% confidence intervals for studies using the MSEL and PLS.

Note: Mean standard score on MSEL: 50 (upper horizontal grey line); mean standard score on PLS: 100 (lower horizontal grey line). PLS: Wisconsin cohort (2014). MSEL: Ben Itzchak cohort, Berry-Kleinman cohort, Lombardo (2015), Paul (2008), I. Smith (2015), Vivanti (2013). Means for the 'poor' and 'good' outcome groups were combined from Lombardo (2015).

remaining studies received a variety of interventions in the community.

Combining the results of seven studies in a random effects meta-analysis, there was an estimated overall increase of 9.4 units (95% CI 4.9–13.8; $I^2$ (heterogeneity) 86.6%, df 0.6, $p<0.001$) on composite receptive language scales observed in a study over time. Substantial heterogeneity was found between studies. In the meta-analysis, a score of 0 indicates children are progressing at rate expected for their age (i.e. norm references). A score over 0 indicates some 'catch up' or improvement relative to an earlier assessment.

For composite expressive language, combining the results of seven studies in a random effects meta-analysis, there was an estimated overall increase of 5.1 units (95% CI 3.3–7.0; $I^2$ 39.0%, df 0.6, $p=0.132$) on composite expressive language scales observed in a study over time.

Moderator analyses were conducted using meta-regressions in an attempt to explain between study variability in receptive and composite expressive language outcomes. The results indicated baseline language ability (composite receptive language $\beta=0.578$, $SE=0.291$, $p=0.64$; composite expressive language $\beta=0.277$, $SE=0.122$, $p=0.84$), age ($\beta=-7.162$, $SE=5.832$, $p=0.705$; $\beta=-1.911$, $SE=3.264$, $p=0.617$), length of follow-up ($\beta=1.24$, $SE=2.486$, $p=0.756$; $\beta=-0.326$, $SE=1.648$, $p=0.861$) and mean sample IQ ($\beta=-5.907$, $SE=8.557$, $p=0.825$; $\beta=-1.965$, $SE=4.538$, $p=0.707$) did not significantly predict the magnitude of change in receptive or composite expressive language.

Change in age-equivalent scores: Four studies provided age-equivalent scores on combined composite expressive and receptive language (Bopp cohort; Chawarska-Klintwall cohort; Stone, 2003; Ziehurt, 2012) and one study provided separate scores for composite expressive/receptive language in children aged under five years at baseline (I. Smith, 2005). There was divergence away from the expected trajectory for chronological age and the participant's language age over time in two of the four studies in studies reporting on combined composite receptive/expressive language (Figure 3(a)). In the studies of children who were under five years at baseline the mean language age was substantially below chronological age at baseline, ranging from 11 months behind at 4.1 years to 2.8 years delay at 4.2 years of age. At follow-up, mean language age ranged from 1.8 years behind at 5.2 years to 3.3 years behind at 6.2 years. One study provided only log adjusted scores which showed an increase in language age for participants over time (Siller, 2013).

One study followed children into adulthood. This study of children from 12 (middle childhood) to 19 years (adolescence) showed a much lower mean language age based on the RDLS/CELF tool than chronological age at baseline with a substantially increasing gap between the two metrics over time. This was reflected in quite flat language trajectories with a 1.4 month language gain over six years for children with an IQ < 70 and a 2.5 year gain in language over a 6–7-year period for children with an IQ > 70 (Sigman & McGovern, 2005).

Change in raw scores: Three studies provided raw scores on clinician-administered tools (Bopp cohort; Hellendoorn, 2015; Ray Subramanian, 2012), detailed in Online Appendix E. Hellendoorn (2015) was the only population-based study. In composite receptive language, mean scores increased from 13.39 to 29.19

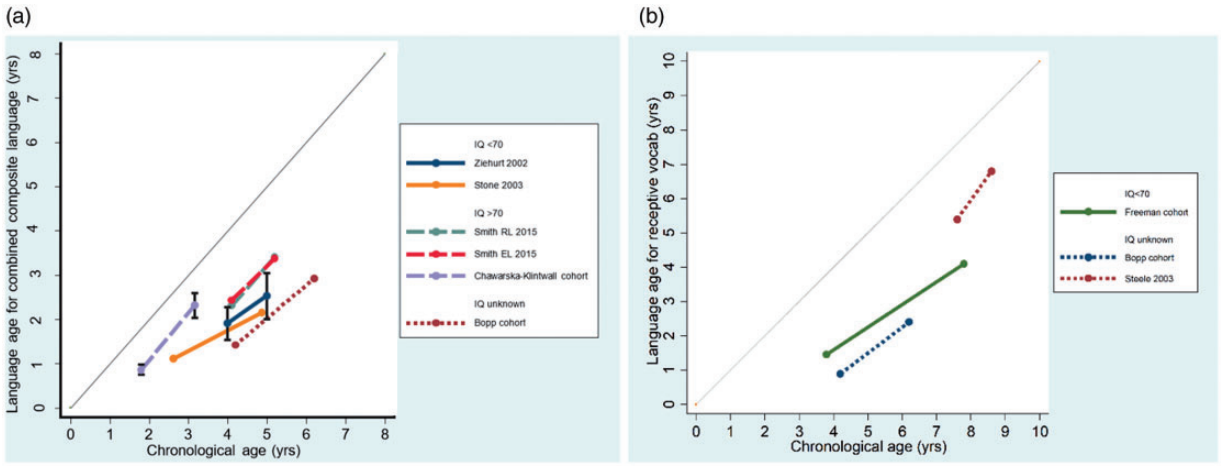


**Figure 3.** Combined composite receptive/expressive language (Figure A) and receptive vocabulary (Figure B) language age relative to chronological age.

Note: The grey line represents a chronological age-equivalent to language age, the line of 'average' development. Stone (2003) only reported composite expressive age-equivalents. I. Smith (2015) reported receptive language (RL) and expressive language (EL) age-equivalent scores separately.

over two years (Bopp cohort), 16.04 to 37.22 over 1.5 years (Hellendoorn, 2015) and 20.35 to 28.44 over one year (Ray Subramanian, 2012). In composite expressive language mean scores increased from 14.38 to 26.05 over two years (Bopp cohort), 17.02 to 36.49 over 1.5 years (Hellendoorn, 2015) and 24.96 to 32.25 over one year (Ray Subramanian, 2012). This indicates improvement in language ability over time in all studies; however, assessment of progress relative to what is expected is not possible with raw scores.

One study could not be presented graphically because it only reported standard scores on the PLS at follow-up (Eaves, 2004). In this study, children with autistic disorder (n = 36) achieved mean scores of 52.9 (SD 16.2) in expressive and 54.9 (SD 17.3) in receptive language at follow-up and those with PDD-NOS achieved mean scores of 63.0 (SD 27.8) in expressive and 64.3 (SD 26.7) receptive language. It was not possible to comment on change over time in this study because age-equivalents were used at baseline and language measures varied across time points.

*Expressive and receptive vocabulary.* Change in standard scores: Three studies reported receptive (Pelicano cohort; Magiati-Moss cohort; Thomas, 2009) and one on expressive vocabulary (Magiati-Moss cohort). In all but one study (Pelicano cohort), that only included children with verbal IQ > 80, baseline scores were substantially lower than age-expected levels.

Two of three studies (Thomas, 2009; Magiati-Moss cohort) reported an increase in standard scores over time, from 48.6 (95% CI 43.1–54.1) to 55.6 (CI 48.4–62.6) over 6.9 years in one study (Magiati-Moss cohort) and from 70.9 (CI 64.4–77.3) to 77.5 (CI 71.3–83.7) over five years in another (Thomas, 2009). The third demonstrated a decrease in standard scores over 2.7 years (Pelicano cohort), where scores decreased from 97.1 (CI 93.7–100.5) to 93.9 (CI 88.7–99.1). The expressive vocabulary study reported that mean standard score decreased from 69.3 (CI 65.1–73.4) to 65.49 (CI 59.7–71.3) (Magiati-Moss cohort).

No vocabulary study demonstrated significant catch up or loss relative to age-matched peers. Despite having lower scores at baseline and follow-up, on average, the children with ASD progressed at a comparable rate to reference norms. In all studies, there was greater variability in standard scores over time, evidenced by wider confidence intervals at follow-up compared to baseline, indicating substantial heterogeneity in language trajectories.

Change in age-equivalent scores: All studies providing data on age-equivalents reported gains in mean receptive vocabulary age over time (Bopp cohort; Freeman cohort; Steele, 2003). Of these three studies, one demonstrated a degree of 'catch up' to chronological age with a gap between chronological and language age of 2.2 years at baseline and 1.8 years at follow-up (Steele, 2003). By contrast, the remaining two studies reported a widening gap between language and chronological age (3.31 years at baseline to 3.8 years at follow-up, Bopp cohort; 2.6 years at baseline to 3.7 years at follow-up, Freeman cohort), Figure 3(b). Only one study followed children into adulthood (Howlin, Goode, Hutton, & Rutter, 2004). This study provided age-equivalents at study end. Participants had an average language age of 8.26 (SD 6.21) at a mean chronological age of 29.33 years (Howlin et al., 2004). Participants in all studies received a variety of interventions in the community rather than the study being an intervention trial.

Change in raw scores: Two studies showed increased mean raw scores on expressive and receptive vocabulary (Bopp cohort; Steele, 2003), detailed in Online Appendix E, suggesting improvement in language over time, although assessment of progress relative to what is expected is not possible with raw scores.

## Parent-report tools

*Adaptive language.* Change in standard scores: Thirteen studies had standard score data able to be extracted from the VABS (Figure 4). A higher score on the VABS communication domain is an indication of better functioning. All studies reported mean baseline scores below the average range. Eighty-five percent of studies (n = 11/13) reported higher scores at follow-up than baseline and 31% (n = 4/13) of these studies reported a statistically significant ($p < 0.05$) improvement in scores from baseline to follow-up (Blacklock, 2013; Landa, 2012; I.-B. Smith, 2015; Sullivan, 2010).

Two studies (Eaves, 2004; Magiati-Moss cohort) reported a decline in scores from baseline to follow-up and one of these was statistically significant ($p < 0.05$) decline (Magiati-Moss cohort). There were no substantial differences between Magiati-Moss cohort and the other studies in participant characteristics (e.g. IQ levels) or methodology to explain the different trajectories. The Magiati-Moss study, however, did have: a higher proportion of individuals with autistic disorder relative to ASD compared to some other studies, a longer follow-up period, and individuals were older at follow-up. In addition, the Magiati-Moss cohort recruited children in the late 1990s; substantially earlier than some of the other studies. The other three studies where participants showed the least amount of language gain recruited participants at the three next earliest time points (Berry-Kleinman cohort; Eaves, 2004; Meyer, 2002). It was not possible to investigate whether type, dose or frequency of intervention may have explained differences in study findings because description of interventions was not detailed within all papers.

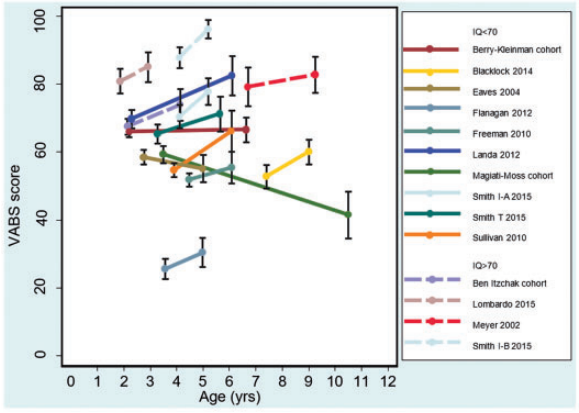


**Figure 4.** Standard scores on the VABS communication subscale at baseline and follow-up.

Note: I.-A. Smith and I.-B. Smith are two cohorts from the same study grouped into IQ < 70 and IQ > 70 (I. Smith, 2015). Participants in the Flanagan (2012) study had substantially lower IQs (Time 2 IQ 39.50 (SD 18.93)) than the participants in other studies. One study presented data by poor and good language groups, here we provide a combined mean (Lombardo, 2015). One study of children with IQ > 70 (mean: 107.03) was unable to be presented graphically. This study reported a mean standard score of 86.44 (SD 16.53) at 8.3 years. At 12.9 years, 12% of children improved in their standard scores, 68% remained unchanged, 20% decreased (Pugliese, 2016).

In all but two studies (Meyer, 2002; I.-B. Smith, 2015), there was greater variability in standard scores between participants at follow-up than at baseline, evidenced by wider CIs around the mean at outcome. Participants in Freeman (2010), Ben Itzchak cohort, Sullivan (2010), Landa (2012), T. Smith (2015), Flanagan (2012) and Blacklock (2013) received an intensive behaviour intervention.

Combining the results of the 14 studies in a random effects meta-analysis, there was an estimated overall increase of 4.0 units on the VABS scale observed in a study over time (95% CI 0.8, 7.4; $I^2$ 82.1%; df 13; $p = 0.016$). There was substantial heterogeneity between studies. Moderator analyses were conducted using meta-regressions in an attempt to explain between study variability in adaptive language outcomes. None of the study-level covariates including VABS score at baseline ($\beta = 0.0611$, SE = 0.128, $p = 0.185$), age ($\beta = 0.366$, SE = 0.9334, $p = 0.344$), length of follow-up ($\beta = -0.529$, SE = 1.64, $p = 0.667$) and reported average IQ (<70 or >70; $\beta = -1.04$, SE = 4.573, $p = 0.561$) provided any insight into the observed heterogeneity in adaptive language outcomes.

Change in age-equivalent scores: One study presented age-equivalents on the VABS for the communication scale. Children were followed from 3.6 years of age for two years and gained 0.73 age-equivalence points per month (Munson-Toth). Two studies provided age-equivalents on the VABS split into receptive and expressive communication. In one study, children were aged 1.8 years at baseline and gained 1.85 years in expressive communication and 1.92 years in receptive communication over 2.5 years (Paul, 2008). In the other study that contained some children without ASD, children gained 6.9 years in receptive communication and 6.7 years in expressive communication over 17 years (Anderson–Lord cohort). This was the only study that followed children into adulthood using the VABS.

Five studies did not provide communication subscale scores from the VABS (Green, 2014; Mosconi, 2009; Stockholm cohort; Pathways cohort), meaning data could not be extracted. In the aforementioned

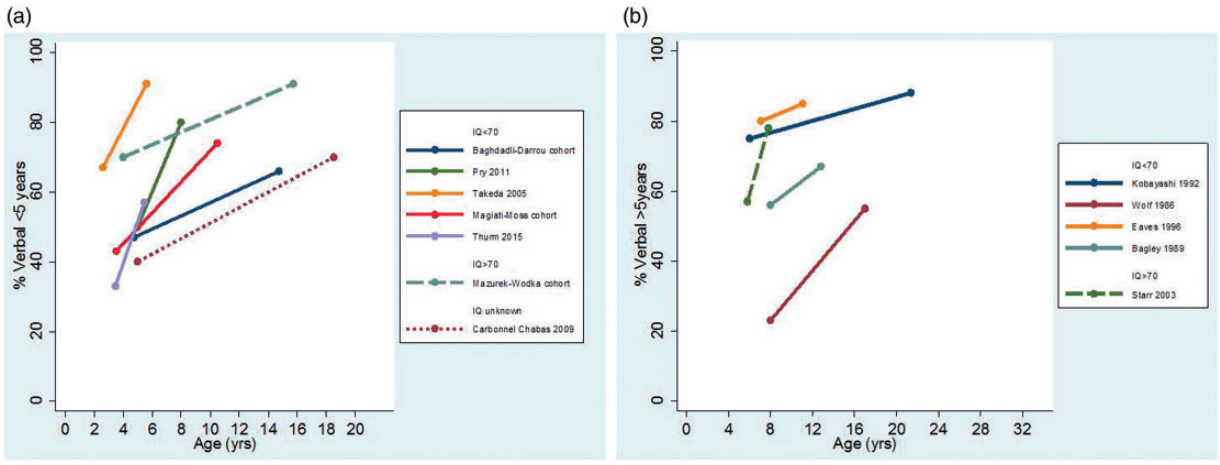


**Figure 5.** Proportion of individuals aged < 5 years (a) and ≥ 5 years (b) who used verbal language at baseline and follow-up.

cases where data was reportedly collected, but not presented, we contacted authors but were not successful in receiving these data.

*Expressive vocabulary.* Change in raw scores: Three studies reported raw data on parent-reported expressive vocabulary using the CDI (Oosterling, 2010; Bopp cohort; Woynaroski-Yoder cohort). The number of words gained from baseline to follow-up ranged from 35 to 76 over 1.25 years in one study (Oosterling, 2010). In one study, an average of 58 words was gained over 4.41 years (Bopp cohort) and in another study (Yoder-Woynaroski cohort), 73 words were gained over 1.3 years (Online Appendix E). Authors of one study were contacted but were not able to provide results from both time points, although this information had been collected (Miniscalco, 2014).

## *Presence or absence of verbal language and phrase language*

*Verbal language.* Eighteen studies reported on the presence of verbal language (nonverbal or verbal) at outcome (Bagley, 1989; Ballaban-Gil, 1996; Carbonnel-Chabas, 2009; Eaves, 1996; Freeman cohort; Howlin et al., 2004; Knorring, 1992; Kobayashi, 1992; Pry, 2011; Sigman & McGovern, 2005; Bennett-Starr cohort; Takeda, 2005; Anderson–Lord cohort; Wolf, 1986; Baghdadli-Darrou cohort; Mazurek-Wodka cohort; Stockholm cohort). Of these studies, four followed the children into adulthood (Ballaban-Gil, 1996; Howlin et al., 2004; Kobayashi, 1992; Wolf, 1986) and three were population-based (Knorring, 1992; Pry, 2011; Stockholm cohort). As before, we included studies reporting on 'minimally verbal' and 'nonverbal' children. We have graphed studies (n = 12) that report both baseline and follow-up data. Seven studies included children aged five years and under at baseline and five studies children over five years at baseline (Figure 5(a,b)). The proportion of children who were verbal at baseline ranged from 40% to 70% and at follow-up 66% to 91%. For children over five years, the proportion of children verbal at baseline ranged from 23% to 80% and at follow-up 55% to 88%. Nineteen to thirty percent of children aged five years and under gained verbal language. For children aged over five years 5–32% gained verbal language over the course of study.

The Howlin cohort followed individuals with an IQ of ≥ 70 from 6.75 years (SD 2.8) to 44.2 years (SD 9.3). In this study, 75% (45/60) of children were verbal and by adulthood 95% (57/60) were verbal. Six studies provided only follow-up data that could be extracted on the proportion of individuals who were verbal (Ballaban-Gil, 1996 (93/99; 94%); Howlin et al., 2004 (61/67; 91%); Knorring, 1993 (25/34; 74%); Freeman cohort (44/53; 83%); Sigman & McGovern, 2005 (23/48; 49%)). One study grouped by those who were nonverbal (25/165; 15%) and minimally verbal (17/165; 10%) at outcome (Stockholm cohort). In addition, Anderson–Lord cohort reported 58% (N = 74) of children who were nonverbal at two years were verbal at nine years. In adults, between 55% and 94% were verbal. In population-based studies, between 74% and 80% were verbal.

Combining the results of the 12 studies in a random effects meta-analysis, it was found 21% of children gained verbal language from baseline to follow-up (95% CI 0.16–0.27; $I^2$ 81.38%; df 10; p = 0.000). There was considerable heterogeneity between studies. Moderator analyses were conducted using meta-regressions in an attempt to explain between study variability in the proportions of children who gained verbal language. None of the study-level covariates including age at baseline (β = −19.105, SE = 10.91, p = 0.123), length of follow-up (β = −3.326, SE = 3.703, p = 0.399) and

reported average IQ (<70 or >70; $\beta = -76.254$, SE = 50.41, p = 0.174) provided any insight into the observed heterogeneity in the proportions of children who gained verbal language.

*Phrase and/or functional language.* Eleven studies reported on participants who gained the ability to use phrases (Anderson–Lord cohort; Jonsdottir, 2007; Kobayashi, 1992; Magiati-Moss cohort; Pry, 2011; Sigman & McGovern, 2005; Thurm, 2015; Wolf, 1986; Howlin cohort; Baghdadli-Darrou cohort; Stockholm cohort). Of these studies, only five followed children into adulthood (Anderson–Lord cohort; Kobayashi, 1992; Sigman & McGovern, 2005; Wolf, 1986; Howlin cohort) and two were population-based (Pry, 2011; Stockholm cohort). Those studies with both baseline and follow-up are presented in Figure 6. The percentage of participants at the end of the study who were able to use phrases ranged between 17% (mean age 7.8 years at follow-up) and 85% (mean age 9 years at follow-up and for children with PDD only (i.e. excluded autistic disorder)). For those children aged over eight years at baseline, 20% of non-phrase speaking children in one study gained the ability to use phrases over a period of 10 years (Wolf, 1986).

Four studies provided only follow-up data that could be extracted on the proportion of individuals using phrases (Howlin cohort 33/68 (49%), Sigman & McGovern, 2005 25/48 (52%), Stockholm cohort 123/165 (75%), Anderson–Lord cohort (autistic disorder 48% and PDD-NOS 85%)). One study reported on children who used functional language at baseline (20%) and at follow-up (47%) (Baghdadli-Darrou cohort). In adults, between 33% and 67% were able to use phrases. In population-based studies, between 48% and 75% were able to use phrases.

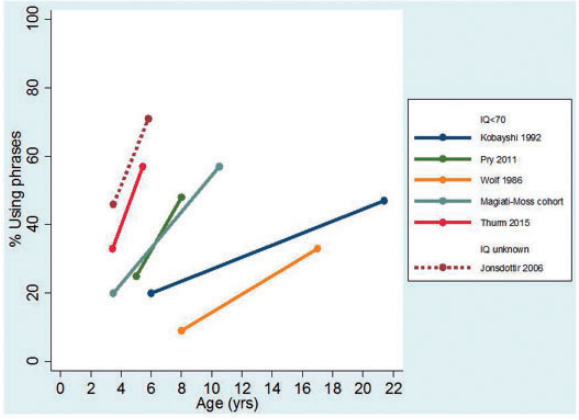


**Figure 6.** Proportion of individuals with ASD who used phrases (and longer) at baseline and follow-up.

Combining the findings from all studies, the proportion of participants at the end of the study duration who were verbal ranged between 55% (median age 20 years at follow-up) and 95% (mean age 44.2 years at follow-up). The proportion of participants using phrases ranged from 33% (mean age at follow-up 17 years) to 85% (mean age at follow-up 9 years).

## Discussion

Language development in individuals with ASD is complex and heterogeneous. Not all children with ASD experience structural and adaptive language difficulties, but the majority do (Levy et al., 2010). A substantial number of studies (n = 54) collectively including a large number of children (N = 5064) have been published that have provided valuable longitudinal information on language outcomes in ASD. To date, however, it has been challenging for clinicians and parents to interpret these findings. This is because there is inconsistency and variability in study methods and participant characteristics. Moreover, the quality of the studies varies, which may contribute to difficulty weighing the significance of the findings from each study.

### *Summary of findings from studies*

Here, we synthesised available information on verbal language outcomes and assessed the quality of included studies. Substantial variability was seen in mean language scores and slopes of the language trajectories across studies. Yet, studies generally reported similar overall findings, whereby mean baseline and outcome scores on standardised language tests were consistently lower for children with ASD than reference norms, with one exception that included only children with IQ > 80 (Pelicano cohort). Moreover, in all included studies, the majority of children with ASD continued to make positive language gains, including children aged over five years at baseline. Language gains occurred across multiple domains, including composite language (receptive and expressive), vocabulary (receptive and expressive), adaptive language and the acquisition of verbal language.

Children under nine years progressed at a rate comparable to reference norms in the majority of studies (n = 11/18; 61%) based on receptive and composite expressive language and receptive and expressive vocabulary mean standard scores. Seven studies showed a significantly faster rate of progress or 'catch up' compared to reference norms in receptive (n = 4/7; 57%) and composite expressive language (n = 3/7; 43%). When studies were combined in the meta-analysis, there was overall gain in composite receptive language (9.3 units) and composite expressive language

(5.1 units) mean scores. Findings from studies that reported age-equivalent scores for composite language and receptive vocabulary were less consistent with evidence of some divergence from an age expected rate of progress in some studies but not others. The majority of studies (n = 8/13; 62%) reporting data on adaptive language in children aged under 11 years demonstrated a rate of progress comparable to age expected norms and 31% (n = 4) of studies reported some 'catch up' to age expected norms, based on standard scores. Only one study showed a significant decrease in standard scores. When studies were combined in the meta-analysis, there was overall gain in adaptive language mean scores (4.0 units). Gains in the aforementioned studies were observed both in children with and without intellectual disability, across different ASD subgroups and across studies where participants accessed intensive behavioural interventions or intervention in the community.

In this review, age, baseline language scores, IQ and length of follow-up did not moderate between study differences in structural or adaptive language growth or the acquisition of verbal language. Our findings contrast with previous studies that have identified IQ and baseline language ability to be a significant predictor of later language outcome (e.g. Ellis Weismer & Kover, 2015; Thurm et al., 2007; Wodka, Mathy, & Kalb, 2013); however, differences may be explained by the varied study methodology and outcomes examined. We were not able to investigate ASD severity as a potential moderator of language growth because of the heterogeneity of tools, versions of tools and ways scores were reported by included studies.

The findings from studies of individuals followed from middle childhood to adulthood are less consistent than for younger children but suggest the same rate of progress experienced during childhood may not be maintained beyond nine years of age. The adaptive and structural language trajectories reported in the few studies conducted during adolescence were flatter for children from middle childhood to adolescence with a growing gap between language ability in children with ASD and reference norms (Pickles, Anderson, & Lord, 2014; Sigman & McGovern, 2005). Similar findings were reported in studies of adults (Howlin et al., 2004). There are a number of potential explanations for slowing of language trajectories into adulthood. There is some evidence of a 'second wave of deficit' or 'second hit' that emerges in the second decade of life that may impact trajectories of development (Minshew & Williams, 2007; Picci & Scherf, 2015). It has been estimated that 30% of children with ASD experience deterioration in functioning over several years following puberty onset. It has been hypothesised that the combination of co-occurring increase in adolescent-specific developmental tasks, increasing complexity of language and social communication required during adolescence, combined with pubertal hormone surges may contribute to this divergence or slowing of developmental trajectories (Picci & Scherf, 2015).

Alternatively, it is possible that more recently published studies including younger cohorts have included individuals diagnosed using broader criteria than previous studies. As such, more recent studies may include individuals with less severe phenotypes and a smaller proportion of children with intellectual disabilities (Keyes et al., 2012). This was evidenced by the studies demonstrating faster rates of development being those more recently published. Finally, interventions, particularly intensive interventions, are more readily available in more recent years and this may have an impact on outcomes on children under eight years. Many funding programs are also focused on early rather than life course interventions, and it is possible that as interventions are reduced progress also decreases. Further research on language trajectories into and throughout adulthood is vital if we are to understand the communication support needs of adults with ASD in the future (Magiati et al., 2014).

### *Considerations for interpreting data in this review*

The purpose of this review was to present currently available information from studies reporting language change over time in ASD. We included a variety of measures so that various language domains could be compared and differences considered. Not all measures in the review are comprehensive language measures, yet were considered important to include as they add to the overall picture of language development and are commonly used with individuals who have ASD. For example, adaptive language tools provide important information on how children are using their language functionally and are an important complement to formal testing which is not always suitable for children with ASD. We have used methods to minimise the risk of a biased assessment of the evidence by being comprehensive in our search, assessing risk of bias and taking IQ of included cohorts in to account.

Here, we reported findings at a group level (group means) as this is how included studies reported their findings, yet substantial variability was seen within cohorts for individual baseline and outcome scores, evidenced by large standard deviations. Mean trajectories will not apply to all children. Some children may make dramatic progress, while others make little progress. Future studies should examine individual variation and present data for important subgroups so we can refine predictions of prognosis. Furthermore, a review

specifically designed to analyse predictors of language outcome will aid our understanding of important factors that may impact language trajectory.

We have used standard scores cross-sectionally to report trajectories here to increase clinical utility yet the shape of trajectories between time points is not known. The potential for regression to the mean when plotting trajectories has been underscored by some authors (Tomblin, Zhang, Buckwalter, & O'Brien, 2003) and consideration should be given to the heterogeneity between studies when interpreting the meta-analysis findings. Furthermore, we used reference norms for comparison and did not require studies to include typically developing children. It is known that some variability in scores over time is not unique to ASD (Conti-Ramsden, St Clair, Pickles, & Durkin, 2012; McKean et al., 2015; Ukoumunne et al., 2012) so ideally studies would include other groups of children so language development in ASD can be placed within a developmental context.

There is some evidence from this review that studies that used age-equivalent scores reported slower rates of development than those studies reporting standard scores and there may be some bias in the way data are presented by studies. Studies of children who are more severely affected may be more likely to report age-equivalents as children may not reach the basal levels required for reporting standard scores and there may be floor effects.

A related consideration is that it can be challenging (and inappropriate) to assess some children (e.g. minimally verbal children) using standardised tools as the children may not have the ability to complete the tasks and/or may not reach the basal level for scoring (Kasari et al., 2013). No studies in this review presented detailed information on the prognosis of language development for these individuals. While it is recognised that there are problems around the psychometric properties of age-equivalent scores, this study included studies that only presented age-equivalents as we wanted to reduce bias in study selection and be inclusive of all studies presenting information on language outcomes, not only those presenting standardised scores. It could be hypothesised that studies presenting raw scores may contain higher proportions of children with more significant language difficulties as it may be standard scores cannot be obtained for these children. Furthermore, it is important to note there may be a proportion of individuals with ASD with severe language impairment who are not represented in studies in this review. The lack of inclusion of these individuals may have resulted in an over-estimate of language gain in summary scores. There is some evidence that children with ASD who have lower language ability at baseline have flatter trajectories compared with children with higher language (Tek et al., 2013). Finally, the VABS is a parent-reported measure of adaptive communication and incorporates pragmatic/higher-order language as well as core structural items. Differences in both the methods of how language is assessed and the items included in the tools should be considered when interpreting the findings in this review.

### *Research implications*

Improvements are needed in both the conduct and report of language outcome studies for ASD cohorts. Recommendations and guidelines have been developed for designing high-quality prognosis studies and for assessing the quality of studies (Hayden et al., 2013). These best practice guidelines were used here to assess risk of bias in included studies. The majority of included studies were rated medium to high risk of bias, with less than 5% at low risk of bias. While the high risk of bias in some studies may be because they were not intentionally planned prognosis studies, best practice methodological approaches for prognosis are crucial for interpreting and weighting the findings of individual studies. It is important that information be collected prospectively on a sample of children diagnosed according to best practice at study commencement. Inclusion criteria for this review stated all children required a diagnostic assessment using established diagnostic criteria at baseline (e.g. DSM, ICD); however, a substantial number (33%) of studies were retrospective and did not report on the children who were not available at follow-up. This may have the potential to bias toward a positive outcome.

Ideally, studies should recruit from a population-based sample or from clinical services that provide services for the broad population of children with ASD, so the individuals are representative of individuals in the general population with ASD. Only five studies derived their samples from a population source, with the rest being selected clinical samples. Clinical samples have been reported to be skewed toward more severely affected individuals and as such the findings may be less transferrable to the full range of individuals with ASD. Data from such studies can still be useful, but application is limited to children or adults with the same types of strengths and difficulties.

Studies should ensure high retention of participants over time and report on differences between participants that were lost to follow-up and those who were not. Of the studies in this review that were prospective, 53% retained more than 80% of participants at follow-up. Few studies provided detailed descriptions of the individuals who were lost to follow-up. Finally, clinicians completing the assessments should be blind to individual's baseline characteristics and diagnosis to

**Table 3.** Key messages.

| |
|---|
| Between 55-95% of individuals who were minimally verbal became verbal at follow up |
| In all but one study children with ASD had mean scores below average at baseline and follow up |
| Approximately 96% of studies showed that children aged under 11 with ASD tracked at comparable rate (or better) of language development to age-expected rates, based on standard scores. This occurred in children with and without intellectual disability and in children who received intensive behavioural interventions and those who received intervention in the community |
| In 40% of studies reporting standard scores children under 11 years demonstrated some 'catch up' to peers (i.e. faster rate of progress than age-expected norms) |
| There was some evidence language development began to slow down after 11 years of age |
| There is a high need for research on adolescent and adult language outcomes |

avoid bias. In this review, only 10% of studies provided information about blinding of clinicians. Individuals in studies reporting age-equivalent scores had slower trajectories than studies reporting standard scores. This may reflect bias, as it is possible that studies may choose to report age-equivalents when they contain participants who are lower functioning or less able to complete standardised tools or obtain standard scores.

### *Implications*

Despite methodological variations, there was consistency in overall findings, which although not fine-tuned for individuals will be useful for clinicians and those caring for people with ASD. Key messages are highlighted in Table 3.

### *Future directions*

It is clear much needs to be done to refine the accuracy of predictions of language outcomes. Understanding how language trajectories may change across the lifespan and the developmental periods where individuals may be more 'vulnerable' to slower progress (e.g. before 2 years and after 10 years) is highly relevant for policy makers and service providers so they can accurately plan future funding, support and service needs for individuals with ASD across the lifespan. Moreover, families and clinicians need clear and accurate information based on their child's characteristics regarding the likely communication outcomes for the children they care for with ASD so they can better anticipate ongoing needs. A more fine-tuned understanding of language trajectories will also enable more tailored interventions.

Very few studies reported individual trajectories or clinical subgroups of ASD. Few studies assessed individuals beyond the age of 11 years using language specific tools. We need information to personalise and apply the findings from the studies to individuals who present at different ages and with a broad range of clinical characteristics. This study has also highlighted an important subgroup with ASD who fail to develop the ability to speak (5–45%). The implications of being unable to speak are substantial in terms of the impact on participation and function and the likely support needs of these individuals relative to children who are able to speak. Evidence-based interventions are sorely lacking for this population but are crucial if we are to prevent the adverse sequelae that are likely to accompany poor language outcomes (Kasari et al., 2014; Paul, Campbell, Gilbert, & Tsiouri, 2013; Tager-Flusberg et al., 2016).

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The researchers acknowledge the Australian NHMRC for salary support through a Practitioner Fellowship #1105008 (A.M); NHMRC Centre of Research Excellence in Speech and Language Neurobiology #1116976 (A.M, A.B). This paper is based on work completed by A.B for her PhD. A.B was supported by an Australian Postgraduate Award scholarship. Infrastructure support was provided by the Victorian Government's Operational Infrastructure Support Program. We wish to thank the William Collie Trust Fund for their financial support. This funding organization was not involved in the development, design, analysis, or interpretation of the study. We wish to thank Kim Jachno for statistical support.

## Supplementary Material

Supplementary material is available for this article online.

## References

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed., text revision). Washington, DC: American Psychiatric Publishing.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: American Psychiatric Publishing.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Association.

Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., ... Pickles, A. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. *Journal of Consulting and Clinical Psychology*, *75*(4), 594–604.

Boucher, J. (2012). Research review: Structural language in autistic spectrum disorder-characteristics and causes. *Journal of Child Psychology and Psychiatry*, *53*(3), 219–233.

Bal, V. H., Katz, T., Bishop, S. L., & Krasileva, K. (2016). Understanding definitions of minimally verbal across instruments: Evidence for subgroups within minimally verbal children and adolescents with autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *57*(12), 1424–1433.

Conti-Ramsden, G., St Clair, M. C., Pickles, A., & Durkin, K. (2012). Developmental trajectories of verbal and nonverbal skills in individuals with a history of specific language impairment: From childhood to adolescence. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1716–1735. doi:10.1044/1092-4388(2012/10-0182)

De Giacomo, A., & Fombonne, E. (1998). Parental recognition of developmental abnormalities in autism. *European Child and Adolescent Psychiatry*, *7*(3), 131–136.

Ellis Weismer, S., & Kover, S. T. (2015). Preschool language variation, growth, and predictors in children on the autism spectrum. *Journal of Child Psychology and Psychiatry*, *56*(12), 1327–1337. doi:10.1111/jcpp.12406

Gernsbacher, M. A., Morson, E. M., & Grace, E. J. (2016). Language and speech in autism. *Annual Review of Linguistics*, *2*, 413–425. doi:10.1146/annurev-linguist-030514-124824

Gillespie-Lynch, K., Sepeta, L., Wang, Y., Marshall, S., Gomez, L., Sigman, M., & Hutman, T. (2012). Early childhood predictors of the social competence of adults with autism. *Journal of Autism and Developmental Disorders*, *42*(2), 161–174. doi:10.1007/s10803-011-1222-0

Gilliam, J. (1995). *Gilliam autism rating scale examiner's manual*. Austin, TX: Pro-Ed.

Hayden, J. A., van der Windt, D. A., Cartwright, J. L., Côté, P., & Bombardier, C. (2013). Assessing bias in studies of prognostic factors. *Annals of Internal Medicine*, *158*(4), 280–286. doi:10.7326/0003-4819-158-4-201302190-00009

Herlihy, L., Knoch, K., Vibert, B., & Fein, D. (2015). Parents' first concerns about toddlers with autism spectrum disorder: Effect of sibling status. *Autism*, *19*(1), 20–28. doi:10.1177/1362361313509731

Hofvander, B., Delorme, R., Chaste, P., Nydén, A., Wentz, E., Ståhlberg, O., ... Leboyer, M. (2009). Psychiatric and psychosocial problems in adults with normal-intelligence autism spectrum disorders. *BMC Psychiatry*, *9*, 35–35. doi:10.1186/1471-244X-9-35

Howlin, P. (2003). Outcome in high-functioning adults with autism with and without early language delays: Implications for the differentiation between autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *33*(1), 3–13.

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *45*(2), 212–229.

Howlin, P., & Moss, P. (2012). Adults with autism spectrum disorders. *Canadian Journal of Psychiatry*, *57*(5), 275–283.

Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research*, *6*(6), 479–493. doi:10.1002/aur.1334

Kasari, C., Kaiser, A., Goods, K., Nietfeld, J., Mathy, P., Landa, R., ... Almirall, D. (2014). Communication interventions for minimally verbal children with autism: A sequential multiple assignment randomized trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *53*(6), 635–646. doi:10.1016/j.jaac.2014.01.019

Kenworthy, L., Wallace, G. L., Powell, K., Anselmo, C., Martin, A., & Black, D. O. (2012). Early language milestones predict later language, but not autism symptoms in higher functioning children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *6*(3), 1194–1202. doi:10.1016/j.rasd.2012.03.009

Keyes, K. M., Susser, E., Cheslack-Postava, K., Fountain, C., Liu, K., & Bearman, P. S. (2012). Cohort effects explain the increase in autism diagnosis among children born from 1992 to 2003 in California. *International Journal of Epidemiology*, *41*(2), 495–503. doi:10.1093/ije/dyr193

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes*, *16*(2–3), 287–308.

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

Levy, S. E., Giarelli, E., Lee, L.-C., Schieve, L. A., Kirby, R. S., Cunniff, C., ... Rice, C. E. (2010). Autism spectrum disorder and co-occurring developmental, psychiatric, and medical conditions among children in multiple populations of the United States. *Journal of Developmental and Behavioral Pediatrics*, *31*(4), 267–275.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H. Jr, Leventhal, B. L., DiLavore, P. C., ... Rutter, M. (2000). 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*, *30*(3), 205–223.

Magiati, I., Tay, X. W., & Howlin, P. (2014). Cognitive, language, social and behavioural outcomes in adults with autism spectrum disorders: A systematic review of longitudinal follow-up studies in adulthood. *Clinical Psychology Review*, *34*(1), 73–86. doi:10.1016/j.cpr.2013.11.002

Mawhood, L., & Howlin, P. (2000). Autism and developmental receptive language disorder – A comparative follow-up in early adult life. I: Cognitive and language outcomes. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(5), 547.

McKean, C., Mensah, F. K., Eadie, P., Bavin, E. L., Bretherton, L., Cini, E., & Reilly, S. (2015). Levers for language growth: Characteristics and predictors of language trajectories between 4 and 7 years. *PloS One*, *10*(8), e0134251. doi:10.1371/journal.pone.0134251

Minshew, N. J., & Williams, D. L. (2007). The new neurobiology of autism – Cortex, connectivity, and neuronal organization. *Archives of Neurology*, *64*(7), 945–950.

Mullen, E. (1995). *Mullen scale of early learning*. Circle Pines, MN: American Guidance Service Inc.

Norrelgen, F., Fernell, E., Eriksson, M., Hedvall, Å., Persson, C., Sjölin, M., ... Kjellmer, L. (2014). Children with autism spectrum disorders who do not develop phrase speech in the preschool years. *Autism*, *19*(8), 934–943. doi:10.1177/1362361314556782

Nyaga, V. N., Arbyn, M., & Aerts, M. (2014). Metaprop: A Stata command to perform meta-analysis of binomial data. *Archives of Public Health*, *72*(1), 39. doi:10.1186/2049-3258-72-39

Paul, R., Campbell, D., Gilbert, K., & Tsiouri, I. (2013). Comparing spoken language treatments for minimally verbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(2), 418–431. doi:10.1007/s10803-012-1583-z

Picci, G., & Scherf, K. S. (2015). A two-hit model of autism: Adolescence as the second hit. *Clinical Psychological Science*, *3*(3), 349–371. doi:10.1177/2167702614540646

Pickett, E., Pullara, O., O'Grady, J., & Gordon, B. (2009). Speech acquisition in older nonverbal individuals with autism: A review of features, methods, and prognosis. *Cognitive and Behavioral Neurology*, *22*(1), 1–21. doi:10.1097/WNN.0b013e318190d185

Pickles, A., Anderson, D. K., & Lord, C. (2014). 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, and Allied Disciplines*, *55*(12), 1354–1362. doi:10.1111/jcpp.12269

Rapin, I., Dunn, M. A., Allen, D. A., Stevens, M. C., & Fein, D. (2009). Subtypes of language disorders in school-age children with autism. *Developmental Neuropsychology*, *34*(1), 66–84.

Rose, V., Trembath, D., Keen, D., & Paynter, J. (2016). The proportion of minimally verbal children with autism spectrum disorder in a community-based early intervention programme. *Journal of Intellectual Disability Research*, *60*(5), 464–477. doi:10.1111/jir.12284

Schopler, E., Reichler, R. J., DeVellis, R. F., & Daly, K. (1980). Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of Autism and Developmental Disorders*, *10*(1), 91–103.

Sigman, M., & McGovern, C. W. (2005). Improvement in cognitive and language skills from preschool to adolescence in autism. *Journal of Autism and Developmental Disorders*, *35*(1), 15–23.

Sparrow, S., Cicchetti, D., & Balla, D. (2005). *Vineland adaptive behaviour scales* (2nd ed.). Circle Pines, MN: American Guidance Service.

Szatmari, P., Bryson, S. E., Boyle, M. H., Streiner, D. L., & Duku, E. (2003). Predictors of outcome among high functioning children with autism and Asperger syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *44*(4), 520–528. doi:10.1111/1469-7610.00141

Tager-Flusberg, H., & Joseph, R. M. (2003). Identifying neurocognitive phenotypes in autism. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1430), 303–314. doi:10.1098/rstb.2002.1198

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. Volkmar, R. Paul, A. Klin & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 335–364). Hoboken, NJ: John Wiley & Sons, Inc.

Tager-Flusberg, H. (2006). Defining language phenotypes in autism. *Clinical Neuroscience Research*, *6*(3–4), 219. doi:10.1016/j.cnr.2006.06.007

Tager-Flusberg, H., & Caronna, E. (2007). Language disorders: Autism and other pervasive developmental disorders. *Pediatric Clinics of North America*, *54*(3), 469–481.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research*, *6*(6), 468–478. doi:10.1002/aur.1329

Tager-Flusberg, H. (2015). Defining language impairments in a subgroup of children with autism spectrum disorder. *Science China. Life Sciences*, *58*(10), 1044–1052. doi:10.1007/s11427-012-4297-8

Tager-Flusberg, H., Plesa Skwerer, D., Joseph, R. M., Brukilacchio, B., Decker, J., Eggleston, B., ... Yoder, A. (2016). Conducting research with minimally verbal participants with autism spectrum disorder. *Autism*, *21*(7), 852–861. doi:10.1177/1362361316654605

Tek, S., Mesite, L., Fein, D., & Naigles, L. (2013). Longitudinal analyses of expressive language development reveal two distinct language profiles among young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*(1), 75–89. doi:10.1007/s10803-013-1853-4

Thurm, A., Lord, C., Lee, L. C., & Newschaffer, C. (2007). Predictors of language acquisition in preschool children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *37*(9), 1721–1734.

Tomblin, J. B., Zhang, X., Buckwalter, P., & O'Brien, M. (2003). The stability of primary language disorder: Four years after kindergarten diagnosis. *Journal of Speech, Language, and Hearing Research*, *46*(6), 1283–1296.

Ukoumunne, O. C., Wake, M., Carlin, J., Bavin, E. L., Lum, J., Skeat, J., ... Reilly, S. (2012). Profiles of language development in pre-school children: A longitudinal latent class analysis of data from the Early Language in Victoria Study. *Child: Care, Health & Development*, *38*(3), 341–349. doi:10.1111/j.1365-2214.2011.01234.x

Whitehouse, A. J. O., Barry, J. G., & Bishop, D. V. M. (2008). Further defining the language impairment of autism: Is there a specific language impairment subtype? *Journal of Communication Disorders*, *41*(4), 319–336.

Wilczynski, N. L., & Haynes, R. B. (2004). Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: An analytic survey. *BMC Medicine*, *2*, 23. doi:10.1186/1741-7015-2-23

Wing, L., Leekam, S. R., Libby, S. J., Gould, J., & Larcombe, M. (2002). The diagnostic interview for social and communication disorders: Background, inter-rater reliability and clinical use. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *43*(3), 307–325.

Wodka, E., Mathy, P., & Kalb, L. (2013). Predictors of phrase and fluent speech in children with autism and severe language delay. *Pediatrics*, *131*(4), e1128–e1134. doi:10.1542/peds.2012-2221

World Health Organization. (2010). *International statistical classification of diseases and related health problems* (10th revision, Vol. 2). Geneva, Switzerland: World Health Organization.