Cochrane Library

Cochrane Database of Systematic Reviews

Long-term prognosis of low language proficiency in children (Protocol)

Hagen ÅM, Rogde K, Lervåg A, Melby-Lervag M, Norbury C

Hagen ÅM, Rogde K, Lervåg A, Melby-Lervag M, Norbury C.
Long-term prognosis of low language proficiency in children (Protocol).
Cochrane Database of Systematic Reviews 2023, Issue 1. Art. No.: CD015268.
DOI: 10.1002/14651858.CD015268.

www.cochranelibrary.com

TABLE OF CONTENTS

[Prognosis Protocol]

Long-term prognosis of low language proficiency in children

Åste M Hagen[1], Kristin Rogde[1], Arne Lervåg[2], Monica Melby-Lervag[1], Courtenay Norbury[3]

[1]Department of Special Needs Education, University of Oslo, Oslo, Norway. [2]Department of Education, University of Oslo, Oslo, Norway. [3]Division of Psychology & Language Sciences, Universtiy College London, London, UK

Contact: Åste M Hagen, ammhage@isp.uio.no.

Editorial group: Cochrane Developmental, Psychosocial and Learning Problems Group.
Publication status and date: Edited (no change to conclusions), published in Issue 1, 2023.

Citation: Hagen ÅM, Rogde K, Lervåg A, Melby-Lervag M, Norbury C. Long-term prognosis of low language proficiency in children (Protocol). *Cochrane Database of Systematic Reviews* 2023, Issue 1. Art. No.: CD015268. DOI: 10.1002/14651858.CD015268.

ABSTRACT

Objectives

This is a protocol for a Cochrane Review (prognosis). The objectives are as follows:

Primary objectives

Population	Children between four and eight years of age, diagnosed with a language disorder, or determined to have low language proficiency before age eight.
Intervention	Not applicable, this is a review of observational studies and will not include active interventions.
Comparator	The comparison group will be peers or siblings with typically developing language skills, that is, no identified language disorder or low language proficiency.

Outcome	We will include the following outcomes if measured when participants are at least 12 years old (adolescents and adults):
	• Proximal outcomes in language and literacy (outcomes within the same domain as the original assessments: omnibus tests of expressive, receptive (including listening comprehension), total language, vocabulary, grammar, narrative or expository discourse, clinical markers, such as non-sense word repetition, sentence repetition, or both)
	• World Health Organization quality of life outcomes across five domains:
	  ○ physical (including general health, sleep and energy, sexual health);
	  ○ psychological (including mental health, self-esteem, memory, learning, and concentration);
	  ○ independence (including activities of daily living, occupational outcomes, dependence on medicinal and non-medicinal drugs and supports, independent living);
	  ○ social relationships (including friendships, romantic relationships, parenthood, peer problems, and anti-social behaviour);
	  ○ environment (including academic outcomes, work satisfaction, financial resources, societal participation in leisure/community activities, safety).

Timing	Studies must have traced the individuals into adolescence, or transition to adulthood or adulthood, or both, thus, when participants are 12 years and older.

Setting	Identification, or assessment and diagnosis using standard diagnostic algorithms. Precise measures and cutoffs on standardised tests will vary from study to study. For instance, Tomblin 1997 used the diagnostic algorithm of < -1.25 standard deviations on two or more language composite scores; and in Norbury 2016, language disorder was defined as scores of −1.5 standard deviations or below on two of five language composites in the absence of intellectual disability, existing medical diagnosis, or both. Identification by speech-language pathologists, educational psychology services, or in the context of research (population studies). Identification includes measures of language, such as vocabulary, grammar, morphosyntax, narrative. It can be expressive measure(s), receptive measure(s), or both. We will exclude identification made on the basis of tests of phonology/speech production, pragmatics, reading, or working memory only. We will include large-scale population studies in which low language proficiency may be determined by a cut-off of at least -1 standard deviation on at least one standard test of language to assess longitudinal relationships between single measures and quality of life outcomes in large community samples (Beitchman 2014; Caspi 2016; Thornton 2021).

The main objective of this review is to assess the long-term prognosis of an early language disorder or low language proficiency (LLP) for children aged four to eight years at baseline, and from age 12 years and up at follow-up, in areas of language and literacy, and broad quality of life outcomes in physical, psychological, independence, social relationships, and environment outcomes. This will include measures of mental and physical health, academic outcome, employment status, financial resources, and societal participation.

We will ask the following research questions: 1) To what extent do children with LLP age four to eight years show higher risk for persistent difficulties with language and literacy into adolescence and adulthood? 2) To what extent do children age four to eight years, with LLP, experience higher risk for poor quality of life across five domains of physical, psychological, independence, social relationships, and environment well-being in adolescence, and adulthood (WHO 2012)?

Secondary objectives

Secondary objectives are: 1) to understand how severity of early language problems affects long-term prognosis and quality of life, and 2) to identify gaps in the extant research. For instance, while the indicative sample of papers consistently report academic and employment outcomes, there is little evidence regarding physical or medical health outcomes.

Investigation of sources of heterogeneity between studies

We expect that there will be substantial heterogeneity between the included studies on the following variables:

- Diagnostic criteria
- Severity of language impairment
- Method of ascertainment (population study versus recruitment from special schools or clinics)
- Year of publication
- Age of outcome measurement
- Inclusiveness of non-verbal IQ (Specific Language Impairment versus Developmental Language Disorder)
- Stability of schooling (e.g. special school consistently versus changing between special school and mainstream classrooms)
- Literacy skills

BACKGROUND

Description of the health condition and context

This prognostic review will examine long-term outcomes of having low language proficiency (LLP) as a child. These children have been described using a multitude of labels and diagnostic criteria, and in many cases, language disorders go undiagnosed and untreated (Bishop 2010; McGregor 2020). Thus, this review will adopt broad inclusion criteria in order to capture all relevant studies. The target population is children who have been identified as having LLP between the ages of four and eight years old. This may include children with a clinical diagnosis of language disorder, those who are receiving speech-language pathology or special education services for poor language, children who screen positive for language disorder in clinical or educational trials, or children identified in population cohort studies on the basis of standardised test scores (cut-off scores may range from -1 standard deviation (SD) below the normative mean to -2 SD below on one or more standardised measures of language). The CATALISE consortium indicated that a diagnosis was only warranted in cases of persistent impairment, and noted there is variation in early language development, which means many very young children may experience delays in their language development that resolve without any additional support (Bishop 2016; Bishop 2017a). Therefore, we will set the lower age limit for our population of interest at age four years, and upper age limit at eight years. This is because some children are not identified, or do not get a diagnosis until they start school, and there is international variation in the age that children start school. In some countries, formal school doesn't start until age seven. Thus, setting the upper age limit at eight years will ensure that we capture most children identified with LLP across different countries and educational systems.

Problems with language are quite common in children. Prevalence estimates of developmental language disorder (including children formerly identified as having 'specific language impairment') in the absence of other biomedical conditions are approximately 7% in English-speaking countries (Norbury 2016; Tomblin 1997). This increases to approximately 10% when children with other neurodevelopmental conditions affecting language, such as autism or intellectual disabilities, are included (Norbury 2016), and may vary according to community socioeconomic factors (Norbury 2021). Language deficits can affect children in different ways. A child may experience problems with expressive language, problems understanding language, or a combination of both. Therefore, the term language disorder refers to challenges in understanding or producing language that affect a child's everyday functioning (Bishop 2016). The term language difficulties is used in a broader way, and often refers to children who are considered at risk of a language disorder, but have not yet been diagnosed (e.g. Dockrell 1998; Hagen 2017).

Different labels have been used throughout the years, and across professional groups and diagnostic manuals (Bishop 2017a). See Table 1 for examples of labels used. Specific language impairment has been a common term in the research literature (Leonard 2014); other terms are: primary language impairment, language delay, developmental language disorder, and developmental dysphasia. This inconsistency in terms is confusing, but what is more serious is that there has been little agreement on which criteria to use, and how to classify language disorders. In 2016, Bishop and colleagues published the first paper from the CATALISE study, the goal of which was to raise awareness of these issues, and create consensus in the field for criteria and terminology (Bishop 2016). Almost 60 experts, representing ten different relevant disciplines, participated in the study. They rated and discussed statements, based on papers from a special issue in a journal on this topic. Based on several rounds of rating, discussions, and revisions of statements, they reached a consensus on 27 statements that specified criteria and classification of language disorders (Bishop 2016). This important work was followed by a study to agree on a common terminology for children's language problems (Bishop 2017b). Using the same panel and methodology as the first phase of the CATALISE study (Bishop 2016), they reached consensus on terminology. They agreed that language disorder refered to a profile of problems that led to functional impairment in everyday life; that is, problems with language of a nature that had significant impact on children's everyday lives (Bishop 2017b). They further agreed that developmental language disorder (DLD) refered to language disorders that were not associated with any known biomedical aetiology (Bishop 2017b).

The Diagnostic and Statistical Manual of Mental Disorders - 5th edition (DSM-5 (APA 2013)) uses the term language disorder. However, the International Statistical Classification of Diseases and Related Health Problems - 11th edition (ICD-11) used developmental language disorder (WHO 2022). The ICD-11 describes four subcategories: DLD with impairment of receptive and expressive language, DLD with mainly impairment of expressive language, DLD with mainly impairment of pragmatic language, and DLD with other specified language impairment.

In this review, we will use the term low language proficiency to refer to children who are either diagnosed with a language disorder, or identified in population cohort studies to have lower language skills. The children may be identified with a cut-off score of at least -1 SD below the mean on one or more standardised measures of language. Diagnosis or assessment will be done by an educational psychology service or in the context of research (the setting). Since we are interested in long-term consequences, we will only include studies that follow participants longitudinally into youth or adulthood. Specifically, we are interested in outcomes that are assessed when children are at least 12 years old.

Health outcomes

Children with LLP may experience persistent difficulties in areas of language and literacy, which may ultimately affect quality of life when they are adults (Dubois 2020; Heckman 2006). Therefore, we will organise the outcomes of interest according to the World Health Organization framework for measuring quality of life (QoL (WHO 2012)). In Table 2, we provide indicative examples of outcome domains, including QoL domains, and their measurement in studies of adolescents and adults with a history of LLP.

Since we are interested in prognosis, we will include only long-term, and not concurrent outcomes in the review. For the prediction horizon, the identification of LLP in included studies must be between the ages of four and eight years, and we will extract follow-up information on the outcomes from the studies when participants are at least 12 years old. We chose 12 years of age, because it signals the beginning of secondary school and adolescence in most countries, and is an important outcome point for education. If possible, we will consider developmental changes

between adolescence and adulthood in studies that report both (Beitchman 2014).

Why it is important to do this review

Prevalence studies have estimated that on average, 7% to 10% of children aged five years to six years have LLP that makes functioning well in school and social settings difficult (Norbury 2016; Tomblin 1997). Children with these challenges are at increased risk of problems in adolescence in the areas of language and literacy (Stothard 1998), and social relationships (Wadman 2011), and are twice as likely as peers with age-appropriate language to experience adverse mental health conditions (Yew 2013). A quote from a young person with LLP contextualises the problems these students may experience both academically and socially: "Nothing really helped me to do better in school. Because I was shy and quiet, sometimes teachers didn't even notice that I was in the classroom. When my mother complained to the [secondary] school about the other students' behaviour towards me, the teachers did nothing to help me. In Year 10 and Year 11, when my attendance was very poor, they were just calling my mum to put the responsibility on her. I preferred to do anything else than going to school." (Palikara 2009). Palikara 2009 followed 64 young people diagnosed with SLI as children, and conducted interviews with them when they were in late adolescents or young adulthood (participants were older than 16 years of age). Of these, 37 were attending a college of further education at the time of the interviews. These college students expressed challenges with the increased academic demands. For instance, one student said: "I had some problems with coursework. It's harder than GCSE. [For GCSE], the work that you had to produce was much shorter. It's longer here. The questions are harder and a lot more research has to be done." (Palikara 2009). About one in six of the interviewees reported that they had encountered difficult peers or had been bullied.

Several studies find that as young adults, people with LLP do not achieve the same level of educational attainment as peers (Conti-Ramsden 2009), and experience challenges with social relationships (Howlin 2000), mental health (Brownlie 2016), and employment status (Heckman 2006; Johnson 2010). Some studies report poorer prognosis as a result of having LLP (e.g. Clegg 2005; Conti-Ramsden 2012; Mawhood 2000), whereas others show a mixed picture (Beitchman 2001; Botting 2016; Clegg 2012). Many studies have found prolonged differences between those with early LLP compared with controls in areas such as academic self-concept, friendships, behaviour, and employment (Conti-Ramsden 2012; Lindsay 2012). In contrast to this, other studies have reported more positive long-term outcomes for individuals with childhood language disorders (Clegg 2012; Records 1992). However, as pointed out by Joffe and Nippold, long-term outcomes are variable, complex, and multifaceted, and several factors can determine risk and resilience (Joffe 2012). There seems to be a general concern about the prognosis for those with childhood LLP, however, studies are somewhat mixed regarding longer-term outcomes, and we do not have a good enough understanding of the consequences of having such challenges early in life. Therefore, it is also challenging to tailor interventions in different phases of life to prevent additional problems.

Based on our knowledge and search in the research field, we only know of two systematic reviews that attempted to summarise the longer-term effects of having problems in the area of language (Dubois 2020; Fisher 2017). The review by Dubois and colleagues includes 15 studies with a total of four cohorts of young adults (ages 18 to 34 years) with DLD (Dubois 2020). They portrayed outcomes in three life areas: education, employment, and independent living. The review provided important insight into the risk of young adults with DLD in these three areas, however, it had some serious limitations. It did not include several cohorts that seem comparable to the included samples (e.g. The Iowa cohort (Tomblin 1997), and the Rutter cohort (Clegg 2005)). It did not examine potential sources of variation in outcomes between cohorts that were included. It conflated studies that examined DLD and studies with population cohorts in which language phenotyping was minimal (i.e. receptive vocabulary only). It also presented the results qualitatively, and did not report a meta-analysis. It did not summarise important results, like persistence of language disorder or mental health. Lastly, the prognosis studies in this field have variable quality, yet Dubois 2020 did not examine study quality in relation to the results in the different studies.

Fisher 2017 undertook a systematic review and meta-analysis of predictors of long-term outcomes among late-talking toddlers. They found that significant predictors of expressive-language outcomes were size of expressive vocabulary, receptive language skills, and socioeconomic status. This is an important contribution to the field, however, it limited outcomes to expressive language only, and only focused on very young children at baseline, i.e. toddlers. Since many of these late-talking toddlers tend to catch up with peers, it is also important to obtain knowledge about the long-term prognosis of children who are slightly older, and are more likely to have persistent language difficulties.

Thus, there seems to be a gap in the field that summarises the longer-term effects of LLP. Compared with related conditions, such as dyslexia and autism, the area of language disorders has been neglected. Although there are longitudinal studies that trace the development and investigate the prognosis of language disorders, there are only two that attempted to summarise these findings: one systematic literature review (Dubois 2020), and one systematic review and meta-analysis (Fisher 2017). Therefore, we currently have neither a good overview of the long-term prognosis of people with LLP, nor the moderators that might relate to a variation in outcomes. Our review will add to the extant literature by documenting a larger number of cohort studies, a broader range of outcomes that are recognised indicators of quality of life, and a broader age range of outcomes, to enable us to document potential age-related changes in outcomes. This will help us to identify gaps in the existing research base, and provide valuable information for families, practitioners, and policymakers on long-term outcomes of LLP.

OBJECTIVES

Primary objectives

Population	Children between four and eight years of age, diagnosed with a language disorder, or determined to have low language proficiency before age eight.
Intervention	Not applicable, this is a review of observational studies and will not include active interventions.
Comparator	The comparison group will be peers or siblings with typically developing language skills, that is, no identified language disorder or low language proficiency.

Outcome	We will include the following outcomes if measured when participants are at least 12 years old (adolescents and adults):
	• Proximal outcomes in language and literacy (outcomes within the same domain as the original assessments: omnibus tests of expressive, receptive (including listening comprehension), total language, vocabulary, grammar, narrative or expository discourse, clinical markers, such as non-sense word repetition, sentence repetition, or both)
	• World Health Organization quality of life outcomes across five domains:
	  ◦ physical (including general health, sleep and energy, sexual health);
	  ◦ psychological (including mental health, self-esteem, memory, learning, and concentration);
	  ◦ independence (including activities of daily living, occupational outcomes, dependence on medicinal and non-medicinal drugs and supports, independent living);
	  ◦ social relationships (including friendships, romantic relationships, parenthood, peer problems, and anti-social behaviour);
	  ◦ environment (including academic outcomes, work satisfaction, financial resources, societal participation in leisure/community activities, safety).

Timing	Studies must have traced the individuals into adolescence, or transition to adulthood or adulthood, or both, thus, when participants are 12 years and older.
Setting	Identification, or assessment and diagnosis using standard diagnostic algorithms. Precise measures and cutoffs on standardised tests will vary from study to study. For instance, Tomblin 1997 used the diagnostic algorithm of < -1.25 standard deviations on two or more language composite scores; and in Norbury 2016, language disorder was defined as scores of −1.5 standard deviations or below on two of five language composites in the absence of intellectual disability, existing medical diagnosis, or both. Identification by speech-language pathologists, educational psychology services, or in the context of research (population studies). Identification includes measures of language, such as vocabulary, grammar, morphosyntax, narrative. It can be expressive measure(s), receptive measure(s), or both. We will exclude identification made on the basis of tests of phonology/speech production, pragmatics, reading, or working memory only. We will include large-scale population studies in which low language proficiency may be determined by a cut-off of at least -1 standard deviation on at least one standard test of language to assess longitudinal relationships between single measures and quality of life outcomes in large community samples (Beitchman 2014; Caspi 2016; Thornton 2021).

The main objective of this review is to assess the long-term prognosis of an early language disorder or low language proficiency (LLP) for children aged four to eight years at baseline, and from age 12 years and up at follow-up, in areas of language and literacy, and broad quality of life outcomes in physical, psychological, independence, social relationships, and environment outcomes. This will include measures of mental and physical health, academic outcome, employment status, financial resources, and societal participation.

We will ask the following research questions: 1) To what extent do children with LLP age four to eight years show higher risk for persistent difficulties with language and literacy into adolescence and adulthood? 2) To what extent do children age four to eight years, with LLP, experience higher risk for poor quality of life across five domains of physical, psychological, independence, social relationships, and environment well-being in adolescence, and adulthood (WHO 2012)?

Secondary objectives

Secondary objectives are: 1) to understand how severity of early language problems affects long-term prognosis and quality of life, and 2) to identify gaps in the extant research. For instance, while the indicative sample of papers consistently report academic and

employment outcomes, there is little evidence regarding physical or medical health outcomes.

Investigation of sources of heterogeneity between studies

We expect that there will be substantial heterogeneity between the included studies on the following variables:

- Diagnostic criteria
- Severity of language impairment
- Method of ascertainment (population study versus recruitment from special schools or clinics)
- Year of publication
- Age of outcome measurement
- Inclusiveness of non-verbal IQ (Specific Language Impairment versus Developmental Language Disorder)
- Stability of schooling (e.g. special school consistently versus changing between special school and mainstream classrooms)
- Literacy skills

METHODS

Criteria for considering studies for this review

Studies will be included in this review if they meet the following criteria.

- Include a sample of children identified as having a language disorder or low language proficiency, using standard diagnostic algorithms (note: precise measures and cut-offs on standardised tests will vary from study to study) between age four and eight years.
- Identification includes measures of language, such as vocabulary, grammar, morphosyntax, and narrative. Cut-off scores on standardised tests will be at least -1 standard deviation (SD) below the normative mean. We will exclude identification made on the basis of tests of phonology/speech production, pragmatics, reading, or working memory only.
- Identification, or assessment and diagnosis is made by speech-language pathologists, educational professionals, clinical/educational psychologists, or researchers (population studies).
- They have assessed individuals at age 12 years or older, on at least one outcome measure of language (vocabulary, grammar, narrative, higher-order language, such as inferencing), or literacy (reading, spelling) skills, or quality of life (QoL) indices (or a combination) in any of the following five domains: physical, psychological, independence, social relationships, and environment well-being.

Types of studies

We will include published or unpublished reports of prospective or retrospective longitudinal studies following participants with low language proficiency (LLP). Studies must report scores on measures of language when participants were between four and eight years, and results on outcome(s) measured when participants were 12 years or older.

A study may or may not include a comparison group. The studies will have two main types of design, either 1) with a comparison group without LLP, or 2) no comparison group. In the first case, the results will be based on differences between groups. The comparison groups can include participants with typical development or other disorders, but no LLP. The groups will be matched on age. In the second design, with no comparison group, we will examine whether severity of LLP can predict long-term outcomes. These studies will most likely report language levels in participants with LLP on a standardised norm referenced test. In these cases, we will use the mean percentile of the sample with LLP as an indicator of severity of the language problems. In cases where data are reported on known characteristics of the population, such as employment, income, etc., results will be compared with this. We will also consider randomised controlled trials (RCTs), but will only extract data from the untreated control group.

We will exclude single-case studies.

Targeted population

The target population is children who were identified as having a language disorder or LLP between the ages of four and eight years. The children may have been identified with a language disorder by clinical or educational services, or they may have been identified as part of research, for instance population cohort studies. The cut-off score has to be at least -1 SD below the normative mean on at least one standardised test of language. We will exclude identification based on tests of phonology/speech production, pragmatics, reading, or working memory only.

We will include 'co-occurring disorders' as defined in CATALISE statement 9: “Co-occurring disorders are impairments in cognitive, sensorimotor or behavioural domains that can co-occur with developmental language disorder (DLD) and may affect pattern of impairment and response to intervention, but whose causal relation to language problems is unclear. These include attentional problems (attention deficit/hyperactivity disorder (ADHD)), motor problems (developmental co-ordination disorder (DCD)), reading and spelling problems (developmental dyslexia), speech problems, limitations of adaptive behaviour and/or behavioural and emotional disorders.” (Bishop 2017b). In these cases, we will consider comorbidity or co-occuring disorders as a moderator of outcome.

We will exclude studies that predominantly include participants with 'differentiating conditions', which according to the CATALISE framework, include “brain injury, acquired epileptic aphasia in childhood, certain neurodegenerative conditions, cerebral palsy, and oral language limitations associated with sensorineural hearing loss (Tomblin 2015), as well as genetic conditions, such as Down Syndrome" (Bishop 2017b). We will also exclude children with autism spectrum disorder, intellectual disability, or both (Harris 2013), because these conditions are commonly linked to genetic or neurological causes (Fitzgerald 2015; Shevell 2001), with the numbers of known aetiology increasing with advances in genetic methods (Bourgeron 2015; Fitzgerald 2015; Shevell 2001).

Types of outcomes to be predicted

We will report two types of outcomes:

Language and literacy. These are proximal measures related to a child’s initial identification and diagnosis.

- Omnibus tests of expressive, receptive (including listening comprehension), total language

- Vocabulary
- Grammar
- Narrative or expository discourse
- Clinical markers, such as nonsense word repetition, sentence repetition, or both

Quality of life (QoL (WHO 2012)): multiple measures across five domains of functioning (see Table 2 for example measures)

- Physical, including general health, sleep and energy, sexual health
- Psychological, including mental health, self-esteem, memory, learning and concentration; independence, including activities of daily living, dependence on medicinal and non-medicinal drugs and supports, independent living
- Level of independence, including mobility and work capacity
- Social relationships, including friendships, romantic relationships, parenthood, peer problems, and anti-social behaviour
- Environment, including academic outcomes, work satisfaction, financial resources, societal participation in leisure/community activities, safety

When a study reports multiple measures for an outcome, we will prioritise standardised and validated measures over researcher-made and non-validated measures, as these are considered more reliable. In the characteristics of included studies table, we will list all measures reported in the study, noting which we selected for analysis and why.

Since we are interested in prognosis, we will only include long-term, not concurrent outcomes. For the prediction horizon, the measurement of language skills, or diagnosis of language disorder had to be done before the children were eight years old, and follow-up information on the outcomes was extracted from the studies when children were at least 12 years old.

Since initial identification and diagnosis is based on language measures, the language and literacy measures are closely related to the identification criteria. Therefore, many of these data will be continuous and normed (see Table 2 for examples of measures). We may also find and include studies that include ordinal data. We anticipate that some QoL data, such as level of independence, may be ordinal or dichotomous. We will present results in a summary of findings table (see Table 3 for an example).

Search methods for identification of studies

Electronic searches

We will search PsycINFO using the strategy in Appendix 1. We will adapt this strategy for the databases listed below, using appropriate indexing terms and syntax:

- Cochrane Central Register of Studies (CENTRAL; current issue) in the Cochrane Library, which includes the Cochrane Developmental, Psychosocial and Learning Problems Specialized Register (from inception onwards);
- MEDLINE Ovid (1946 onwards);
- MEDLINE In-Process and Other Non-Indexed Citations Ovid (1946 onwards);
- MEDLINE Epub Ahead of Print Ovid (1946 onwards);
- Embase Ovid (1974 onwards);
- CINAHL Plus EBSCOhost (Cumulative Index to Nursing and Allied Health Literature; 1937 onwards);
- ClinicalTrials.gov (2000 onwards; clinicaltrials.gov/);
- ERIC EBSCOhost (1966 onwards);
- Education Abstracts (H.W. Wilson) EBSCOhost (1983 onwards);
- Education Database Proquest (1988 onwards);
- Linguistics and Language Behavior Abstracts ProQuest (LLBA; 1973 onwards);
- PsycINFO Ovid (1806 onwards);
- Scopus Elsevier (1970 onwards);
- Science Citation Index-Expanded Web of Science, Clarivate (1970 onwards);
- Social Sciences Citation Index Web of Science, Clarivate (1970 onwards);
- Conference Proceedings Citation Index-Science Web of Science, Clarivate (1990 onwards);
- Conference Proceedings Citation Index-Social Science and Humanities Web of Science, Clarivate (1990 onwards);
- Emerging Sources Citation Index Web of Science, Clarivate (2015 onwards);
- ProQuest Dissertations & Theses Global (1743 onwards);
- OpenGrey (1980 onwards; www.opengrey.eu);
- Google Scholar (scholar.google.com/). As Google Scholar does not have a limit on the number of hits, we will screen the first 500 references that are most relevant for our search (all available years).

We will not limit the searches by date, publication type, study type, publication status, or language. We will seek translation of non-English studies, using translate.google.com or a professional translation service, as appropriate. Before publication, we will conduct a search to ensure none of our included studies have been corrected or retracted.

Searching other resources

In addition to searching in electronic databases, we will also:

- search for relevant studies in the reference lists of included studies, and any relevant systematic reviews identified in our searches;
- post on list servers for the Society of Scientific Studies of Reading and The Society for Research on Learning Disorders;
- contact experts in the field to determine if they know of any additional eligible studies.

Data collection

Selection of studies

Two of the review authors (ÅMH and KR), will independently conduct the initial screening of titles and abstracts identified through the electronic searches, and remove clearly irrelevant articles.

We will obtain the full-text articles of all potentially relevant studies, and of those whose relevance cannot be determined from the abstract. Two authors (ÅMH and KR) will independently review all full-text articles for eligibility. We will report Kappa statistics. The two authors will resolve disagreements through discussion, or if

required, consultation with a third review author (MM-L). We will collate multiple reports of the same study, so the study, rather than the report, is the unit of interest. We will outline the study selection process in a PRISMA study flow diagram (Moher 2009).

Data extraction and management

We will code study variables that relate to contextual information (e.g. country, year of publication), methods (e.g. clinical sample, number of participants, type of comparison group, age at study onset and follow-up, diagnosis, type of outcome measures), and risk of bias (e.g. attrition, loss to follow-up, participation, and outcome measures). We will extract data from included studies using a data extraction form inspired by Brignell 2022. See Appendix 2 for details about study variables, and Appendix 3 for risk of bias assessment.

We will use the **ch**ecklist for critical **a**ppraisal and data extraction for systematic **r**eviews of prediction **m**odelling **s**tudies (CHARMS) to develop the form (Moons 2014). We will pilot the data extraction form on several studies, and make necessary edits. Two review authors (ÅMH and KR) will independently extract all data, and a third review author (MM-L) will check the data. The three authors (ÅMH, KR and MM-L) will resolve disagreements through discussion, or if required, consult with a fourth review author (AL or CN).

Assessment of risk of bias in included studies

We will assess the risk of bias in each included study to examine the quality of the studies. Adapted from the protocol of Brignell 2022, we will assess risk of bias by examining three main domains: participation, attrition, and outcome measurement. Two review authors (ÅMH and KR) will independently code all studies as high, moderate, low, or unclear risk of bias according to the coding scheme presented in Appendix 3. The coding scheme is an adaption of the QUIPS tool (Hayden 2013), and includes several items for each of the three domains of participation, attrition, and outcome. Because we are not investigating prognostic factors, we excluded the three domains (prognostic factor; measurement, study confounding; statistical analysis and reporting) from the QUIPS tool from our risk of bias assessment.

The two authors (ÅMH and KR) will discuss discrepancies after their independent assessment, and if needed, consult with one of the other authors in the team (MM-L, AL, or CN). When discrepancies are resolved, they will rate each risk of bias category as high, medium, low, or unclear risk of bias.

If information needed to code risk of bias is not available in a study report, we will attempt to contact the authors by email to ask for further information. If the authors are not available, or are unable or unwilling to provide the needed information, we will code the risk of bias as unclear.

Measures of association or predictive performance measures to be extracted

In the planned review, it is likely that the studies will use a variety of outcome measures. As outlined above, we have organised these multiple outcomes as proximal to initial identification (language, literacy, or both), and outcomes related to five domains of quality of life (WHO 2012). These outcomes might be related to several measures of association or predictive performance measures to be extracted, or to the global measures of quality of life, which implicate both performance and the range of supports available. We are interested in whether there are differences between those identified as having LLP between the ages of four and eight and a comparison group without LLP, in outcomes from age 12 to adulthood on both kinds of outcomes. Thus, the main predictor here will be binary, i.e. LLP or no LLP. However, the criteria for LLP will vary across studies, and we will take this into account as a moderator.

Dealing with missing data

To deal with missing data, we will first contact authors to try to get the data needed to estimate an effect size. If data are reported in terms of significance tests or in formats other than what we are looking for, we will transform them. To extract data from plots, we will use the WebPlot Digitizer (Rohatgi 2022). If data are reported in significance tests (z, t, one way ANOVA), we will convert them to either Pearson's r or Cohen's d, depending on the nature of the study. If data are reported in either odds ratio or bivariate regression analysis, we will transform them to r or d (Borenstein 2021). Missing data in the meta-analysis are not likely to be missing at random. Therefore, in line with Cochrane recommendations, we will impute the missing data with replacement values, and treat these as if they were observed (Higgins 2022).

Assessment of heterogeneity

We expect sources of heterogeneity to originate from the following:

- diagnostic criteria;
- severity of language impairment;
- method of ascertainment (population study versus recruitment from special schools or clinics);
- year of publication;
- age of outcome measurement;
- inclusiveness of non-verbal IQ (specific language impairment versus developmental language disorder);
- stability of schooling or educational pathways (e.g. special school consistently versus changing between special school and mainstream classrooms);
- literacy skills.

To get an overview of the true variation between studies, we will calculate Tau^2, I^2, and prediction intervals. We will use Tau^2 to examine the magnitude of the variations in effect sizes between studies. When using Cohen's d, Tau^2 is on the same metric as the effect itself. Its interpretation is: If a mean effect is zero and Tau^2 is 0.3, a rough estimate of the range of true study effects is the mean effect ± 2 SD (two times Tau^2), that is, d from 0.6 to 0.6. I^2 was used to investigate the proportion of variation in the effect sizes that reflected true variation rather than a sampling error (Borenstein 2021).

Quantifying heterogeneity between studies might also inform the potential performance in validation of risk prediction models (Chen 2020; Debray 2017).

Assessment of reporting deficiencies

We will use funnel plots to examine publication bias, Eggers test and Pet/pees methods (Debray 2018), since these methods currently seem to perform best under most circumstances (Carter 2019). We will generate funnel plots separately for each outcome with at

least 10 studies. We will look for asymmetry in the funnel plots to determine publication bias. If funnel plots resemble symmetrical funnels, we can assume the absence of bias.

Data synthesis

Data synthesis and meta-analysis approaches

It is likely that the included studies will use different types of estimates. The majority of studies are likely to use correlational coefficients, i.e. correlations between language skills and later outcomes. However, some are also likely to use Cohen's d for group differences between language group and control group. We will first transform correlations to Fishers z to get a normal distribution, and then back to r for presentation purposes. The methods' literature does not recommend transforming all statistics to a common metric, for instance Cohen's d (Borenstein 2021). Therefore, if we get at least two studies, we will calculate separate overall effect sizes, based on the metric that is used in the original paper. If this is not possible, we will use Cohen's d as an effect size estimate, and transform those that do not report this. Finally, as mentioned above, some studies are also likely to report results in different formats or effect sizes, for instance t-tests, ANOVA, odds ratio, or regression coefficients. If possible, we will transform these to standardised group differences or correlations, depending on the design of the study.

We will undertake the statistical analyses in Robumeta, R package (Fisher 2015), and use a random-effects model, since there are likely to be true differences between effect sizes, not only sampling error. We will summarise the results with the pooled estimate (the average prognostic factor effect), its 95% confidence interval (CI), the estimates of I^2 and Tau^2 (heterogeneity), and a 95% prediction interval for the prognostic effect in a single population. Robumeta will also correct for dependencies in the data. It is likely that dependencies will be present on several levels, for example one study might have more than one outcome, several studies might be conducted by the same lab, and so on.

If we are unable to calculate a mean effect size for some outcomes, we will evaluate these studies qualitatively, and provide a narrative report. In this case, we will qualitatively code the evidence according to the following schema, which is adapted from Neligan 2021: if multiple studies (at least two) with low risk of bias show evidence of a strong relationship between baseline (i.e. detected language disorder or LLP) and outcome, we will code as a 'strong relationship between baseline and outcome'. We will code a 'moderate relationship between baseline and outcome' when this is consistent across multiple (at least two) high risk of bias studies, or one low risk of bias study. We will code a 'limited relationship between baseline and outcome' if only one high risk of bias study is available. If there are inconsistent findings across studies, we will code an 'unclear relationship between baseline and outcome', and finally, if we do not detect a relationship, we will code it as 'no relationship between baseline and outcome'.

It is important to distinguish population studies that aim to identify children with LLP, and those in which language phenotyping follows clinically accepted protocols from population studies that include language as part of their descriptive battery, but primary study questions are not about language. Therefore, we will undertake meta-analyses for the prospective studies. For the retrospective studies, we will present meta-coefficients for each study separately, if possible; if not, we will provide a narrative summary.

Assessing quality of evidence and summary of findings tables

Adapted from Brignell 2022, we will judge the overall quality of evidence for all outcomes using the GRADE approach (Iorio 2015). We will rate the quality of the evidence for risk of bias, inconsistency, indirectness, imprecision, publication bias, prognostic effect, and dose-response gradient (Appendix 4). We will classify the quality of the evidence as high, moderate, low, or very low (Appendix 5). We will present a summary of our results in a summary of findings table; see Table 3 for an example.

Subgroup analysis and investigation of heterogeneity

Subgroups for analysis will be the following variables:

- Diagnostic criteria: the cut-off score(s) used as inclusion criteria might differentially relate to the long-term prognosis of low language proficiency; therefore, we will include language severity as a moderator of outcome. We plan to conduct subgroup analyses on the categories of SDs below the mean: -1, -1.25, -1.5, and -2.
- Severity of language impairment: we aim to examine whether severity of language impairment reflects the long-term prognosis of language proficiency. By examining differences in standard deviation units, we can examine the severity of impairment as a continuous moderator.
- Method of ascertainment: how participants are selected for a sample may differ from study to study. That means that study samples may be systematically different from each other, so we will examine the method of ascertainment using the following categories:
  - population study;
  - recruitment from special schools or clinics.
- Year of publication:
  - Studies older than 1980;
  - Studies 1980 to 2000;
  - Studies 2000 to current.
- Age of outcome measurement: long-term prognosis of language proficiency may differ in the timing of the outcome measurements, therefore, we will undertake subgroup analyses using the categories:
  - adolescence (age 12 to 18);
  - transition (age 19 to 24);
  - adulthood (age 25 and up).
- Inclusiveness of non-verbal IQ: samples may differ in the criteria related to inclusiveness of nonverbal IQ, therefore, we will undertake subgroup analyses on:
  - non-verbal IQ included (e.g. specific language impairment);
  - non-verbal IQ excluded (e.g. developmental language disorder).
- Educational pathways:
  - mainstream only;
  - special class or school only;
  - mainstream then special school;
  - variable (back and forth).
- Literacy skills: whether the participants are characterised with literacy problems or not may be differently associated with long-

term prognosis in language proficiency, so we will undertake subgroup analyses on the categories:
  - reported literacy problems;
  - no reported literacy problems.
- Presence of known, co-occuring conditions:
  - ADHD;
  - dyslexia;
  - motor deficits;
  - other learning disorders.

We will use meta-regression with both continuous variables and with dummy coding for categories (Fisher 2015).

Sensitivity analysis

We will use a 'one study removed' sensitivity analysis to determine whether the overall estimates between studies are influenced by outlier studies. This is done by performing a series of meta-analyses that leave out one by one of the studies in the original meta-analysis. By removing one study, you can observe how much the results change, and how that particular study influences your results (Borenstein 2021). Using this approach, we will consider effect sizes that fell outside the 95% confidence interval of the average effect size to be outliers. We will set correlations between different outcomes within the same domain to 0.4, based on previous studies of correlations between the outcome tasks of interest. However, we will conduct a sensitivity analysis to ensure that the results were robust across several levels of correlations (Fisher 2015). We will also use sensitivity analysis to examine the impact from different levels of dependencies in the data.

ACKNOWLEDGEMENTS

The authors would like to thank the editorial team of Cochrane Developmental, Psychosocial and Learning Problems for the support and advice during the development of this protocol. A special thank to Margaret Anderson for help with the search strategy and the editors for helpful comments.

The Developmental, Psychosocial and Learning Problems Editorial Team is grateful to the following peer reviewers for their time and comments: Dorothy Bishop, Professor of Developmental Neuropsychology, University of Oxford (UK); Maria Luisa Lorusso, PhD, Unit of Neuropsychology of Developmental Disorders, Scientific Institute IRCCS E Medea, Bosisio Parini (Italy); Genna White, parent of a child with developmental language disorder and speech and language therapist (UK); and Robert Wolff, Kleijnen Systematic Reviews Ltd (UK). They would also like to thank Victoria Pennick for copyediting the review.

REFERENCES

Additional references

APA 2013

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

Baddeley 1994

Baddeley A. Working memory: the interface between memory and cognition. In: Schacter DL, Tulving E, editors(s). Memory Systems 1994. London, UK: MIT Press, 1994:351–67.

Beitchman 2001

Beitchman JH, Wilson B, Johnson CJ, Atkinson L, Young A, Adlaf E, et al. Fourteen-year follow-up of speech/language-impaired and control children: psychiatric outcome. *Journal of the American Academy of Child & Adolescent Psychiatry* 2001;**40**(1):75-82. [DOI: 10.1097/00004583-200101000-00019] [PMID: 11195567]

Beitchman 2014

Beitchman JH, Brownlie EB, Bao L. Age 31 mental health outcomes of childhood language and speech disorders. *Journal of the American Academy of Child and Adolescent Psychiatry* 2014;**53**(10):1102-10.e8. [DOI: 10.1016/j.jaac.2014.07.006] [PMID: 25245354]

Bishop 2010

Bishop DV. Which neurodevelopmental disorders get researched and why? *PLoS One* 2010;**5**(11):e15112. [DOI: 10.1371/journal.pone.0015112] [PMCID: PMC2994844] [PMID: 21152085]

Bishop 2016

Bishop DV, Snowling MJ, Thompson PA, Greenhalgh T, Catalise Consortium. CATALISE: a multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLoS One* 2016;**11**(7):e0158753. Erratum in: PLoS; 2016; 11(12):e0168066. DOI: 10.1371/journal.pone.0168066. [DOI: 10.1371/journal.pone.0158753] [PMCID: PMC4938414] [PMID: 27392128]

Bishop 2017a

Bishop DV. Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *International Journal of Language & Communication Disorders* 2017;**52**(6):671-80. [DOI: 10.1111/1460-6984.12335] [PMCID: PMC5697617] [PMID: 28714100]

Bishop 2017b

Bishop DV, Snowling MJ, Thompson PA, Greenhalgh T, Catalise2 Consortium. Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: terminology. *Journal of Child Psychology and Psychiatry* 2017;**58**(10):1068-80. [DOI: 10.1111/jcpp.12721] [PMCID: PMC5638113] [PMID: 28369935]

Borenstein 2021

Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to Meta-Analysis. 2nd edition. Chichester (UK): John Wiley & Sons, Ltd, 2021. [ISBN: 978-1-119-55835-4]

Botting 2016

Botting N, Durkin K, Toseeb U, Pickles A, Conti-Ramsden G. Emotional health, support, and self-efficacy in young adults with a history of language impairment. *British Journal of Developmental Psychology* 2016;**34**(4):538-54. [DOI: 10.1111/bjdp.12148] [PMCID: PMC5082521] [PMID: 27226087]

Bourgeron 2015

Bourgeron T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nature Reviews Neuroscience* 2015;**16**:551-63. [DOI: 10.1038/nrn3992]

Brignell 2022

Brignell A, Harwood RC, May T, Woolfenden S, Montgomery A, Iorio A, et al. 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] [PMCID: PMC9516883 (available on 2023-09-28)] [PMID: 36169177]

Brownlie 2004

Brownlie EB, Beitchman JH, Escobar M, Young A, Atkinson L, Johnson C, et al. Early language impairment and young adult delinquent and aggressive behavior. *Journal of Abnormal Child Psychology* 2004;**32**(4):453-67. [DOI: 10.1023/B:JACP.0000030297.91759.74] [PMID: 15305549]

Brownlie 2007

Brownlie EB, Jabbar A, Beitchman JH, Vida R, Atkinson LR. Language impairment and sexual assault of girls and women: findings from a community sample. *Journal of Abnormal Child Psychology* 2007;**35**(4):618-26. [DOI: 10.1007/s10802-007-9117-4] [PMID: 17340176]

Brownlie 2016

Brownlie EB, Bao L, Beitchman J. Childhood language disorder and social anxiety in early adulthood. *Journal of Abnormal Child Psychology* 2016;**44**(6):1061-70. [DOI: 10.1007/s10802-015-0097-5] [PMID: 26530522]

Carter 2019

Carter EC, Schönbrodt FD, Gervais WM, Hilgard J. Correcting for bias in psychology: a comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science* 2019;**2**(2):115-44. [DOI: 10.1177/2515245919847196]

Caspi 2016

Caspi A, Houts RM, Belsky DW, Harrington H, Hogan S, Ramrakha S, et al. Childhood forecasting of a small segment of the population with large economic burden. *Nature Human Behaviour* 2016;**1**:0005. [DOI: 10.1038/s41562-016-0005] [PMCID: PMC5505663] [PMID: 28706997]

Chen 2020

Chen L. Overview of clinical prediction models. *Annals of Translational Medicine* 2020;**8**(4):71. [DOI: 10.21037/atm.2019.11.121] [PMCID: PMC7049012] [PMID: 32175364]

Clegg 2005

Clegg J, Hollis C, Mawhood L, Rutter M. Developmental language disorders – a follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry* 2005;**46**(2):128-49. [DOI: 10.1111/j.1469-7610.2004.00342.x] [PMID: 15679523]

Clegg 2012

Clegg J, Ansorge L, Stackhouse J, Donlan C. Developmental communication impairments in adults: outcomes and life experiences of adults and their parent. *Language, Speech, and Hearing Services in Schools* 2012;**43**(4):521-35. [DOI: 10.1044/0161-1461(2012/11-0068)] [PMID: 22826372]

Conti-Ramsden 2009

Conti-Ramsden G, Durkin K, Simkin Z, Knox E. Specific language impairment and school outcomes. I: identifying and explaining variability at the end of compulsory education. *International Journal of Language & Communication Disorders* 2009;**44**(1):15-35. [DOI: 10.1080/13682820801921601] [PMID: 18608604]

Conti-Ramsden 2012

Conti-Ramsden G, Durkin K. Language development and assessment in the preschool period. *Neuropsychology Review* 2012;**22**(4):384–401. [DOI: 10.1007/s11065-012-9208-z] [PMID: 22707315]

Conti-Ramsden 2018

Conti-Ramsden G, Durkin K, Toseeb U, Botting N, Pickles A. Education and employment outcomes of young adults with a history of developmental language disorder. *International Journal of Language and Communication Disorders* 2018;**53**(2):237-55. [DOI: 10.1111/1460-6984.12338] [PMCID: PMC5873379] [PMID: 29139196]

Debray 2017

Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;**356**:i6460. [DOI: 10.1136/bmj.i6460] [PMID: 28057641]

Debray 2018

Debray TP, Moons KG, Riley RD. Detecting small-study effects and funnel plot asymmetry in meta-analysis of survival data: a comparison of new and existing tests. *Research Synthesis Methods* 2018;**9**(1):41-50. [DOI: 10.1002/jrsm.1266] [PMCID: PMC5873397] [PMID: 28975717]

Dockrell 1998

Dockrell J, Lindsay G. The ways in which speech and language difficulties impact on children's access to the curriculum. *Child Language Teaching and Therapy* 1998;**14**(2):117-33. [DOI: 10.1177/026565909801400201]

Dubois 2020

Dubois P, St-Pierre M-C, Desmarais C, Guay F. Young adults with developmental language disorder: a systematic review of education, employment, and independent living outcomes. *Journal of Speech, Language, and Hearing Research* 2020;**63**(11):3786-800. [DOI: 10.1044/2020_JSLHR-20-00127] [PMID: 33022192]

Fisher 2015

Fisher Z, Tipton E. Robumeta: an R-package for robust variance estimation in meta-analysis. March 2015. Available at arxiv.org/pdf/1503.02220.pdf. [DOI: 10.48550/arXiv.1503.02220]

Fisher 2017

Fisher EL. A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers. *Journal of Speech, Language, and Hearing Research* 2017;**60**(10):2935-48. [DOI: 10.1044/2017_JSLHR-L-16-0310] [PMID: 28915512]

Fitzgerald 2015

Fitzgerald TW, Gerety SS, Jones WD, van Kogelenberg M, King DA, McRae J. Large-scale discovery of novel genetic causes of developmental disorders: the deciphering developmental disorders study. *Nature* 2015;**519**:223-8. [DOI: 10.1038/nature14135]

Hagen 2017

Hagen ÅM, Melby-Lervåg M, Lervåg A. Improving language comprehension in preschool children with language difficulties: a cluster randomized trial. *Journal of Child Psychology and Psychiatry* 2017;**58**(10):1132-40. [DOI: 10.1111/jcpp.12762] [PMID: 28671266]

Harris 2013

Harris JC. New terminology for mental retardation in DSM-5 and ICD-11. *Current Opinion in Psychiatry* 2013;**26**(3):260-2. [DOI: 10.1097/YCO.0b013e32835fd6fb]

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] [PMID: 23420236]

Heckman 2006

Heckman JJ. Skill formation and the economics of investing in disadvantaged children. *Science* 2006;**312**(5782):1900-2. [DOI: 10.1126/science.1128898] [PMID: 16809525]

Higgins 2022

Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 6.3 (updated February 2022). Cochrane, 2022. Available from training.cochrane.org/handbook.

Howlin 2000

Howlin P, Mawhood L, Rutter M. Autism and developmental receptive language disorder – a follow-up comparison in early adult life. II: social, behavioural, and psychiatric outcomes. *Journal of Child Psychology and Psychiatry* 2000;**41**(5):561-78. [DOI: 10.1111/1469-7610.00643] [PMID: 10946749]

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] [PMID: 25775931]

Joffe 2012

Joffe VL, Nippold MA. Progress in understanding adolescent language disorders. *Language, Speech, and Hearing Services in Schools* 2012;**43**(4):438-44. [DOI: 10.1044/0161-1461(2012/12-0052)] [PMID: 22826370]

Johnson 2010

Johnson CJ, Beitchman JH, Brownlie EB. Twenty-year follow-up of children with and without speech-language impairments: family, educational, occupational, and quality of life outcomes. *American Journal of Speech-language Pathology* 2010;**19**(1):51-65. [DOI: 10.1044/1058-0360(2009/08-0083)] [PMID: 19644128]

Kohnert 2009

Kohnert K, Windsor J, Ebert KD. Primary or "specific" language impairment and children learning a second language. *Brain and Language* 2009;**109**(2-3):101-11. [DOI: 10.1016/j.bandl.2008.01.009] [PMCID: PMC2709989] [PMID: 18313136]

Leonard 2014

Leonard LB. Children with Specific Language Impairment. 2nd edition. Cambridge (MA): MIT Press, 2014.

Lindsay 2012

Lindsay G, Dockrell J. The relationship between speech, language and communication needs (SLCN) and behavioural, emotional and social difficulties (BESD). Research Report No.: DFE-RR247-BCRP6. London, UK: Department for Education; 2012 May. assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/219632/DFE-RR247-BCRP6.pdf.

Mawhood 2000

Mawhood L, Howlin P, Rutter M. 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* 2000;**41**(5):547-59. [DOI: 10.1111/1469-7610.00642] [PMID: 10946748]

McGregor 2020

McGregor KK. How we fail children with developmental language disorder. *Language, Speech, and Hearing Services in Schools* 2020;**51**(4):981-92. [DOI: 10.1044/2020_LSHSS-20-00003] [PMCID: PMC7842848] [PMID: 32755505]

Moher 2009

Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine* 2009;**6**(7):e1000097. [DOI: 10.1371/journal.pmed.1000097] [PMCID: PMC2707599] [PMID: 19621072]

Moons 2014

Moons KG, De Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Medicine* 2014;**11**(10):e1001744. [DOI: 10.1371/journal.pmed.1001744] [PMCID: PMC4196729] [PMID: 25314315]

Neligan 2021

Neligan A, Adan G, Nevitt SJ, Pullen A, Sander JW, Bonnett L, et al. Prognosis of adults and children following a first unprovoked seizure. *Cochrane Database of Systematic Reviews* 2021, Issue 1. Art. No: CD013847. [DOI: 10.1002/14651858.CD013847]

Njiokiktjien 1990

Njiokiktjien C. Developmental dysphasia: clinical importance and underlying neurological causes. *Acta Paedopsychiatric* 1990;**53**(2):126-37. [PMID: 1704678]

Norbury 2016

Norbury CF, Gooch D, Wray C, Baird G, Charman T, Simonoff E, et al. The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *Journal of Child Psychology and Psychiatry* 2016;**57**(11):1247-57. [DOI: 10.1111/jcpp.12573] [PMCID: PMC5082564] [PMID: 27184709]

Norbury 2021

Norbury C, Griffiths SL, Vamvakas G, Baird G, Charman T, Simonoff E, et al. Socioeconomic disadvantage is associated with prevalence of developmental language disorders, but not rate of language or literacy growth in children from 4 to 11 years: evidence from the Surrey Communication and Language in Education Study (SCALES); March 2021. Available at ssrn.com/abstract=3814832. [DOI: 10.31234/osf.io/nyv7z]

Palikara 2009

Palikara O, Lindsay G, Dockrell JE. Voices of young people with a history of specific language impairment (SLI) in the first year of post-16 education. *International Journal of Language & Communication Disorders* 2009;**44**:56–78. [DOI: 10.1080/13682820801949032]

Records 1992

Records N, Tomblin JB, Freese PR. The quality of life of young adults with histories of specific language impairment. *American Journal of Speech-language Pathology* 1992;**1**(2):44-53. [DOI: 10.1044/1058-0360.0102.44]

Rohatgi 2022 [Computer program]

WebPlotDigitizer. Rohatgi A, Version 4.6. Pacifica (CA): Ankit Rohatgi, 2022. Available at automeris.io/WebPlotDigitizer.

Shevell 2001

Shevell MI, Majnemer A, Rosenbaum P, Abrahamowicz M. Etiologic determination of childhood developmental delay. *Brain Development* 2001;**23**:228–35. [DOI: 10.1016/s0387-7604(01)00212-1]

Stothard 1998

Stothard SE, Snowling MJ, Bishop DV, Chipchase BB, Kaplan CA. Language-impaired preschoolers: a follow-up into adolescence. *Journal of Speech, Language, and Hearing Research* 1998;**41**(2):407-18. [DOI: 10.1044/jslhr.4102.407] [PMID: 9570592]

Sunderajan 2019

Sunderajan T, Kanhere SV. Speech and language delay in children: prevalence and risk factors. *Journal of Family Medicine and Primary Care* 2019;**8**(5):1642-6. [DOI: 10.4103/jfmpc.jfmpc_162_19] [PMCID: PMC6559061] [PMID: 31198730]

Thornton 2021

Thornton E, Patalay P, Matthews D, Bannard C. Does early child language predict internalizing symptoms in adolescence? An investigation in two birth cohorts born 30 years apart. *Child Development* 2021;**92**(5):2106-27. [DOI: 10.1111/cdev.13615] [PMID: 34213009]

Tomblin 1997

Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research* 1997;**40**(6):1245–60. [DOI: 10.1044/jslhr.4006.1245] [PMCID: PMC5075245] [PMID: 9430746]

Tomblin 2015

Tomblin JB, Harrison M, Ambrose SE, Walker EA, Oleson JJ, Moeller MP. Language outcomes in young children with mild to severe hearing loss. *Ear and Hearing* 2015;**36**:76S-91S. [DOI: 10.1097/AUD.0000000000000219]

Wadman 2011

Wadman R, Durkin K, Conti-Ramsden G. Close relationships in adolescents with and without a history of specific language impairment. *Language, Speech, and Hearing Services in Scools* 2011;**42**(1):41-51. [DOI: 10.1044/0161-1461(2010/10-0003] [PMID: 20679413]

Whitehouse 2009

Whitehouse AJ, Watt HJ, Line EA, Bishop DV. Adult psychosocial outcomes of children with specific language impairment, pragmatic language impairment and autism. *International Journal of Language & Communication Disorders* 2009;**44**(4):511-28. [DOI: 10.1080/13682820802708098] [PMCID: PMC2835860] [PMID: 19340628]

WHO 2012

World Health Organization. Programme on Mental Health: WHOQOL User Manual; 1998, revised 2012. Available at apps.who.int/iris/handle/10665/77932.

WHO 2022

World Health Organization. ICD-11. International Classification of Diseases 11th Revision. Available from icd.who.int/en (accessed 8 March 2022).

Winstanley 2018

Winstanley M, Durkin K, Webb RT, Conti-Ramsden G. Financial capability and functional financial literacy in young adults with developmental language disorder. *Autism & Developmental Language Impairments* 2018;**3**:2396941518794500. [DOI: 10.1177/2396941518794500] [PMCID: PMC6188087] [PMID: 30370337]

Yew 2013

Yew SG, O'Kearney R. Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments: meta-analyses of controlled prospective studies. *Journal of Child Psychology and Psychiatry* 2013;**54**(5):516-24. [DOI: 10.1111/jcpp.12009] [PMID: 23082773]

ADDITIONAL TABLES

Table 1. Examples of different labels used

Labels	Definition	Reference
Specific language impairment (SLI)	an unexplained and serious deficit in spoken language ability that affects their social and academic well-being	Leonard 2014
Primary language impairment	used as a synonym for SLI; refers to individuals with unexpectedly low language performance relative to otherwise typical development	e.g. Kohnert 2009
Language delay	a language delay in young children (usually ages 2 to 7 years); involves late development of language abilities relative to milestones for their age	e.g. Sunderajan 2019
Developmental dysphasia	a disorder involving difficulties speaking and understanding spoken words	e.g. Njiokiktjien 1990
Language problems/ language difficulties	used in a broader way than 'language disorder', refers to children who are considered at risk of a language disorder, but have not yet been diagnosed	e.g. Dockrell 1998; Hagen 2017
Language disorder	a profile of problems that leads to 'functional impairment in everyday life'	DSM-5 (APA 2013); see also Bishop 2017b

Table 1. Examples of different labels used *(Continued)*

Developmental language disorder (DLD) with subcategories[a]	language disorders that are not associated with any known biomedical aetiology	ICD-11 (WHO 2022); see also Bishop 2017b
Low language proficiency (LLP)	individuals who are either diagnosed with a language disorder or identified in population cohort studies to have lower language skills.	used in this review

DSM-5: Diagnostic and Statistical Manual of Mental Disorders 5th edition; ICD-11: International Statistical Classification of Diseases and Related Health Problems 11th Revision
[a]DLD with impairment of receptive and expressive language, DLD with mainly impairment of expressive language, DLD with mainly impairment of pragmatic language, and DLD with other specified language impairment

Table 2. Outcome domains with examples of measurements and studies

Outcome domains	Examples of indicators and measures	Reference
Language and Literacy	Language (e.g. WAIS - Verbal IQ; British Picture Vocabulary Scales; Expressive One Word Picture Vocabulary); reading (e.g. Wechsler Objective Reading Dimensions; Woodcock Reading Mastery Tests); non-verbal IQ (e.g. WAIS - non-verbal IQ)[a]	Clegg 2005; Conti-Ramsden 2018; Johnson 2010
QoL Domain I Physical	Pain and discomfort (e.g. % injury insurance claims); general health (e.g. % obese); sleep and rest	Caspi 2016
QoL Domain II Psychological	Thinking, learning, memory, (e.g. Doors & People (Baddeley 1994)); self-esteem; negative feeling (e.g. Social Interaction Anxiety Scale)	Brownlie 2016; Clegg 2005
QoL Domain III Level of independence	Mobility; activities of daily living; dependence on medical and non-medical substances; work capacity (e.g. duration of unemployment)	Caspi 2016; Clegg 2005; Conti-Ramsden 2018; Johnson 2010; Whitehouse 2009
QoL Domain IV Social relationships	Personal relationships (e.g. family, marital status, number of children, friendships); social support (more formal support than personal, e.g. social worker, access to organised social clubs); anti-social behaviour (e.g. Youth Self Report (Delinquency/Aggression), Conner's Rating Scale (Conduct subscale)	Brownlie 2004; Caspi 2016; Clegg 2005; Johnson 2010; Whitehouse 2009
QoL Domain V Environment	Freedom, physical safety and security (e.g. % criminal court convictions; experienced sexual assault); work and finance (e.g. % in full/part-time work, nature of work, salary, % satisfied with job); transport (e.g. interview about independent travel)	Brownlie 2007; Caspi 2016; Clegg 2005; Conti-Ramsden 2018; Johnson 2010; Whitehouse 2009; Winstanley 2018

IQ: intelligence quotient; QoL: quality of life; WAIS: Wechsler Adult Intelligence Scale
[a]Non-verbal IQ is placed with language and literacy outcomes because it is often tied to diagnosis of language disorders.

Table 3. Summary of findings: long-term prognosis of having low language proficency as a child

Population:

Setting:

Outcomes	Indicators	Effect size (95% CI)	Number of participants	Quality of the evidence (GRADE)	Comments

Table 3. Summary of findings: long-term prognosis of having low language proficency as a child *(Continued)*

	(length of follow-up)	(studies)
Language and literacy		
QoL Domain I Physical		
QoL Domain II Psychological		
QoL Domain III Level of independence		
QoL Domain IV Social relationships		
QoL Domain V Environment		

CI: confidence intervals; QoL: quality of life

APPENDICES

Appendix 1. Search strategy for PsycINFO Ovid

1 language disorders/ and (deficit$ or delay$ or difficult$ or disorder$ or impair$ or problem$ or proficien$).tw.

2 specific language impairment/

3 Speech Language Pathology/

4 language delay/

5 language development/ and (deficit$ or delay$ or difficult$ or disorder$ or impair$ or problem$ or proficien$).tw.

6 (language adj1 (deficit$ or delay$ or difficult$ or disorder$ or impair$ or problem$ or proficien$)).tw.

7 communication disorders/

8 ((language or anomia or aphasia or dysphasi$) adj1 development$).tw.

9 late talk$.tw.

10 ((DLD or PLI or SLD or SLI) and language).tw.

11 or/1-10

12 epidemiology/

13 prognosis/

14 prediction/

15 longitudinal studies/

16 prospective studies/

17 followup studies/

18 risk factors/

19 retrospective studies/

20 academic achievement prediction/

21 occupational success prediction/

22 predict$.tw.

23 (prognosis or prognostic).tw.

24 (follow$ up$ or followup$).tw.

25 longitudinal$.tw.

26 prospectiv$.tw.

27 cohort$.tw.

28 ((adult$ or longterm or long term or later) adj outcome$).tw.

29 or/12-28

30 11 and 29

Appendix 2. Data extraction scheme

Study details	Definition
Study number (ID)	-
Author	First author (surname and first initial)
Year of publication	-
Description of study design	Study description, method, prospective cohort, retrospective cohort, cross-sectional cohort, assessment of outcome, assessment of predictive factors, analysis or discussion of confounders, controlled, with/without intervention
Country of origin	-
Study population/group	Clinical description of participants
Number of participants/clinical group	Number (N)
Comparison group	Comparison group described
Number of participants/comparison group	Number (N)
Inclusion/exclusion criteria	Participants who were eligible for study: description of diagnosis or criteria
Diagnosis at baseline	Description
Known comorbidities	Description of comorbid diagnosis
Diagnostic criteria	Diagnosis or cut-off
Diagnostic tool/measure	Language outcome
Timing of diagnosis	Prior to study, at baseline, etc.
Age of diagnosis	Participants mean age at diagnosis, age range, and SD
Language of participants	Instructional language (language spoken in school)

(Continued)

Bilingualism	Whether home language is different than instructional language
Participants age of study onset	Participants mean age at study onset, age range, and SD
Participants age of study closure	Participants mean age at last follow-up, age range, and SD
Study attrition	Number of participants lost to follow-up; reasons for loss to follow-up
Type of outcomes	-
Study approach and outcomes	When outcomes were measured
Cognitive ability/IQ	Outcome; measure used
Outcome type	Type of outcome(s) e.g. language skills; math skills; SES; education
Outcome measure	Name of test
Outcome instrument	E.g. Self-report; norm-referenced test
Change in diagnosis	Outcome: measure and criteria used
Period of follow-up in years	Length of follow-up for the study
Notes	-
Data	-

Footnotes

ID: identifier;**IQ:** intelligence quotient;**N:** number; **SES:** socioeconomic status

Appendix 3. Risk of bias coding scheme

This risk of bias scheme is based on the Hayden 2013 identification of potential bias in studies of prognostic research. Bias related to participation, attrition, and outcome measurement are particular relevant for this review, and questions developed in the Hayden study will be used to inform our judgements of risk of bias.

Table 1 below describes the potential bias, and guides how we will assess the bias types into high, moderate, low, or unclear risk. The information in Table 1 is modified from the Hayden 2013 tool for questions, developed to advise the judgements of risk of bias in prognostic research (the Quality in Prognosis Studies (QUIPS)).

Table 2 below shows the risk of bias coding scheme that the authors will report.

Table 1. Guidelines for the assessment of risk of bias

Potential bias and questions to guide the assessment*	**Risk of bias ratings**			
	High	**Moderate**	**Low**	**Unclear**
Participation How is the sample described at study onset?	Evidence that the participation rate is low or that the study sample is different from the population of interest.	Indication that the study sample might be different from the	Information reported indicates that the study sample adequately	*General rule*: contact study authors if information is unclear or not reported in the study to enable us to

(Continued)

How is the participation inclusion process described (e.g. in terms of diagnostic criteria)? How is the recruitment described? Are adequate descriptions of the inclusion and exclusion criteria reported?	The described recruitment process indicates that the included sample might differ from the population of interest.	population of interest.	represents the population of interest. Evidence that authors used standardised tests of language to make the diagnosis. High participation and reports of recruitment processes that ensure that the characteristics in the sample are similar to non-participants in the same population.	assess bias. If unable to obtain the required information, bias categories should be coded as unclear. Describe issues that are unclear.
Attrition What is the withdrawal rate? (i.e. have many participants withdrawn?) Is there evidence that outcomes might be biased because the participants completing the study represent a selected group? (i.e. what are the reasons for not completing the study, and is there a risk for systematic differences between persons completing the study and persons lost to follow-up?)	Large withdrawal rate and risk for systematic differences that may bias the prognostic associations from study onset to follow-up. Reported information states that persons who completed the study are likely to differ from those lost to follow-up.	Some withdrawal rate, or possible differences (or both) between persons at follow-up and persons lost to follow-up cannot be ruled out.	The participants at follow-up are likely to represent persons enrolled in the study. The study reports 'complete follow-up'. Evidence that the participants who are missing have random explanations.	*General rule*: contact study authors if information is unclear or not reported in the study to enable us to assess bias. If unable to obtain the required information, bias categories should be coded as unclear.
Outcome measurement Is a clear definition of the outcome provided? Is the method of outcome measurement adequately valid and reliable? (e.g. blinded assessment) Is the method and setting of outcome measurement the same for all study participants?	Reported information indicates that the measurement of the outcome at follow-up is likely to be different from the measures used at study onset. Persons assessing the participants were not blind to the participants' results at study onset. Clear issues related to the measurement itself or to method and setting for outcome assessment.	There are issues with the measurement of the outcome or method that indicate possible differences between follow-up and study onset.	Evidence that the outcome measurement is valid and reliable.	*General rule*: contact study authors if information is unclear or not reported in the study to enable us to assess bias. If unable to obtain the required information, bias categories should be coded as unclear.

Table 2. Risk of bias coding scheme

Publication	Participation	Attrition	Outcome measurement

(Continued)

"First author (year)"	*Rating*: high, moderate, low, unclear	*Rating*: high, moderate, low, unclear	*Rating*: high, moderate, low, unclear
	Description: description of what the decision was based upon, or short reference to the reported information, or both.	*Description*: description of what the decision was based upon, or short reference to the reported information, or both.	*Description*: description of what the decision was based upon, or short reference to the reported information, or both.
"First author (year)"	"	"	"
	"	"	"

Appendix 4. GRADE assessment for judging overall quality of evidence

This table is adapted and modified from Iorio 2015.

	Domain	Description
Grade down	Risk of bias	If most evidence is from studies with moderate or high risk of bias for most of the bias domains
	Inconsistency	If large I^2 value, indicating significant heterogeneity Unexplained heterogeneity or variability in results across studies, with differences in results that are not clinically meaningful
	Indirectness	If 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	If there is no precise estimate of the effect size in the meta-analysis, and confidence intervals are excessively wide, or they overlap with the value of no effect Also, if there is imprecision across studies: few studies and insufficient sample size (< 100 cases reaching follow-up), or no justification provided for small sample size
	Publication bias	If indications of publication bias
Grade up	Large prognostic effect	If moderate or large effect reported by most studies or in pooled findings in the meta-analysis
	Dose-response gradient	If 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)

Appendix 5. Levels of quality

Table taken from Brignell 2022 and Iorio 2015.

Quality level	Definition

(Continued)

High	Very confident that the true prognostic effect lies close to that of the estimate.
Moderate	Moderately confident that the true prognostic effect is likely to be close to the estimate, but there is a possibility that it is substantially different.
Low	Limited confidence in the estimate.
Very low	Very little confidence in the estimate.

WHAT'S NEW

Date	Event	Description
20 January 2023	Amended	Correcting title

HISTORY

Protocol first published: Issue 1, 2023

CONTRIBUTIONS OF AUTHORS

The contributions of authors for the protocol are as follows.

Åste M Hagen (ÅMH) coordinated the protocol, drafted the Background and parts of the Methods section, notably Criteria for considering studies for this review (types of studies, targeted population, types of outcomes to be predicted), Search methods for identification of studies, and Data collection (selection of studies, data extraction and management, and assessment of risk of bias in included studies).

Kristin Rogde (KR) drafted the appendices, parts of the Methods section, notably Search methods, and commented on the rest of the draft.

Monica Melby-Lervåg (MM-L) drafted large parts of the Methods section, notably Data collection (measures of association or predictive performance measures to be extracted, dealing with missing data, assessment of heterogeneity, assessment of reporting deficiencies), and Data synthesis (data synthesis and meta-analysis approaches, subgroup analysis and investigation of heterogeneity, sensitivity analysis). She also commented on the rest of the draft.

Arne Lervåg (AL) consulted and commented on the draft.

Courtenay Norbury (CN) consulted and co-wrote the Background and the Methods sections.

ÅMH is the guarantor for the review. KR and ÅMH developed the search strategy.

DECLARATIONS OF INTEREST

Åste M Hagen (ÅMH) has declared that she has no conflicts of interest.

Kristin Rogde (KR) has declared that she has no conflicts of interest.

Monica Melby-Lervåg (MM-L) has declared that she has no conflicts of interest.

Arne Lervåg (AL) has declared that he has no conflicts of interest.

Courtenay Norbury (CN): reports being the Principal Investigator of the Surrey Communication and Language in Education Study (SCALES), which is eligible for inclusion in the review; the study was funded by a grant from the Economic and Social Research Council (2018 to 2021), and by a grant from Wellcome (2012 to 2016). The funders of these grants had no role in the study design, conduct, methods, data analysis, reporting, or publication of this protocol for the review.

SOURCES OF SUPPORT

Internal sources

- Department of Special Needs Education, University of Oslo, Norway

  Support for the preparation of the protocol and review to be carried out during office hours for ÅMH, KR, and MM-L
- Department of Education, University of Oslo, Norway

  Support for the preparation of the protocol and review to be carried out during office hours for AL

External sources

- The Research Council of Norway, Norway

  This review is part of a project, Better Equipped, which is funded by the Research Council of Norway (grant number 324207). The funder had no role in the study design, conduct, methods, data analysis, reporting, or publication of this protocol for the review.