Screening and Treatment for Developmental Delay in Early Childhood (ages 1-4): Systematic Review

Original Review: June 6th, 2014

Integrated Addendum Review: October 28th, 2015

McMaster Evidence Review and Synthesis Centre Team:

Rachel Warren, Meghan Kenny, Donna Fitzpatrick-Lewis, Muhammad Usman Ali, Maureen Rice, Andy Bayer, Sharon Peck-Reid, Donna Ciliska, Diana Sherifali, Parminder Raina

McMaster University, Hamilton Ontario Canada

Evidence Review Clinical Expert:

Dr. Teresa Bennett

CTFPHC Working Group:

Marcello Tonelli (Chair)

CTFPHC Working Group Members:

Paula Brauer, Patricia Parkin, Kevin Pottie, Brett Thombs

External Members: Denis Leduc

PHAC Scientific Research Manager:

Lesley Dunfield and Wendy Martin

Abstract

Background: Developmental delay (DD), defined as a failure to reach age-expected milestones, has been recognized as the most common form of disability in children aged 0 to 4 in Canada. This systematic review was produced for the Canadian Task Force on Preventive Health Care (CTFPHC) to provide guidelines on the screening of children for DD. The last CTFPHC guideline on this topic was published in 1994.

Purpose: The purpose of this staged review is to synthesize evidence on the effectiveness and harms of screening (Stage I) and treatment (Stage II) of DD, and to provide information regarding test properties of screening methods (Stage III) for children who are not otherwise identified as being at high risk of or suspected to have DD.

Review Approach and Inclusion Criteria: This was a staged review. For Stage I (Screening) RCTs and controlled clinical trials on effectiveness and harms of screening for children aged 1-4 who were not high risk or suspected of having DD were sought. In the absence of good-quality evidence, Stage II (Treatment) was initiated, as this indirect evidence may be used to inform a screening recommendation. Stage II involved three parts. We searched first for systematic reviews on behavioural or psychological treatment of children aged 1-6 diagnosed with DD or autism spectrum disorder or autism disorder. Next, we searched for RCTs on behavioural or psychological treatment of children aged 1-6 diagnosed with DD. Due to a paucity of RCT evidence on the pre-specified outcomes of interest, a third search (Stage II Addendum) was undertaken to identify RCT evidence with domain-specific outcomes. For Stage III (Test Properties) studies of any design were sought that assessed test properties of Canadian relevant screening methods for DD, ASD, and AD.

Data Sources: Stage I: Medline, Embase and PsychINFO (no beginning date limitations through February 24th, 2014); Stage II: Medline, Embase, PsychINFO and the Cochrane Database of Systematic Reviews (systematic reviews) (2009 to March 4th, 2014) and Medline, Embase, Cochrane Central and PsycINFO (RCTs) (2000 to March 25th, 2014); Stage II (Addendum): Medline, Embase, Cochrane Central and PsycINFO (randomized controlled trials [RCTs]) (2000 to June 16th, 2015); Stage III: Medline, Embase, and PsycINFO (no beginning date limitations to March 13th, 2014).

Study Selection: The titles and abstracts of papers considered for the key questions and sub-questions were reviewed independently by two reviewers; any article marked for inclusion by either team member went on to full-text screening. Full-text review was done independently by two people with consensus required for inclusion or exclusion.

Data Abstraction: Review team members extracted data about the population, study design, intervention, analysis and results for outcomes of interest. One team member completed full abstraction, followed by a second team member who verified all extracted data and ratings. We assessed study quality using Cochrane's Risk of Bias tool and the GRADE framework (Stage I), AMSTAR (Stage II) and Cochrane RoB (Stage II Addendum) or QUADAS-II (Stage III). For the contextual questions, inclusion screening and abstraction were done by one person.

Analysis: A meta-analysis could not be performed due to the paucity of studies reporting on effectiveness of DD screening (Stage I) such as referral rate, time to intervention referral and academic performance. The forest plots display the effect size data reported by the studies. For effectiveness and harms of treatment (Stage II) a review of reviews was done and results were reported narratively; RCT evidence was meta-analysed when possible. The screening test properties data were extracted or calculated across studies from reported sensitivity, specificity and prevalence in 2 x 2 contingency tables (true positive, true negative, false positives and false negatives). Extracted test properties data were meta-analyzed where possible using exact binomial rendition of the bivariate mixed-effects regression model modified for synthesis of screening test data.

Results: For effectiveness of screening (Stage I), two trials reporting on the referral rate, time to referral and academic performance outcomes were included; these outcomes were assessed using GRADE. These trials could not be meta-analysed. We found no studies addressing the screening outcomes of cognitive function, incidence of mental health conditions, overall quality of life, survival, or functionality as an adult.

One moderate quality trial (n= 2,103) reported referral data for mixed gender children <30 months who were screened for DD using Ages and Stages Questionnaire-II (ASQ-II). This study reported significantly more referrals to early intervention than the control group with a relative risk (RR) of 1.95 (95% CI 1.49 to 2.54) in the intervention group with office support (participants met with trained office staff to complete the ASQ screening tool with the use of props) and an RR of 1.71 (95% CI 1.30 to 2.25) in the intervention group without office support. The authors found a 70% shorter time to referral in the intervention group with office support (Rate Ratio of 0.30 [95% CI 0.19 to 0.48]), and a 64% shorter time to referral for the intervention group without office support (Rate Ratio of 0.36 [95% CI 0.23 to 0.59]), both compared to the control group.

One low quality trial (n=11,440) screening mixed gender children aged 15 months at entry for language delay with the VroegTijdige Onderkenning Ontwikkelingsstoornissen (VTO) Language Screening instrument, reported no differences between groups in academic performance outcomes at age eight years with an RR of 0.71 (95% CI 0.48 to1.04) of attending a special school; an RR of 0.99 (95% CI 0.81 to1.21) of repeating a grade; an RR of 1.26 (95% CI 0.89 to1.80) of repeating a grade because of language problems in regular primary school; an RR of 0.88 (95% CI 0.63 to1.23) of being below the 10^{th} percentile of oral tests; an RR of 1.00 (95% CI 0.72 to1.40) of being below the 10^{th} percentile of reading tests in grade 2; and an RR of 0.68 (95% CI 0.41 to 1.13) of being below the 10^{th} percentile of spelling tests in grade 2.

For effectiveness of treatment (Stage II), we found no systematic reviews or RCTs addressing treatment outcomes of academic performance, survival, and functionality as an adult. Five systematic reviews on treatment of Autism Spectrum Disorder included the outcomes of cognitive function, quality of life and harms; results were mixed.

Four systematic reviews reported on the results of behavioural interventions: two systematic reviews (8 unique studies; n=329) found significant differences in cognitive function between groups; one systematic review (one study; n=24) found non-significant changes and one systematic review did not include any unique studies. One systematic review on

acupuncture/acupressure plus conventional treatment (4 studies; n=179) reported no significant differences in three studies; one study found greater improvement in the intervention group.

Two systematic reviews (one on behavioural interventions and one on acupuncture/acupressure) examined quality of life. The systematic review on acupuncture/acupressure interventions found no studies reporting on quality of life; one systematic review on behavioural treatment showed a statistically significant difference in daily living skills. Two systematic reviews examined harms of treatment (one on behavioural interventions and one on acupuncture/acupressure. No adverse events were reported in the systematic review on behavioural interventions (5 studies; n=200); adverse events as a result of acupuncture/acupressure (9 studies; n=357) included initial crying for fear or pain, superficial bleeding, crying or irritability, and worsening of sleep. Three of these studies did not report any harms.

For effectiveness of domain-specific treatment (Stage II Addendum) three RCTs (n=239) provided evidence on the outcome of language impairment; these outcomes were assessed using GRADE. The pooled estimate across three studies showed a significant improvement in language impairment for intervention group as compared to controls with an SMD of 0.81 (95% CI 0.02 to 1.6; I^2= 84%). This body of evidence was rated moderate quality. One RCT (n=155) provided data for the outcome of social and personal activities of daily living (adaptive functioning). The effect estimate showed no difference between intervention and control groups with a mean difference (MD) of 0.60 in the Vineland Adaptive Behaviour Scale (95% CI -3.05 to 4.25). This body of evidence was rated low quality.

For test properties (Stage III), 17 observational studies were found reporting on test properties of screening tests for ASD, AD, DD, and PDD. Nine cohort studies (n=70,816) using the MCHAT as a screening tool for ASD showed a sensitivity of 78.0% (95% CI 64.0% to 88.0%), a specificity of 69.0% (95% CI 47.0% to 85.0%), a positive predictive value of 30.3% (95% CI 17.3% to 50.4%) and 94.8% (95% CI 88.3% to 97.6%) and a negative predictive value of 94.8% (95% CI 88.3% to 97.6%). Four cohort studies (n=1,001) using ASQ as a screening tool for DD showed a median sensitivity of 55.0% (range 47.1% to 66.7%), a median specificity of 86.0% (range 38.6% to 94.3%), a median positive predictive value of 41.4% (range 23.2% to 71.4%) and a median negative predictive value of 84.9% (range 70.8% to 95.6%).

Conclusion: In children aged 1 to 4 years of age without suspected DD, the evidence that screening for DD improves outcomes is inconclusive. For direct evidence on effectiveness of screening, there is a paucity of studies reporting on long term outcomes. Indirect treatment evidence is mixed, though there is some evidence to support the effect of treatment on language-specific outcomes. The current evidence on various tools for screening for ASD, AD, DD and PDD is limited.

Table of Contents

List of Acronyms

AAP	American Academy of Pediatrics
ABA	Applied Behavioural Analysis
ABI	Applied Behavioural Intervention
AD	Autistic Disorder
ADHD	Attention Deficit Hyperactivity Disorder
ADOS	Autism Diagnostic Observation Schedule
ADOS-T	Autism Diagnostic Observation Schedule Toddler Module
AHRQ	Agency for Healthcare Research and Quality
AKQ	Addendum Key Question
AMSTAR	A Measurement Tool to Assess Systematic Reviews
ARR	Absolute Risk Reduction
ASD	Autism Spectrum Disorders
ASQ	Ages and Stages Questionnaire
ASQ-II	Ages and Stages Questionnaire 2nd Edition
BDI	Battelle Developmental Inventory
BINS	Bayley Infant Neurodevelopmental Screener
BSID	Bayley Scale of Infant Development
CD	Communication Delay
CDC	Centers for Disease Control and Prevention
CHAT	Checklist for Autism in Toddlers
CI	Confidence Interval
CPS	Canadian Paediatric Society
CSBS-DP	Communication and Symbolic Behaviour Scales – Developmental Profiles
CTFPHC	Canadian Task Force on Preventive Health Care
DD	Developmental Delay
DDST	Denver Developmental Screening Test
DSM-IV	Diagnostic and Statistical Manual 4th Edition
DSM-5	Diagnostic and Statistical Manual 5th Edition
EI	Early Intervention
EIBI	Early Intensive Behaviour Intervention
ESAT	Early Screening of Autistic Traits Questionnaire
FU-MCHAT	Follow-up Modified Checklist for Autism in Toddlers
GRADE	Grading of Recommendations, Assessment, Development and Evaluations
IQ	Intelligence Quotient
KQ	Key Question
M-CHAT	Modified Checklist for Autism in Toddlers
MSEL	Mullen Scales of Early Learning
NDDS	Nipissing District Developmental Screen
NNS	Number Needed to Screen
NPV	Negative Predictive Value
NSCH	National Survey of Children's Health
OCD	Obsessive Compulsive Disorder
ODD	Oppositional Defiant Disorder
OR	Odds Ratio

PDD	Pervasive Developmental Disorder
PEDS	Parents' Evaluation of Developmental Status
PEP-R	Psychoeducational Profile Revised
PLS	Preschool Language Scale
PPV	Positive Predictive Value
PRESS	Peer Review of Electronic Search Strategies
QUADAS II	Quality Assessment of Diagnostic Accuracy Studies 2nd Edition
RCT	Randomized Controlled Trial
RR	Relative Risk
SCQ	Social Communication Questionnaire
SD	Standard Deviation
SES	Socioeconomic Status
SIGN	Scottish Intercollegiate Guidelines Network
SMD	Standardized Mean Difference
SRS	Social Responsiveness Scale
STAT	Screening Test for Toddlers
US	United States of America
USPSTF	United States Preventive Services Task Force
VABS	Vineland Adaptive Behavior Scale
VTO	VroegTijdige Onderkenning Ontwikkelingsstoornissen
WHO	World Health Organization
WISC	Wechsler Intelligence Scale for Children
WPPSI	Wechsler Preschool and Primary Scale of Intelligence

Chapter 1: Introduction

Purpose and Background

This systematic review will be used by the Canadian Task Force on Preventive Health Care (CTFPHC) to inform recommendations on screening for developmental delay (DD) in children aged 1 to 4 years who are not suspected of or at high risk for DD in a primary care setting. Infants younger than 1 year without signs or symptoms are less likely to be assessed for DD. Delays in children older than 4 years may be identified in schools. The CTFPHC[i] updated its previous review (1990) on well-baby care in 1994 and found fair evidence to assess developmental milestones at each well-baby visit in the guideline on Well-Baby Care in the First 2 Years of Life (B recommendation).[1] The 1994 CTFPHC recommendation on Preschool Screening for Developmental Problems[2] found good evidence to recommend against the use of the Denver Developmental Screening Test (DDST)[3] in asymptomatic preschool children. There was insufficient evidence for other screening tools.

This systematic review synthesizes the effectiveness and harms of screening for DD in children; synthesizes the evidence for effectiveness and harms of treatment for DD; evaluates test properties for selected screening tests; and answers contextual questions.

Definition

DD is defined as a failure to reach age-expected milestones in any one of four domains: 1. gross and fine motor skills, 2. speech and language, 3. social and personal activities of daily living, and 4. performance and cognition (i.e.: ≥25% or 1.5-2 standard deviations discrepancy from age-expected milestones). Global DD is a significant delay in two or more of the four main developmental domains.[4, 5] Although autistic disorder (AD) and autism spectrum disorder (ASD), as defined by the Diagnostic and Statistical Manual 4th Edition (DSM-IV),[6] are separate but interrelated concepts, for the purposes of this review they are considered part of DD.

Prevalence and Burden of Developmental Delay

In 2001, a Statistics Canada report stated that DD (global and domain specific) was the most common disability in children aged 0 to 4 years in Canada, experienced by 1% of this population.[7] Of children identified with DD, 59% have intellectual delay, 54% have physical delay and 38% have other delays inclusive of speech delay.[7] In 2006, Statistics Canada reported that developmental delay was significantly higher in boys (1.3%) aged 0 to 4 years than girls (0.8%) aged 0 to 4 years.[8] Additional recent surveys suggest that 1% to 3% of children are affected with global DD and 5-10% have a delay in a single domain.[4, 9] Children with global DD

[i] In 1990 the CTFPHC was known as the Canadian Task Force on the Periodic Health Examination.

often develop learning, behavioural, or emotional problems and may be at higher risk for other health problems.[10]

Risk Factors

The most frequently reported characteristics that are associated with an increased risk of developmental delay include factors relating to birth (low birth weight, multiple birth, premature birth), exposure to infection or high levels of environmental toxins during pregnancy, family history of DD, and parental use of alcohol and/or tobacco. Less commonly reported risk factors for DD include birth order, larger family size, illnesses experienced during childhood, and parental educational attainment.[11]

Rationale and Strategies for Screening

The preschool years form a period of intense developmental change across multiple domains. Reliably determining the clinical severity of DD in children younger than age 5 years using standardized measures is therefore challenging. This is underscored in the Diagnostic and Statistical Manual, 5th Edition (DSM-5)[12] which emphasizes that a diagnosis of global DD may change after age 5 when standardized scores (e.g., Intelligence Quotient [IQ]) are more stable in young children and more requirements for other diagnoses (e.g. Intellectual Disability) may be met.

Nevertheless, it has been reported that DD in some children can be improved with early identification and early intervention. [4, 10, 13] Interventions may include parenting programs, early learning centres, speech and language programs, and physical or occupational therapy.[14, 15] Interventions may influence school readiness, which in turn could increase rate of high school graduation, which in turn could increase employability.[16, 17] Finally, lower IQ is associated with higher all-cause mortality – raising the hypothesis that interventions which increase IQ may also lead to reduced mortality in adulthood.[18]

Screening for DD was identified by family physicians as a topic of interest, especially because there is a perceived lack of resources available for interventions directed at management of DD. Such interventions could be initiated or monitored at the level of primary care.

Current Clinical Practice

In Canada, recommendations for an enhanced 18-month well-baby visit, which includes using the Nipissing District Developmental Screen (NDDS) and Rourke Baby Record as a surveillance tool to monitor for DD have been made in Ontario.[10, 19] An Expert Panel on the 18-month well-baby visit determined that it would be efficient to expand the 18-month visit to include among other things, a more detailed assessment of the child's stage of development. The panel drew support for this decision from a retrospective review/audit in a family practice that instituted an enhanced 18-month well-baby visit. Results from this review indicated that the additional evaluation components identified concerns that required follow-up. This enhanced well-baby visit has been supported by the

Canadian Paediatric Society (CPS).[15] A 2011 Canadian national scan indicated that most provinces are interested in 18-month monitoring, and some provinces (Alberta, Manitoba, Nova Scotia, Saskatchewan) are undertaking or have completed pilot programs of screening.[15]

Other National Guidelines

In 2006, the United States Preventive Services Task Force (USPSTF) assessed screening for speech and language delay in preschool children and found insufficient evidence for the use of screening instruments in children up to 5 years of age to detect speech and language delay in primary care.[20] These guidelines were updated in 2015, and the same conclusions were reached.[21]

In 2007, The Scottish Intercollegiate Guidelines Network (SIGN) recommended against population screening for autism spectrum disorders (ASD).[22] The SIGN report states that as part of child health surveillance, health professionals should be aware of early signs of ASD and may use the Modified Checklist for Autism in Toddlers (M-CHAT)[23] to identify children who may be at risk for autism. The USPSTF is currently finalizing a plan for guideline development on screening for ASD in children between the ages of 12 and 36 months.[24]

The AAP recommends screening for DD using a standardized tool at 9, 18 and 24 or 30 months of age [25] and screening for autism at 18 months and 24 months.[26] Rates of referral to early intervention following positive screening were assessed following a pilot program based on the AAP recommendations, and results showed that these rates varied (from 48% to 78%) and did not meet the recommendation that every child who screens positive during DD screening be referred for early intervention. The screening test differed across the various centres involved in the pilot project, and included the Parents' Evaluation of Developmental Status (PEDS), Ages and Stages Questionnaire 2nd Edition (ASQ-II), and DDST.[27]

Chapter 2: Methods

Review Approach

We conducted a systematic literature search to address the effectiveness and harms of screening for global and domain specific DD, the effectiveness and harms of treatment for global and domain specific DD, and the test properties of commonly used screening tests. The screening tests included general tests, tests for a single domain, as well as tests for ASD and AD. A separate search was conducted for contextual questions.

This was a staged review. Stage I included the identification of studies that addressed screening effectiveness in children 1 to 4 years who were not suspected of having DD or at risk for DD. In the absence of identifying good-quality evidence (randomized controlled trials or controlled clinical trials) for the outcomes of interest, Stage II was initiated. Stage II included a search for systematic reviews and randomized controlled trials (RCTs) that addressed treatment effectiveness or harms, as this indirect evidence may be used to inform a screening recommendation. Due to a paucity of RCT evidence in Stage II, an addendum which expanded our outcomes of interest was added to our staged strategy in order to ensure that our search was exhaustive. Stage III involved a search for the test properties of validated screening tests or tests in use that assessed DD.

Many of the methods outlined in this section are used consistently throughout our reviews (see http://canadiantaskforce.ca/ for details of past reviews).

Analytic Framework and Key Questions

The analytic framework for this review is presented in Figure 1. The framework provides a visual representation of the parameters of the review, including population, interventions and outcomes of interest.

KEY QUESTIONS (KQ)

Stage I:

1. What is the effectiveness of screening children aged 1 to 4 years without suspected DD to improve outcomes? (outcomes of interest: referral rates for early intervention; time to referral to early intervention; cognitive function; academic performance; incidence of mental health conditions; overall quality of life; survival; functionality as an adult)
   a. What is the optimal interval for screening for DD?
2. What is the incidence of harms of screening children aged 1 to 4 years without suspected DD?

Stage II:

3. What is the effectiveness of treatment for children diagnosed with DD to improve outcomes? (outcomes of interest: cognitive function; academic performance; incidence of mental health conditions; overall quality of life; survival; functionality as an adult)
4. What is the incidence of harms of treatment for children diagnosed with DD?

Stage II (Addendum):

Addendum KQ1. What is the effectiveness of treatment for children diagnosed with DD to improve outcomes in gross and fine motor skills, language impairment, adaptive functioning, intellectual disability (IQ), learning disability (academic testing) and academic underachievement?

Addendum KQ2. What is the incidence of any harms of treatment for children diagnosed with DD?

Stage III:

5. What is the sensitivity, specificity, positive and negative predictive values, and likelihood ratios of the various screening tests to assess DD in children aged 1 to 4 years who are not already suspected of having DD?

CONTEXTUAL QUESTIONS

1. What is the cost-effectiveness and feasibility of screening for DD in preschool children aged 1 to 4 years?
2. What are parent or primary caregiver values and preferences for screening for DD in preschool children aged 1 to 4 years?
3. What is the evidence for a higher burden of disease, a differential treatment response, differential performance of screening for DD, or barriers to implementation of screening for DD in subgroups? Subgroups include: Aboriginal, rural or remote populations, low socioeconomic status, drug or alcohol dependency, or other ethnic populations.

Study Selection

The titles and abstracts of papers considered for the key questions and sub questions were reviewed in duplicate by members of the synthesis team; any article marked for inclusion by either team member went on to full-text rating. Full-text inclusion was done independently by two people. All disagreements were resolved through discussions and consensus. For papers located in the contextual questions search, title and abstract screening was done by one person.

Stage I – Screening

Search Strategy

For Stage I key questions 1 and 2, we searched Medline, Embase and PsychINFO with no beginning date limitations to September 16th, 2015 (Appendix 1). This search was peer-reviewed using the Peer Review of Electronic Search Strategies (PRESS) format.[28]

Inclusion and Exclusion Criteria

Population

The population of interest was children aged 1-4 years of age who were not suspected of or at high risk of DD. High risk has been defined as those born prematurely (gestational age less than 37 completed weeks at birth) or with low birth weight (birth weight less than 2,500 g) and/or children with other known disorders that may be associated with or affect development. We also excluded studies of children over 4 years, studies of case finding in children in whom DD was suspected or children at high risk for DD, and studies on screening for hearing or vision problems (as these are usually identified through specific hearing and vision screening tests). Samples must have a mean age that falls under our upper age limit.

Language

The published results of studies had to be available in either English or French.

Interventions

Screening with any test, tool, or questionnaire used to screen for DD; including tools for specific domains, tools for general DD, and tools for AD and ASD. We excluded the DDST as previous CTFPHC guidelines found good evidence recommending against its use.[2]

Settings

Settings were limited to primary care settings and public health clinics. Studies conducted in school settings were not included.

Study Design and Comparison Groups

Randomized controlled trials (RCTs) and controlled cohort studies with comparison groups that did not receive screening were eligible. Any study design (with or without comparison groups) was considered acceptable to answer the questions on harms.

Outcomes

The outcomes of interest included clinically relevant changes in:

- referral rates for early intervention

- time to referral to early intervention
- cognitive function
- academic performance
- incidence of mental health conditions (diagnosis or symptoms), as defined by DSM-IV[6] including anxiety; depression; oppositional defiant disorder (ODD); obsessive-compulsive disorder (OCD)
- overall quality of life
- survival
- functionality as an adult (including employment, criminality, and independence)

To answer the question on harms of screening outcomes included parental anxiety and stigma (labeling).

Timeframe

There was no minimum follow-up time necessary for inclusion in our evidence summary.

Data Abstraction

For each study used to answer any KQ, review team members extracted data about the population, the study design, the intervention, the analysis and the results for outcomes of interest. For each study, one team member completed full abstraction (study characteristics, risk of bias assessment, outcome data) using electronic forms housed in a web-based systematic review software program.[29] A second team member then verified all extracted data and ratings; disagreements were resolved through discussion and/or third party consultation when consensus could not be reached.

Assessing Risk of Bias

Arriving at a Grading of Recommendations, Assessment, Development and Evaluations (GRADE) rating for a body of evidence (see next section) requires a preliminary assessment of the risk of bias or study limitations for the individual studies. All studies included to answer the effectiveness of screening question were assessed using the Cochrane Risk of Bias tool.[30] This rating tool covers six domains: sequence generation; allocation concealment; blinding of participants, personnel and outcome assessors; incomplete outcome reporting; selective outcome reporting; and other risk of bias. Information to determine risk of bias was abstracted from the primary methodology paper for each study and any other relevant published papers. For each study, one team member completed the initial ratings which were then verified by a second person; disagreements were resolved through discussion and/or third party consultation when consensus could not be reached. To assign a high or low risk of bias rating for a particular domain we looked for explicit statements or other clear indications that the relevant

methodological procedures were or were not followed. In the absence of such details we assigned unclear ratings to the applicable risk of bias domains.

To determine the overall risk of bias rating for an outcome group we considered all domains. However, greater emphasis was placed on the assessments of the four areas of randomization, allocation, blinding of outcome assessment and selective reporting because those represented most significant sources of introducing bias to RCTs on DD and hence could lead to biased estimates of outcome findings and conclusions. Table 2a summarizes the risk of bias ratings applied to the studies included in this stage of the review (Stage I).

Assessing Strength or Quality of the Evidence

The strength of the evidence was determined based on the GRADE system of rating the quality of evidence using GRADEPro software.[31, 32] This system of assessing evidence is widely used and is endorsed by over 40 major organizations including the World Health Organization (WHO), Centers for Disease Control and Prevention (CDC), and the Agency for Healthcare Research and Quality (AHRQ).[33] The GRADE system rates the quality of a body of evidence as high, moderate, low or very low; each of the four levels reflects a different assessment of the likelihood that further research will impact the estimate of effect (i.e., high quality: further research is unlikely to change confidence in the estimate of effect; moderate quality: further research is likely to have an important impact on confidence in the estimate of effect and may change the estimate; low quality: further research is very likely to have an important impact on confidence in the estimate of effect and is likely to change the estimate; very low quality: the estimate of effect is very uncertain).[33]

A GRADE quality rating is based on an assessment of five domains: (1) risk of bias (limitations in study designs), (2) inconsistency (heterogeneity) in the direction and/or size of the estimates of effect, (3) indirectness of the body of evidence to the populations, interventions, comparators and/or outcomes of interest, (4) imprecision of results (few participants/events/observations, wide confidence intervals [CIs]), and (5) indications of reporting or publication bias. Grouped RCTs begin with a high quality rating which may be downgraded if there are serious or very serious concerns across the studies related to one or more of the five conditions. For this review, key data were entered into the GRADEPro software along with the quality assessment ratings to produce two analytic products for each outcome and the comparisons of interest: (1) a GRADE Evidence Profile Table and (2) a GRADE Summary of Findings Table.

Data Analysis

The aggregate quality of the body of evidence for each outcome was assessed using the GRADE evidence rating approach based on number and type of studies, study risk of bias, consistency of results across studies, precision and directness of evidence.

A meta-analysis could not be performed due to a paucity of studies reporting on effectiveness of DD screening. The forest plots display the effect size data reported by the studies. The subgrouping in forest plots was based on the type of screening method and outcome measures used.

For the effectiveness of DD screening that showed a statistically significant effect, we added the estimates of absolute risk reduction (ARR) and number needed to screen (NNS). The NNSs were calculated using the absolute numbers presented in the GRADE tables. GRADE estimates the absolute number per million using the control group event rate and risk ratio with the 95% confidence interval.

Stage II – Treatment

Search Strategy

For Stage II key questions 3 and 4, two searches were performed. We first searched Medline, Embase, PsychINFO and the Cochrane Database of Systematic Reviews from 2009 to September 16th, 2015 (Appendix 2) for systematic reviews (Appendix 2). Next we searched Medline, Embase, Cochrane Central and PsycINFO from 2000 to September 16th, 2015 for RCTs (Appendix 3). As we had located a sufficient number of moderate to high quality (n= 5) systematic reviews on treatment for AD/ASD, we limited our search to RCTs on global or domain specific DD.

Inclusion and Exclusion Criteria

Population

For Stage II the population of interest was children beginning intervention for DD between the ages of 1-6 years, in order to factor in time from diagnosis to treatment. Excluded from this question were children with externalizing disorders (conduct disorders, ODD and Attention Deficit Hyperactivity Disorder) as these disorders are typically identified in children of school age, and children with hearing and vision problems (as these are usually identified through specific hearing and vision screening tests).

Language

The published results of studies had to be available in either English or French.

Interventions

Interventions for DD using behavioural or psychological interventions were included.

Settings

Settings were limited to primary care settings and public health clinics. Studies conducted in school settings were not included.

Study Design and Comparison Groups

Systematic reviews and RCTs using comparison groups receiving usual care or no intervention were considered. Systematic reviews that included various types of comparison groups were included but only data from studies with no treatment/usual care comparators were extracted. To answer the question on harms of treatment, a usual care or no intervention comparison group was not required.

Outcomes

The outcomes of interest included clinically relevant changes in:

- cognitive function
- academic performance
- incidence of mental health conditions (diagnosis or symptoms), as defined by DSM-IV[6] including anxiety; depression; oppositional defiant disorder (ODD); obsessive-compulsive disorder (OCD)
- overall quality of life
- survival
- functionality as an adult (including employment, criminality, and independence)

To answer the question of harms of treatment, any harms arising from DD treatment were included.

Timeframe

There was no minimum follow-up time necessary for inclusion in our evidence summary.

Data Abstraction

For each study used to answer any KQ, review team members extracted data about the population, the study design, the intervention, the analysis and the results for outcomes of interest. For each study, one team member completed full abstraction (study characteristics, risk of bias/quality assessment, outcome data) using electronic forms housed in a web-based systematic review software program. [29] A second team member then verified all extracted data and ratings; disagreements were resolved through discussion and/or third party consultation when consensus could not be reached.

Assessing Risk of Bias

RCTs

RCTs were to be assessed using the Cochrane Risk of Bias tool (as described in Stage I). Because no studies met our inclusion criteria for this question, no risk of bias assessment was performed on Stage II RCTs.

Assessing Strength or Quality of the Evidence

Systematic Reviews

The methodological quality of the included systematic reviews was assessed using A Measurement Tool to Assess Systematic Reviews (AMSTAR).[34] AMSTAR is a valid and reliable instrument used to assess the methodological quality of systematic reviews.[34] The AMSTAR checklist is comprised of eleven questions on the design of the review; the selection and extraction of studies; the search strategy; the inclusion of study lists; the assessment and reporting of scientific quality; the methods used to combine studies; the assessment of publication bias and the statement of conflict of interest. Reviews that were rated 11 to 9 using AMSTAR were considered high methodological quality systematic reviews; studies rated 8 to 6 were considered moderate methodological quality and reviews rated 5 to 0 were considered low methodological quality. The quality of the included studies within the systematic reviews was not assessed.

RCTs

Strength or quality of RCTs were to be assessed using GRADE (as described in Stage I). Because no studies met our inclusion criteria for this question, GRADE was not performed on Stage II RCTs.

Data Analysis

Systematic Reviews

For effectiveness and harms of treatment, a review of reviews was done. Results from systematic reviews that received high scores (9, 10 or 11/11) on AMSTAR were reported narratively. We also considered moderate-high methodological quality systematic reviews (rated 8) that focused solely on our population of interest (children aged 1-6). Other moderate methodological quality reviews and low methodological quality were not included in narrative summary.

If a systematic review provided a meta-analysis of studies meeting our criteria we reported those data. If there was no meta-analysis with studies meeting our inclusion criteria we looked at the effect sizes and provided a median and range, where possible. For studies reporting standardized mean difference (SMD), we used Cohen's rule: an SMD of 0.2 or less indicates a very small effect size, a value between 0.2 and 0.5 indicates a small effect, a value between 0.5 and 0.8 indicates a medium effect and a value of 0.8 or larger indicates a large effect.[35] Results were reported narratively. When possible we provided overall sample sizes for the studies of interest from the systematic reviews.

RCTs

We intended to perform a meta-analysis on Stage II RCTs. Because no studies met our inclusion criteria for this question a meta-analysis was not performed.

Stage II – Treatment (Addendum)

Due to the absence of RCT evidence on DD treatment for children based on important and critical outcomes listed above, an addendum was added to search for domain specific outcomes (see below for complete list) in the DD treatment RCT literature.

Search Strategy

For Stage II (Addendum) key questions 1 and 2 we searched Medline, Embase, Cochrane Central and PsycINFO for RCTs from 2000 to September 16th, 2015 (Appendix 4). We limited our search to RCTs on global or domain specific DD.

Study Selection

The titles and abstracts of papers considered for the key questions were reviewed in duplicate by members of the synthesis team; any article marked for inclusion by either team member went on to full-text rating. Full-text inclusion was done independently by two people. All disagreements were resolved through discussions and consensus.

Inclusion and Exclusion Criteria

Population

Children diagnosed with domain specific developmental delay (DD) in one or more of the domains (gross and fine motor skills; speech and language; social and personal activities of daily living; performance and cognition). The treatment intervention had to have been initiated in children between the ages of 1-6 years.

Intervention

Behavioural, pharmacological and psychological interventions for DD were included.

Comparator

RCTs using no treatment comparison groups were included. In the absence of no treatment comparators, we included usual care comparison groups.

Outcomes

To answer AKQ1, the outcomes of interest included clinically relevant changes in:

- Gross and fine motor skills
- Language impairment
- Adaptive functioning
- Intellectual disability (IQ)
- Learning disability (academic testing)
- Academic underachievement

There was no inclusion restriction on the type of tool by which the outcomes were measured.

We considered any harm (AKQ2) of treatment for DD to answer the question on harms of treatment.

Data Abstraction and Quality Assessments

For each study used to answer any KQ, review team members extracted data about the population, the study design, the intervention, the analysis and the results for outcomes of interest. For each study, one team member completed full abstraction (study characteristics, risk of bias assessment, outcome data) using electronic forms housed in a web-based systematic review software program.[29] A second team member then verified all extracted data and ratings; disagreements were resolved through discussion and/or third party consultation when consensus could not be reached. RCTs were assessed using the Cochrane Risk of Bias tool[30] (see Table 2 for summary ratings).

Assessing Strength or Quality of Evidence

The strength of the evidence was determined based on the GRADE system of rating the quality of evidence (as described in Stage I). Table 2b summarizes the risk of bias ratings applied to the studies included in this stage of the review

Data Analysis

The aggregate quality of the body of evidence for each outcome was assessed using the GRADE evidence rating approach based on number and type of studies, study risk of bias, consistency of results across studies, precision and directness of evidence.

For the continuous outcomes of benefit of treatment of developmental delay such as language impairment, gross & fine motor skills, and adaptive functioning (socialization) we utilized change from baseline to immediate post-treatment data (means, standard deviations) and extracted data were meta-analyzed when appropriate. The DerSimonian and Laird random effects model with inverse variance (IV) method was utilized to generate the summary measures of effect in the form of mean difference (MD) if outcome was reported using a single outcome measure or standardized mean difference (SMD) if outcome was reported using multiple outcome measure.[36] For studies where the same outcome was reported using different outcome measures or scales, we selected the primary and validated outcome measure. A weighted composite score was computed for studies where primary outcome measure was reported using multiple sub-scales. MD and SMD were calculated using change from baseline data [i.e., mean difference between pre-treatment (baseline) and post-treatment (final/end-point) values along with the standard deviation (SD) for both intervention and control groups]. For studies that did not report SD, we calculated this value from the reported standard error (SE) of the mean, or from the 95% confidence intervals (CI) using equations provided in the Cochrane Handbook for Systematic Reviews of Interventions.[30] The Cochran's Q (α=0.05) was employed to detect statistical heterogeneity and I^2 statistic to quantify the magnitude of statistical heterogeneity between studies where I^2 30% to 60% represents moderate and I^2 50% to 90% represents substantial heterogeneity across studies.[37] Analyses were performed using Review Manager (Version 5.3) and GRADE pro software packages.[32, 38] When studies did not provide data necessary for pooling, results are described narratively.

Stage III –Test Properties

Search Strategy

For Stage III key question 5, we searched Medline, Embase, and PsycINFO with no beginning date limitations to September 16th, 2015 (Appendix 5). A grey literature search was also performed to search for Canadian-specific tests including the NDDS and the Rourke Baby Record (Rourke), using an Advanced Google search limited to Canada with the above specific search terms.

Inclusion and Exclusion Criteria

Population

For Stage III the population of interest was children aged 1-4 years of age who were not suspected of or at high risk of DD. High risk was defined as those born prematurely (gestational age less than 37 completed weeks at birth) or with low birth weight (birth weight less than 2, 500 g) and/or children with other known disorders that may be associated with or affect development. We also excluded studies of children over 4 years, studies of case finding in children in whom DD was suspected or who were at high risk for DD, and studies of screening for hearing or vision problems (as these are usually identified through specific hearing and vision screening tests).

Language

The published results of studies had to be available in either English or French.

Index Test

Screening with any test, tool, or questionnaire used to screen for DD in Stage I (including tools for specific domains, tools for general DD and tools for AD or ASD) was included. After consultation with our clinical expert, we also included tests, tools, or questionnaires currently in use in Canada: M-CHAT; Checklist for Autism in Toddlers (CHAT); Communication and Symbolic Behaviour Scales – Developmental Profiles of Infant-Toddler Checklist (CSBS-DP) and Infant-Toddler Checklist; Social Communication Questionnaire (SCQ); Early Screening of Autistic Traits Questionnaire (ESAT); Screening Test for Toddlers (STAT); Social Responsiveness Scale (SRS); NDDS; PEDS; Ages and Stages Questionnaire (ASQ); Bayley Infant Neurodevelopmental Screener (BINS); and Rourke.

Settings

Settings were limited to primary care settings and public health clinics.

Study Design

For the question on test properties, acceptable study designs included RCTs, cohort and case-control studies.

Reference Standard

To answer the test properties question the outcomes of interest included diagnosis of ASD, AD, DD or sub-domain of DD. There is no widely acknowledged gold standard test for the screening for DD. As such we included any study in which there was a diagnosis of DD by means of clinical or diagnostic evaluations using Bayley Scale of Infant Development (BSID) or BSID-II; Wechsler Preschool and Primary Scale of Intelligence (WPPSI); Vineland Adaptive Behavior Scale (VABS); Preschool Language scale (PLS); Autism Diagnostic Observation Scale (ADOS) or Autism Diagnostic Observation Scale Toddler Module (ADOS-T); Mullen Scales of Early Learning (MSEL); Battelle Developmental Inventory (BDI); or any combination of the above tests and/or clinical interview or judgment. Also included were studies in which the test(s) used for the diagnosis was not reported by the study authors.

Data Abstraction

For each study used to answer any KQ, review team members extracted data about the population, the study design, the intervention, the analysis and the results for outcomes of interest. For each study, one team member completed full abstraction (study characteristics, risk of bias assessment, outcome data) using electronic forms housed in a web-based systematic review software program.[29] A second team member then verified all extracted data and ratings; disagreements were resolved through discussion and/or third party consultation when consensus could not be reached.

Assessing Risk of Bias

Risk of bias for each study was assessed using Quality Assessment of Diagnostic Accuracy Studies 2nd Edition (QUADAS-II).[39] The purpose of the QUADAS tool is to determine the quality of the included primary diagnostic studies. Four domains were assessed including: 1) patient selection; 2) the index test used; 3) the reference standard applied; and 4) the flow of participants through the study and the timing of the delivery of the index test and reference standard. Risk of bias was assessed in each domain, and the first three domains were also rated on concerns regarding applicability. Signaling questions were used to assist reviewers in identifying items of concern related to study design that could impact interpretation of risk of bias.[40]

Data Analysis

The screening test properties data was extracted or calculated across studies from reported sensitivity, specificity and prevalence in 2 x 2 contingency tables (true positive, true negative, false positives and false negatives). Extracted test properties data was meta-analyzed where possible using exact binomial rendition[41] of the bivariate mixed-effects regression model modified for synthesis of screening test data. [42-44] Summary sensitivity, specificity, and the corresponding positive likelihood and negative likelihood ratios are derived as functions of the estimated model parameters. Areas under the SROC curves were used as a measure of the screening performance of the test and computed using empirical Bayes approach to fitting the hierarchical summary receiver operating curve (HSROC) model. [41, 43, 44] The summary estimates of negative and positive predictive values were estimated using summary estimates of sensitivity and specificity at a representative pre-test probability (pooled prevalence) obtained from included studies (see chapter 10, Section 10.2.3.Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy[45]). For outcomes where meta-analysis could not be performed, the sensitivity and specificity, likelihood ratios and positive and negative predictive values with 95% confidence intervals, were recalculated for each primary study from the contingency tables and reported descriptively.

For harms of DD screening such as false positive rate (over referral) was calculated as 1 - specificity and false negative rate as 1- sensitivity using data from either pooled estimates or 2 x 2 contingency tables. The studies were primarily grouped based on type of developmental delay such as ASD, AD, any DD and sub-domains of DD, and further sub-grouped based on screening test type such the MCHAT alone; CHAT; CSBS; ASQ; BINS; PEDS; and STAT

Chapter 3: Results

Summary of the Included Studies

The results section is separated into three sections to address the three stages of this review. We found no studies addressing the screening outcomes of cognitive function, incidence of mental health conditions, overall quality of life, survival, and functionality as an adult. We found no studies addressing treatment outcomes of academic performance; survival; and functionality as an adult. Overall, a total of 31 studies[46-76] were identified to answer the key questions that met the inclusion criteria for this review; two trials[46, 47] reported on the effectiveness of screening, no studies reported on harms of screening (Stage I); 5 systematic reviews[48-52] and no RCTs reported on the effectiveness of treatment, and 2 systematic reviews[49, 50] and no RCTS reported on the harms of treatment (Stage II); 3 RCTs[53-55] reported on the effectiveness of treatment on domain specific outcomes (Stage II addendum) and 17 studies[56-59, 61-63, 65, 67, 69-76] reported on test properties (Stage III) (Figures 2a-e: Search Results).

Results for Key Questions: Stage I (KQ1 and KQ2)

Summary of the Literature Search for Stage I (KQ1 and KQ2)

The search for key questions 1 and 2 located 8,104 unique citations screened at title and abstract; 400 were reviewed at full text (Figure 2a). Two studies were located and included from our search. Nineteen systematic reviews on related topics were identified by our team. The reference lists of these systematic reviews were also searched; no papers were added to our database as a result. See Table 1a for Characteristics of Included Studies. See Evidence Set 1 for detailed results.

KQ 1. What is the effectiveness of screening children aged 1 to 4 years without suspected DD to improve outcomes?

Referral Outcomes

One RCT provided evidence for referral rates and time to referral in children <30 months who were screened for DD using ASQ-II at 9 months, 18 months and 30 months.[46] This 2013 study took place in the United States of America (US) and included 2,103 mixed-gender children who were randomly allocated to the office support group (mean age 10.5 [standard deviation (SD) 8.2] months), no office support group (mean age 10.5 [SD 8.1] months) or usual care group (mean age 10.4 [SD 8.6] months). Those families allocated to the office support group met with trained office staff to complete the ASQ screening tool with the use of props; those families in the no office support group completed the ASQ without support of office staff or the use of props. A child was considered screen positive if they scored <2 SDs for age on any of the five developmental domains, and could be referred to early intervention (EI) services. Children in the control group who failed the usual care developmental screen (milestones consisting of 8-10

questions from 4 domains) could also be referred to EI services. For the outcome of referral rate to intervention, the screening arm with office support showed significantly more referrals to early intervention (19.9%) than the control group (10.2%) with a relative risk (RR) of 1.95 (95% CI 1.49 to 2.54). The absolute risk increase was 9.67%. The number needed to screen (NNS) for one child to be referred was 10 (95% CI 6 to 20). The referral rates were also significantly more (17.5%) for the screening without office support group (RR 1.71; 95% CI 1.30 to 2.25) as compared to the control group. The absolute risk increase was 7.24%. The NNS for one child to be referred was 14 (95% CI 8 to 33).

The authors found a 70% shorter time to referral (number of days=181) in the intervention group with office support (Rate Ratio of 0.30 [95% CI 0.19 to 0.48]), and a 64% shorter time to referral (number of days=234) for the intervention group without office support (Rate Ratio of 0.36 [95% CI 0.23 to 0.59]), both compared to the control group (number of days=467). The GRADE ranking for outcomes of time to referral and referral rates (for both screening with office support and screening without office support) was MODERATE. This study was downgraded on Indirectness due to participant age at entry under 12 months.

Academic Performance

One RCT provided data on academic performance in children screened for language delay.[47] This 2007 study, set in six geographic regions in the Netherlands, included 11,440 mixed gender children aged 15 months at the beginning of the study (mean age not reported). Intervention children were screened twice using the VroegTijdige Onderkenning Ontwikkelingsstoornissen (VTO) Language Screening instrument and control children received usual care (usual monitoring system by physicians). Children were screened at 18 and 24 months of age with a final score ranging from 0 to 7 (sum of scores on both screens). Children with a total score of ≤ 2 were referred for additional assessment to confirm language delay. Post-screening, the study did not offer an intervention and did not indicate whether children received interventions elsewhere. The study reported academic performance outcomes at age eight, such as: special school attendance; repeating a grade; repeating a grade because of language problems; being below the 10^{th} percentile of oral tests; being below the 10^{th} percentile of reading tests; and being below the 10^{th} percentile of spelling tests. The results showed no differences between groups with a RR of 0.71 (95% CI 0.48 to1.04) of attending a special school (Intervention=2.7%; Control=3.7%); an RR of 0.99 (95% CI 0.81 to1.21) of repeating a grade (Intervention=14.4%; Control=14.1%; an RR of 1.26 (95% CI 0.89 to1.80) of repeating a grade because of language problems in regular primary school (Intervention=6.1%; Control=4.9%); an RR of 0.88 (95% CI 0.63 to1.23) of being below the 10^{th} percentile of oral tests (Intervention=8.8%; Control=9.7%); an RR of 1.00 (95% CI 0.72 to1.40) of being below the 10^{th} percentile of reading tests in grade 2 (Intervention=4.7%; Control=4.7%); and an RR of 0.68 (95% CI 0.41 to1.13) of being below the 10^{th} percentile of spelling tests in grade 2 (Intervention=2.8%; Control=4.2%). The GRADE ranking for all outcomes for academic performance was LOW. This body of evidence was

downgraded for potential risk of bias due to insufficient information on allocation concealment and blinding of participants and on Imprecision due to effect estimate including null value.

KQ1a. What is the optimal interval for screening for DD?

Our search did not locate any studies meeting our inclusion criteria reporting on optimal interval for screening for DD.

KQ 2. What is the incidence of harms of screening children aged 1 to 4 years without suspected DD?

Our search did not locate any studies meeting our inclusion criteria reporting on harms. Please see Stage III results for data on false positives and false negatives.

Results for Key Questions: Stage II (KQ 3 and 4)

Summary of the Literature Search for Stage II (KQ3 and 4)

Two separate searches were performed to answer key questions 3 and 4 in Stage II. Our first search was limited to systematic reviews and yielded 1,887 unique citations that were screened at title and abstract; 187 were reviewed at full text (Figure 2b). Five systematic reviews were included. Because these systematic reviews focused on autism/ASD, we conducted a targeted search limited to RCTs on global or domain specific DD. This second search yielded 5,099 unique citations that were screened at title and abstract; 370 were reviewed at full text (Figure 2c). No RCTs meeting our inclusion criteria were found (see Evidence Set 2).

KQ 3. What is the effectiveness of treatment for children diagnosed with developmental delay to improve outcomes?

Outcomes for Systematic Reviews

A total of 5 systematic reviews[48-52] reported on the effectiveness of treatment for children for the outcomes of interest: five reported on cognitive function[48-52], two reported on quality of life[49, 50]. The population of interest in each of these systematic reviews was children with autism/ASD. No systematic reviews reporting on academic performance, incidence of mental health conditions, survival or functionality as an adult (including employment, criminality, and independence) were found. Three of the included systematic reviews[48-50] were rated as high methodological quality (9, 10 or 11/11). We also included two moderate methodological quality systematic reviews[51, 52] (each rated 8/11) as they were focused exclusively on children aged 1-6. See Table 4 for more details on the included systematic reviews. The systematic reviews included a variety of control conditions but we focused only on included studies that compared with a no intervention or usual care control. There were multiple overlapping studies reporting on cognitive function between two of the systematic reviews (see Table 4 for details).[49, 51] In this systematic review we only considered the unique RCTs or CCTs that met our inclusion criteria.

Cognitive Function

Five systematic reviews[48-52], with 13 unique included studies meeting our inclusion criteria[77-89] reported on the outcome of cognitive function among children with ASD; four focused on various types of behavioural interventions including early intensive behavioural interventions (EIBI)[49], applied behavioural interventions (ABI) and applied behavioural analysis (ABA)[51, 52] and interventions focused on training parents in management of their children's ASD symptoms and related functioning [48] and one systematic review examined the effects of acupuncture/acupressure plus conventional treatment.[50] All behavioural interventions reporting on cognitive outcomes included only children who met our age criteria (1 to 6 years of age). We did not include measurements of non-verbal IQ for this outcome.

The results of the systematic reviews on behavioural interventions found significant differences between groups in the two reviews measuring cognitive function with composite IQ.[49, 51] One systematic review of ABAs (3 unique studies; n=129) found large effect sizes with a median SMD of 1.34 (0.60 to 2.08).[51] Another systematic review (5 unique studies; n=172) of EIBIs completed a meta-analysis and reported a medium effect size of 4 pooled studies, SMD 0.76 (95% CI 0.040 to 1.11; $I^{2=}$21%).[49] One additional included study (n=28) was not pooled (g=0.74)[82].

One behavioural systematic review (1 study; n=24[77]) reported non-significant changes in cognitive outcomes and developmental/intellectual gains.[48] One behavioural systematic review met our inclusion criteria, but did not include any unique studies that were not considered above.[52]

One systematic review on acupuncture/acupressure plus conventional treatment (4 studies[86-89], n=179) included children aged 3 to 18 years old.[50] Three studies[86-88] reported on acupuncture and each reported no significant difference between intervention and control groups using either Griffiths Mental Developmental Scale or Leiter International Performance Scale-Revised. One study[88] found greater improvement in the intervention group using the Chinese Version of Psychoeducational Profile (MD 10.75, 95% CI 3.82 to 17.68, P=0.002). One study[89] reported on acupressure and found no significant differences in developmental aspects (no overall score provided).

Quality of Life

Two systematic reviews examined quality of life measurements in children.[49, 50] One systematic review on acupuncture/acupressure found no studies reporting standardized quality of life measures.[50] One systematic review (5 studies[78-82], n=199) on early EIBI with children under 6 years old pooled four studies and reported a statistically significant difference (SMD 0.55 [95% CI 0.24 to 0.87], p<0.001) in daily living skills. One study (n=28) was not pooled (g=-0.03).[49]

Outcomes for RCTs

Our search did not locate any RCTs meeting our inclusion criteria reporting on the effectiveness of treatment for DD for the outcomes of interest.

KQ4. What is the incidence of harms of treatment for children diagnosed with DD?

Outcomes for Systematic Reviews

Harms

Two systematic reviews reported on harms of treatment for ASD/Autism using EIBI[49] and acupuncture interventions.[50] The participants of interest for all systematic reviews were children with ASD. The age of participants ranged from 3 to 18 years; one systematic review included children with a mean age under 6.[49]

A systematic review on EIBI (5 studies[78-82]; n=200) reported no deterioration on primary outcomes or adverse events as a result of treatment in any of the studies.[49] A systematic review on acupuncture or acupressure (9 studies; n=357[86-94]) plus conventional treatment found two studies[86, 93] reporting initial crying for fear and possible pain, one study[87] reporting superficial bleeding or crying and irritability occurring during acupuncture treatment and one study[94] found worsening of sleep patterns. [50] There was no information on the number of participants experiencing these effects or whether there was a difference between treatment and control groups. Three studies[88, 90, 92] on acupuncture and two studies[89, 91] (n=32) on acupressure plus conventional treatment did not report any harms.

Outcomes for RCTs

Our search did not locate any RCTs meeting our inclusion criteria reporting harms.

Results for Key Questions: Stage II Addendum (AKQ1 and AKQ2)

Summary of the Literature Search for Stage II Addendum (AKQ1 and AKQ2)

The search located 5,099 unique citations; we also identified one additional citation by searching a recently published on topic systematic review.[95] Of the 5,099 unique citations screened at title and abstract; 370 were screened at full text. At full text, 367 citations were excluded. Three RCTs were included. Please see Figure 2d for details. See Table 1b for Characteristics of Included Studies. See Evidence Set 3 for detailed results.

AKQ1. What is the effectiveness of treatment for children diagnosed with DD to improve outcomes in gross and fine motor skills, language impairment, adaptive functioning, intellectual disability (IQ), learning disability (academic testing) and academic underachievement?

A total of three RCTs reported on the effectiveness of treatment for children for the outcomes of interest.[53-55] Of the three studies, three reported on changes in language impairment [53-55] and one reported on changes in social and personal activities of daily living (adaptive functioning).[55] All of the studies included mixed gender samples, with mean ages ranging from 24.7 to 66.68 months. One of the studies[53] had an intervention length < 6 months while two of the studies[54, 55] had an intervention length > 6 months. Follow-up was immediate post-intervention in all studies. Two of the studies[54, 55] were conducted in Europe, and one study[53] was conducted in Canada.

Language impairment Outcomes

Three RCTs (n=117 Intervention; n=122 Control) provided data for the outcome of language impairment. We included studies with a no treatment comparator – one of the studies used a waiting list comparison group[54], one used a watchful waiting comparison group[55] and one used a no intervention comparator.[53] The language impairment was assessed using DIBELS, SETK-2, and auditory comprehension & expressive language. The pooled estimate across three studies showed a significant improvement in language impairment for intervention group as compared to controls with an SMD of 0.81 (95% CI 0.02 to 1.6; I^2= 84%).The GRADE ranking for the outcome of language delay was MODERATE. This body of evidence was downgraded for potential risk of bias due to insufficient information on sequence generation, allocation concealment and a high risk of bias associated with blinding, selective outcome reporting and other risk of bias (baseline characteristics, pre-hoc power analysis, sample size < 30 per arm).

Adaptive Functioning Outcomes

One RCT (n=71 Intervention; n=84 Control) provided data for the outcome of social and personal activities of daily living (adaptive functioning) with a watchful waiting comparison group.[55] The adaptive functioning or socialization was assessed using VABS. The effect estimate showed no difference between intervention and control groups with a mean difference (MD) of 0.60 (95% CI -3.05 to 4.25).The GRADE ranking for the outcome of adaptive functioning was LOW. This body of evidence was downgraded for potential risk of bias due to insufficient information on sequence generation and high risk of bias associated with blinding, and Imprecision due to effect estimate including null value.

No studies reporting on gross and fine motor skills or performance and cognition outcomes using no treatment control groups or usual care control groups were identified.

AKQ2. What is the incidence of harms of treatment for children diagnosed with DD?

No studies were found that reported harms of treatment interventions.

Results for Key Questions: Stage III (KQ 5)

Summary of the Literature Search for Stage III (KQ5)

The search for the key questions located 1,815 unique citations which were screened at title and abstract; 404 were reviewed at full text (Figure 2e). Fourteen systematic reviews were identified by our team. The reference lists of on-topic systematic reviews were searched; no papers were

added to our database as a result. We also identified one additional study through a hand search of the grey literature. After full-text screening 21 studies were included for key question 5 of Stage III.

KQ5. What are the sensitivity, specificity, PPV, NPV, and likelihood ratios of the various screening tests to assess DD in children aged 1 to 4 years who are not already suspected of having developmental delay?

We assessed these papers with QUADAS II (Table 5). Across the four domains the included cohort and case control studies varied in terms of risk of bias and applicability ratings. The reference standard was primarily clinical or diagnostic evaluation. The tests included in this section are: the M-CHAT, CHAT, CSBS, SCQ, BINS, ASQ, PEDS and STAT. We did not find any studies on test properties of NDDS or Rourke meeting our inclusion criteria.

Four domains using QUADAS tool (QUADAS-II) across 17 test properties studies were assessed including: 1) patient selection; 2) the index test used; 3) the reference standard applied; and 4) the flow of participants through the study and the timing of the delivery of the index test and reference standard. For patient selection domain; 3 (17.6%) studies were rated as low risk, 6 (35.3%) as unclear risk and 8 (47.1%) as high risk for concerns regarding risk of bias; while 13 (76.5%) studies were rated as low risk, 1 (5.9%) as unclear risk, and 3 (17.6%) as high risk for concerns regarding applicability. For index test domain; 8 (47.1%) studies were rated as low, 8 (47.1%) as unclear risk and 1 (5.9%) as high risk for concerns regarding risk of bias; while all (100%) studies were rated as low risk for concerns regarding applicability. For reference standard domain; 7 (41.2%) studies were rated as low, 8 (47.1%) as unclear and 2 (11.8%) as high risk for concerns regarding risk of bias; while 16 (94.1%) studies were rated as low risk and 1 (5.9%) as unclear risk for concerns regarding applicability. For risk of bias due to flow of participants through the study and the timing of the delivery of the index test and reference standard domain; 8 (47.1%) studies were rated as low risk, 3 (17.6%) as unclear risk and 6 (35.3%) as high risk for concerns regarding risk of bias.

Eleven observational studies (cohort and case-control) were found reporting on the test properties of screening tools including MCHAT, [56, 57, 63, 69-72, 74, 75] SCQ, [57, 58, 73] and PEDS[70]used to screen Autism Spectrum Disorder. Four observational studies (cohort and case-control) were found reporting on the test properties of screening tools including MCHAT,[69] CHAT[67, 68] and STAT,[65] used to screen Autism disorder. Four observational studies (cohort and case-control) were found reporting on the test properties of screening tools including ASQ[59, 61, 62, 76] and PEDS[59] used to screen any Developmental Delay. For a more detailed overview of the test properties for each of these screening tests please see Evidence Set 4. Only the results for test properties data on screening tests examined in multiple studies or large cohort studies is reported here.

Autism Spectrum Disorders (ASD)

Nine cohort studies with a total sample of 70,816 using MCHAT as a screening tool for ASD were found. [56, 57, 63, 69-72, 74, 75] The reference standard was primarily clinical or diagnostic evaluation. The bivariate pooled analysis for MCHAT showed a sensitivity of 78.0% (95% CI 64.0% to 88.0%) and a specificity of 69.0% (95% CI 47.0% to 85.0%). The PPV and NPV using pooled estimate for prevalence across nine studies were 30.3% (95% CI 17.3% to 50.4%) and 94.8% (95% CI 88.3% to 97.6%) respectively. The area under curve was 0.81 (95% CI 0.77 to 0.84). The false positive rate was 31.0% (95% CI 15.0% to 53.0%) and the false negative rate was 22.0% (95% CI 12.0% to 36.0%).

Small cohort and case-control studies (sample size < 100) also assessed other screening tools for ASD such as SCQ [57, 58, 73] and PEDS [70] and the data is presented in Evidence Set 4.

Autism Disorder (AD)

One large cohort study with a total sample of 1,851 using MCHAT as a screening tool for AD was found.[69] The study showed a sensitivity of 62.5% (95% CI 35.4% to 84.8%) and a specificity of 83.2% (95% CI 81.4% to 84.9%). The PPV and NPV were 3.1% (95% CI 1.5% to 5.7%) and 99.6% (95% CI 99.2% to 99.9%) respectively. The false positive rate was 16.8% (95% CI 15.1% to 18.6%) and the false negative rate was 37.5% (95% CI 15.2% to 64.6%).

One small case-control study assessed STAT[65] and one assessed CHAT [67] as screening tool for AD and the test properties data is presented in Evidence Set 4.

Developmental Delay (DD)

Four cohort studies with a total sample of 1,001 using ASQ as a screening tool for DD were found.[59, 61, 62, 76] The reference standard varied across studies therefore results are presented descriptively. The studies showed a median sensitivity of 55.0% (range 47.1% to 66.7%) and a median specificity of 86.0% (range 38.6% to 94.3%). The median PPV and NPV were 41.4% (range 23.2% to 71.4%) and 84.9% (range 70.8% to 95.6%) respectively. The median false positive rate was 14% (range 5.7% to 61.4%) and the median false negative rate was 45.0% (range 33.3% to 52.9%).

One fair size cohort study with a total sample of 331 using PEDS as a screening tool for DD was found.[59] The study showed a sensitivity of 41.1% (95% CI 24.7% to 59.3%) and a specificity of 89.3% (95% CI 85.1% to 92.5%). The PPV and NPV were 30.4% (95% CI 17.7% to 45.8%) and 92.9% (95% CI 89.3% to 95.7%) respectively. The false positive rate was 10.7% (95% CI 7.5% to 14.9%) and the false negative rate was 58.9% (95% CI 40.7% to 85.3%).

Results for Contextual Questions

We located 583 studies in our contextual questions searched; 12 studies were included to answer the subgroups question.

CQ1. What is the cost-effectiveness and feasibility of screening for DD in preschool children aged 1 to 4 years?

Our search did not locate any studies on the question of cost-effectiveness and feasibility of screening for DD in preschool children aged 1 to 4 years.

CQ2. What are parent or primary caregiver values and preferences for screening for DD in preschool children aged 1 to 4 years?

Our search did not locate any studies on the question of primary caregiver values and preferences for DD screening in preschool children aged 1 to 4 years.

CQ3. What is the evidence for a higher burden of disease, a differential treatment response, differential performance of screening for DD, or barriers to implementation of screening for DD in subgroups? Subgroups include: Aboriginal population, rural or remote populations, low socioeconomic status, drug or alcohol dependency, or other ethnic populations.

We found no studies with data relevant to the Canadian context addressing rural or remote populations, drug or alcohol dependency or other ethnic populations.

Aboriginal populations

One Canadian study [96] compared a Canadian First Nations population to the general U.S. population with the aim of evaluating the applicability of the ASQ in Mohawk children. Seventeen teachers and parents of 282 Mohawk children between 9 and 66 months participated between 2006 and 2009 at the Child and Family Center in Mohawk Territory, Quebec. The authors found that there was little difference between the U.S. general population and Canadian Mohawk children on scores on the ASQ domains. All correlations on the domains were significant ($r=0.46-0.87$, $p<0.01$), showing that scores on the ASQ were very similar between the two groups. The authors conclude further research into the use of the ASQ tool with Mohawk children is needed as they caution that this sample is not representative of the wider Mohawk community.

Socioeconomic status

There were mixed results in the five studies that examined the association between socioeconomic status (SES) and DD.[97-101] Two studies, one of 136 children between the ages of 9 and 12 months in Brazil[97] and one of 206 families from a largely urban community in the

US,[100] looked at SES and language delay. Low SES was found to be positively related in the first study ($p<0.05$),[97] but unrelated in the latter study ($p>0.05$)[100] to indicators of language delay.

A group of studies investigated the relationship between parental education and DD. One American study[98] looked at the National Survey of Children's Health (NSCH) data and randomly selected 91, 642 children ages 18 months to 5 years old. Parent/guardian education of less than high school was related to greater odds of possible (Odds Ratio [OR]: 1.67; 95% CI 1.09 to 2.55) and probable (OR 3.52; 95% CI 2.46 to 5.03) DD. Conversely, in another US study, 3,680 children with ASD were identified within a sample of 557,689 eight year old children living in one of multiple study areas.[99] In this study, higher adult educational achievement, defined as the percentage of adults with a bachelor's degree, was more likely in children with ASD (30.3%) versus those children without the diagnosis (24.8%).[99] Two additional studies investigated the relationship between parental education and DD and found it unrelated to results on a DD screening test in 136 children aged 9 to 12 months in one Brazilian study[97] and having no effect on the likelihood of 16, 223 children aged 10 month to 4 year olds receiving a developmental questionnaire at health check- ups ($p<0.05$) in one US study.[101]

The SES subcategory of income was generally negatively associated with DD, although there were some mixed results. One study using NSCH data of 91, 642 children ages 18 months to 5 years old, showed that as incomes decreased there were significant increase in the odds of having possible (OR 1.13; 95% CI 1.03 to 1.24) and probable (OR 1.46; 95% CI 1.33 to 1.61) DD.[98] Another study showed that 136 children in Brazil between the ages of 9 and 12 months who lived in a borrowed space had a greater chance of scoring lower on the Bayley Scale Receptive Communication subscale compared to renting or owning a home ($p=0.009$).[97] Those owning a flush toilet ($p=0.02$ for cognitive, $p=0.03$ for receptive communication), radio ($p=0.03$), television ($p=0.02$) or cell phone ($p=0.02$ for cognitive, $p=0.03$ for gross motor) also had higher scores on some subscales of the BSID. Per capita family income was found to be unrelated to scores on the Receptive Communication subscale ($p>0.05$). In a US study of 557,689 children 8 years of age across all race groups, there were fewer ASD cases in a high poverty area compared to a control group ($p>0.05$).[99] In this study, median household income was higher across all race groups in ASD cases compared to controls ($p>0.05$).[99]

Finally, only one study examined the potential role of SES as a barrier to developmental screening: an analysis of the NSCH data of 10 month to 4 year olds revealed an insignificant relationship between poverty and accessing a developmental questionnaire in the American health system (OR 0.86; 95% CI 0.61 to 1.23 for <100% of poverty line).[101]

Chapter 4: Discussion, Limitations and Conclusion

Discussion

Effectiveness of Screening for DD

The evidence on the effectiveness of screening for DD in improving cognitive, academic and adaptive functioning outcomes in children 1-4 years old is scant. We found one study that reported higher and earlier intervention rates among the children screened for DD.[46] Referral to intervention is an intermediate outcome and does not necessarily reflect better outcomes. Long-term outcomes related to the early intervention programs were not reported. Another Dutch study reported on academic performance outcomes in children screened at 15 months for speech and language delay.[47] Although this study provided longer term outcomes with a follow-up of 81 months, screening did not show a significant improvement in academic performance, including attending a special school; repeating a grade; being below the 10^{th} percentile of oral tests; being below the 10^{th} percentile of reading tests; and being below the 10^{th} percentile of spelling tests at this time point. Ideally, the intermediate outcome of early referral leads to improved long term outcomes. There is no evidence that would confirm this. Furthermore, our search did not uncover any evidence on the harms associated with screening for DD.

Effectiveness of Treatment for DD

We found no systematic reviews reporting on academic performance, survival or functionality as an adult. We found no systematic reviews reporting on any outcomes of interest for global or domain specific DD. Evidence from the systematic reviews on treatment for autism/ASD are inconclusive on cognitive outcomes (IQ). Of the four systematic reviews focused on behavioural interventions, two reviews found significant differences between intervention (EIBI and ABA) and control in cognitive outcomes, while two reviews (parent-mediated ASD intervention and ABI) found no significant differences. There were varying sample sizes and effect sizes. We are unable to make a conclusive statement on the effectiveness of treatment of ASD.

Our original targeted search for RCTs on global or domain specific DD found no studies meeting our inclusion criteria. In light of absence of RCT evidence for our initial strategy to evaluate the effectiveness of DD treatment for children based on important and critical outcomes such as cognitive function, academic performance, incidence of mental health conditions, overall quality of life, survival, and functionality as an adult; we considered other relevant outcomes mentioned in literature i.e. gross and fine motor skills, language impairment, adaptive functioning, intellectual disability (IQ), learning disability (academic testing) and academic underachievement. We found three RCTs for the outcome of language impairment, one RCT for adaptive functioning and no RCTs for remaining outcomes. One RCT examining the improvement in adaptive functioning or socialization showed no difference in effect between intervention and control groups. For the outcome of language impairment (n=239), the pooled

estimate (SMD 0.81, 95% CI 0.02 to 1.6) and two out of three included RCTs (n=37 and n=47 respectively) showed significant benefit of DD treatment for young children as compared to controls.

Test Properties for Screening Tools for DD

Our search for test properties did not yield any studies on NDDS or Rourke. For screening for ASD, the evidence from large and fair size cohort studies showed that apart from MCHAT, all other screening tools such as CHAT, CSBS, SCQ and PEDS showed poor performance with low sensitivity (range: 35.1% to 69.2%). This low sensitivity implies a high level of false negatives and a large proportion of ASD cases would not be identified at population level. Both sensitivity and specificity are important metrics of a screening tool but a high sensitivity is most desirable in order to rule out risk of ASD more accurately. The pooled analysis across 9 cohort studies for MCHAT as a screening tool for ASD showed modest sensitivity of 78% (64% to 88%) and acceptable level of discrimination with AUC 0.81 (0.77 to 0.84). [56, 57, 63, 69-72, 74, 75] However, the specificity of 69% (47% to 85%) was relatively low which implies a high false positive rate and further assessment on larger numbers without the actual disorder at population level. The evidence for MCHAT as a screening tool is further limited by high variability across studies in terms of prevalence, length of follow-up, samples size and selection, study quality, and diagnostic evaluation and case ascertainment.

For screening for AD, DD and PDD the current evidence from large and fair size cohort studies showed that screening tools such as MCHAT, CHAT, SCQ, BINS and PEDS performed poorly as screening tools with low sensitivity (range: 31.8% to 62.5%), however most of the evidence is from single study outcomes and hence limited by paucity of studies reporting the test properties outcomes.

Contextual Questions

The contextual questions search found no evidence about cost-effectiveness and feasibility of screening or parent or primary caregiver preferences and values and limited information on burden of delay, with no studies relevant to the Canadian context addressing remote or rural populations, alcohol or drug dependent population or other ethnic populations. The only study on burden of delay in Canadian Aboriginal populations compared a Mohawk population in Canada to a U.S general population sample and found little difference between the U.S. general population and Canadian Mohawk children on scores on the ASQ domains. The authors concluded that their findings support further study into the use of the ASQ tool with Mohawk children. However, caution should be used when applying these findings to a Canadian context.

Implications for future research

Further research, specifically in cognitive function, academic and adaptive function, as well as the potential harms of screening is needed to provide more conclusive evidence regarding the effectiveness of screening for DD in children 1 to 4 years of age. We found no peer-reviewed evidence on some commonly used tools in Canada such as NDDS or Rourke. Robust evidence is needed to determine whether the continued use of these tools is clinically relevant and appropriate. There are a number of questions of interest for this review which could not be answered due to a lack of evidence. More research is needed to help answer the questions of what screening intervals are most effective, result in the least harm, and help inform clinical practice regarding screening in this young population.

There is also lack of evidence on treatment of global/domain specific DD (effectiveness or harms) for outcomes of interest (cognitive function, academic performance, incidence of mental health conditions, overall quality of life, survival, functionality as an adult) in both RCTs and high quality systematic reviews.

For the domain specific treatment outcomes (gross and fine motor skills, language impairment, adaptive and academic functioning), the generalizability and applicability of the findings at a population level is limited by multiple factors, such as the paucity of evidence, small sample sizes, variability across trials in intervention mode of delivery (i.e. individual or parent based), intervention delivery experts (i.e. research staff, speech language pathologists, and pediatric neurologists), intensity and content of intervention, length of follow-up, age of children, and outcomes measures used to assess language impairment. Overall, the current evidence offers little information and direction on specific factors and elements associated with efficacious interventions and warrant the need of future high quality research with a more standardized approach to affirm these findings.

The current evidence on various tools for screening for ASD, AD, DD and PDD is limited and does not warrant their use in routine surveillance for these disorders in early childhood at population level. MCHAT as screening tool for ASD provided some promising results in terms of modest sensitivity and acceptable level of discrimination but current evidence is limited by high variability across studies, therefore future high quality research evaluating the accuracy of MCHAT as screening tool with a more standardized approach and utilizing rigorous standards, adequate sample size and follow-up is needed.

Limitations

There is one screening study that reports on the intermediate outcome of referrals. Due to the fact that there are no long-term outcomes in this study, we cannot draw conclusions about the effectiveness of screening for referral for treatment. This review also focused only on the effectiveness of interventions on the specific outcomes of cognitive functioning, academic

performance, incidence or symptoms of mental health conditions, overall quality of life, survival and functionality as an adult. Several of the interventions identified in the process of this review target other developmental or symptom domains such as language ability or ASD symptoms, and their effectiveness with respect to these outcomes was not captured in this review. We are limited by the outcomes of interest for this review. There is also insufficient evidence to answer the question of the optimal interval and harms of screening for DD. Publication bias and methodological inconsistency could not be assessed due to lack of evidence.

Limitations of systematic reviews included in the treatment section (Stage II) included lack of reporting of individual study sample sizes in the included studies, limited outcomes (not all outcomes of interest to this review were reported), and varied ages of participants. Furthermore, the results of the systematic reviews must be interpreted with caution, as authors often expressed concern for the methodological quality of their included studies. Although methodological quality of the systematic reviews was undertaken in this review, we did not assess the quality of the individual included studies within the systematic reviews and therefore we are unable to incorporate these concerns into our findings. Limitations of the RCTs found in the treatment section addendum included risk of bias, inadequate sample sizes and small number of studies for the RCTs. Both RCTs and systematic reviews limited the population to children with a diagnosis rather than children with symptoms. The search dates for the treatment section of the review may have also limited the results.

For test properties there is a lack of gold standard and even the clinical diagnosis that was often used as the reference standard was not applied consistently as a different battery of tests was used. Due to these limitations, we suggest caution when interpreting the meaning of the results. Finally, for this review, only publications in English and French were considered for inclusion.

Conclusion

The evidence on screening for developmental delay in children aged 1 to 4 years of age without suspected DD to improve cognitive, educational and adaptive functioning outcomes is inconclusive. Further research on effectiveness and harms with longer term outcomes is needed to inform decisions about screening and screening intervals. Indirect evidence on treatment is also mixed and further high quality RCT research is needed.

References

1. Feldman W. Well-baby care in the first 2 years of life. In: Examination Canadian Task Force on the Periodic Health, editor. *Canadian Clinical Guide to Clinical Preventive Health Care*. Ottawa, ON: Health Canada; 1994. p. 258-66.

2. Feightner JW. Preschool screening for developmental problems. In: Examination Canadian Task Force on Periodic Health, editor. *The Canadian guide to clinical preventive health care*. Ottawa, ON: Health Canada; 1994. p. 289-96.

3. Denver Developmental Materials Inc. [Internet]. Denver developmental materials, Inc. Available from: http://denverii.com/.

4. Bellman M, Byrne O, Sege R. Developmental assessment of children. BMJ. 2013; 346:e8687. Available from: http://www.bmj.com/content/346/bmj.e8687.

5. Poon JK, LaRosa AC, Pai GS. Developmental delay timely identification and assessment. 2010; 47(5):415-22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20519787.

6. American Psychiatric Association. Diagnostic and statistical manual of mental disorders 4th Edition ed. Washington, DC: American Psychiatric Press Inc.; 2000.

7. Statistics Canada. A profile of disability in Canada 2001. Ottawa, ON; 2002 2003. Catalogue no. 89-577-XIE. Available from: http://www.statcan.gc.ca/pub/89-577-x/pdf/4228016-eng.pdf.

8. Canada EaSD. Disability in Canada: A 2006 Profile. Ottawa, ON; 2013. Available from: http://www.esdc.gc.ca/eng/disability/arc/disability_2006.shtml.

9. Blanchard LT, Gurka MJ, Blackman JA. Emotional, developmental, and behavioral health of American children and their families: a report from the 2003 National Survey of Children's Health. Pediatrics. 2006; 117(6):e1202-12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16740820.

10. Ontario Children's Health Network, Ontario College of Family Physicians, Services Ontario Ministry of Children and Youth. Getting it right at 18 months...making it right for a lifetime. Report of the expert panel on the 18 month well baby visit. 2005. Available from: http://www.children.gov.on.ca/htdocs/English/documents/topics/earlychildhood/getting_it_right_18_months.pdf.

11. National Center on Birth Defects and Developmental Disabilities [Internet]. Developmental disabilities: facts about developmental disabilities. Atlanta, GA: Centers for Disease Control. Available from: http://www.cdc.gov/ncbddd/developmentaldisabilities/facts.html.

12. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM). 5th Edition ed. Arlington, VA: American Psychiatric Publishing; 2013.

13. Palomo-Alvarez C, Puell MC. Relationship between oculomotor scanning determined by the DEM test and a contextual reading test in schoolchildren with reading difficulties. Graefes Arch Clin Exp Ophthalmol. 2009; 247(9):1243-9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/?term=19347678.

14. Mackrides PS, Ryherd SJ. Screening for developmental delay. Am Fam Physician. 2011; 84(5):544-9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21888305.

15. Williams R, Clinton J, Bennett S, Hertzman C, Leduc D. Getting it right at 18 months: in support of an enhanced well-baby visit. Paediatr Child Health. 2011; 16(10):647-50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23204907.

16. Anderson LM, Shinn C, Fullilove MT, Scrimshaw SC, Fielding JE, Normand J, et al. The effectiveness of early childhood programs: a systemdevelopment atic review. Am J Prev Med. 2003; 24(Suppl 3):32-46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12668197.

17. Trefler D. Quality is free: A cost-benefit analysis of early child development initiatives. Paediatr Child Health. 2009; 14(10):681-4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21119819.

18. Whalley LJ, Deary IJ. Longitudinal cohort study of childhood IQ and survival up to age 76. BMJ. 2001; 322(7290):819. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11290633.

19. NDDS Canada [Internet]. Nipissing district development screen. North Bay, ON: Nipissing District Developmental Screen Intellectual Property Association; 2011. Available from: http://www.ndds.ca/canada.html.

20. U. S. Preventive Services Task Force. Screening for speech and language delay in preschool children: recommendation statement. Pediatrics. 2006; 117(2):497-501. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16452370.

21. Siu AL, on behalf of the US Preventive Services Task Force. Screening for Speech and Language Delay and Disorders in Children Aged 5 Years or Younger: US Preventive Services Task Force

Recommendation Statement. Pediatrics. 2015; 136(2). Available from: http://pediatrics.aappublications.org/content/136/2/e474.full.pdf.

22. SIGN Scottish Intercollegiate Guidelines Network. Assessment, diagnosis and clinical interventions for children and young people with autism spectrum disorders. A national clinical guideline. Edinburgh, UK; 2007. Available from: http://www.sign.ac.uk/pdf/sign98.pdf.

23. Robins DL, Fein D, Barton ML, Green JA. The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. J Autism Dev Disord. 2001; 31(2):131-44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11450812.

24. U.S. Preventive Services Task Force [Internet]. Update Summary: Autism spectrum discorder in young children: Screening. 2013. Available from: http://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryDraft/autism-spectrum-disorder-in-young-children-screening.

25. Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics. 2006; 118(1):405-20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16818591.

26. Johnson CP, Myers SM. Identification and evaluation of children with autism spectrum disorders. Pediatrics. 2007; 120(5):1183-215. Available from: http://pediatrics.aappublications.org/content/120/5/1183.full.pdf+html.

27. King TM, Tandon SD, Macias MM, Healy JA, Duncan PM, Swigonski NL, et al. Implementing developmental screening and referrals: lessons learned from a national project. Pediatrics. 2010; 125(2):350-60. Available from: http://pediatrics.aappublications.org/content/125/2/350.

28. Sampson M, McGowan J, Lefebvre C. PRESS: Peer Review of Electronic Search Strategies. Ottawa, ON; 2008. Available from: https://www.cadth.ca/media/pdf/477_PRESS-Peer-Review-Electronic-Search-Strategies_tr_e.pdf.

29. Distiller (DistillerSR systematic review software) [computer program]. Ottawa, ON: Evidence Partners. Available from: http://systematic-review.net/.

30. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]: The Cochrane Collaboration; 2011. www.cochrane-handbook.org.

31. GRADE working group. Published place unknown: GRADE working group. Available from: http://www.gradeworkinggroup.org/.

32. GRADEpro. [Computer program]. Version 3.2 for Windows. Available from: http://ims.cochrane.org/revman/other-resources/gradepro/download.

33. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008; 336(7650):924-6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/?term=18436948.

34. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007; 7:10. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1810543/pdf/1471-2288-7-10.pdf.

35. Cohen J. Statistical power analysis for the behavioral sciences: Academic press; 2013.

36. Wallis SA. Binomial confidence intervals and contingency tests: mathematical fundamentals and the evaluation of alternative methods. J Quant Linguist. 2013; 20(3):178-208. Available from: http://www.ucl.ac.uk/english-usage/staff/sean/resources/binomialpoisson.pdf.

37. The Cochrane Statistical Methods Group. Analysing data and undertaking meta-analyses. In: Deeks J. J., Higgins J. P., Altman D. G., The Cochrane Statistical Methods Group, editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 510 [updated March 2011]*: The Cochrane Collaboration; 2011.

38. Review Manager (RevMan) [Computer Program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration. Available from: http://tech.cochrane.org/revman/download.

39. University of Bristol [Internet]. QUADAS-2. Bristol, UK: University of Bristol. Available from: http://www.bris.ac.uk/quadas/quadas-2/.

40. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011; 155(8):529-36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22007046.

41. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. J Clin Epidemiol. 2006; 59(12):1331-2; author reply 2-3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17098577.

42. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med. 2002; 21(4):589-624. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11836738.

43. Van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. Stat Med. 1993; 12(24):2273-84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7907813.

44. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005; 58(10):982-90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16168343.

45. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Cochrane handbook for systematic reviews of diagnostic test accuracy. 2010. Available from:

46. Guevara JP, Gerdes M, Localio R, Huang YV, Pinto-Martin J, Minkovitz CS, et al. Effectiveness of developmental screening in an urban setting. Pediatrics. 2013; 131(1):30-7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23248223.

47. van Agt HME, van der Stege HA, de Ridder-Sluiter H, Verhoeven LTW, de Koning HJ. A cluster-randomized trial of screening for language delay in toddlers: effects on school performance and language development at age 8. Pediatrics. 2007; 120(6):1317-25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18055682.

48. Oono IP, Honey EJ, McConachie H. Parent-mediated early intervention for young children with autism spectrum disorders (ASD). Cochrane Database Syst Rev. 2013; (4):CD009774. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23633377.

49. Reichow B, Barton EE, Boyd BA, Hume K. Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD). Cochrane Database Syst Rev. 2012; (10):CD009260. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23076956.

50. Cheuk DK, Wong V, Chen WX. Acupuncture for autism spectrum disorders (ASD). Cochrane Database Syst Rev. 2011; (9):CD007849. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21901712.

51. Virués-Ortega J. Applied behavior analytic intervention for autism in early childhood: meta-analysis, meta-regression and dose-response meta-analysis of multiple outcomes. Clin Psychol Rev. 2010; 30(4):387-99. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20223569.

52. Spreckley M, Boyd R. Efficacy of applied behavioral intervention in preschool children with autism for improving cognitive, language, and adaptive behavior: a systematic review and meta-analysis. J Pediatr. 2009; 154(3):338-44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18950798.

53. Hund-Reid C, Schneider P. Effectiveness of phonological awareness intervention for kindergarten children with language impairment. Can J Speech Lang Pathol Audiol. 2013; 37(1):6-25. Available from: http://cjslpa.ca/download.php?file=2013_CJSLPA_Vol_37/No_01_1-127/Hund-Reid_Schneider_CJSLPA_Spring_2013.pdf.

54. Buschmann A, Jooss B, Rupp A, Feldhusen F, Pietz J, Philippi H. Parent based language intervention for 2-year-old children with specific expressive language delay: a randomised controlled trial. Arch Dis Child. 2009; 94(2):110-6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18703544.

55. Glogowska M, Roulstone S, Enderby P, Peters TJ. Randomised controlled trial of community based speech and language therapy in preschool children. BMJ. 2000; 321(7266):923-6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11030677.

56. Robins DL, Casagrande K, Barton M, Chen C-MA, Dumont-Mathieu T, Fein D. Validation of the modified checklist for Autism in toddlers, revised with follow-up (M-CHAT-R/F). Pediatrics. 2014; 133(1):37-45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24366990.

57. Corsello CM, Akshoomoff N, Stahmer AC. Diagnosis of autism spectrum disorders in 2-year-olds: a study of community practice. J Child Psychol Psychiatry. 2013; 54(2):178-85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22905987.

58. Maljaars J, Noens I, Scholte E, van Berckelaer-Onnes I. Evaluation of the criterion and convergent validity of the Diagnostic Interview for Social and Communication Disorders in young and low-functioning children. Autism. 2012; 16(5):487-97. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21690082.

59. Limbos MM, Joyce DP. Comparison of the ASQ and PEDS in screening for developmental delay in children presenting for primary care. J Dev Behav Pediatr. 2011; 32(7):499-511. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21760526.

60. Pierce K, Carter C, Weinfeld M, Desmond J, Hazin R, Bjork R, et al. Detecting, studying, and treating autism early: the one-year well-baby check-up approach. J Pediatr. 2011; 159(3):458-6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/?term=21524759.

61. Gollenberg AL, Lynch CD, Jackson LW, McGuinness BM, Msall ME. Concurrent validity of the parent-completed Ages and Stages Questionnaires, 2nd Ed. with the Bayley Scales of Infant Development II in a low-risk sample. Child Care Health Dev. 2010; 36(4):485-90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20030657.

62. Rydz D, Srour M, Oskoui M, Marget N, Shiller M, Birnbaum R, et al. Screening for developmental delay in the setting of a community pediatric clinic: a prospective assessment of parent-report questionnaires. Pediatrics. 2006; 118(4):e1178-86. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17015506.

63. Eaves LC, Wingert H, Ho HH. Screening for autism: agreement with diagnosis. Autism. 2006; 10(3):229-42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16682396.

64. Wetherby AM, Woods J, Allen L, Cleary J, Dickinson H, Lord C. Early indicators of autism spectrum disorders in the second year of life. J Autism Dev Disord. 2004; 34(5):473-93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15628603.

65. Stone WL, Coonrod EE, Turner LM, Pozdol SL. Psychometric properties of the STAT for early autism screening. J Autism Dev Disord. 2004; 34(6):691-701. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15679188.

66. Hess CR, Papas MA, Black MM. Use of the Bayley Infant Neurodevelopmental Screener with an environmental risk group. J Pediatr Psychol. 2004; 29(5):321-30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15187171.

67. Scambler D, Rogers SJ, Wehner EA. Can the checklist for autism in toddlers differentiate young children with autism from those with developmental delays? J Am Acad Child Adolesc Psychiatry. 2001; 40(12):1457-63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11765292.

68. Baird G, Charman T, Baron-Cohen S, Cox A, Swettenham J, Wheelwright S, et al. A screening instrument for autism at 18 months of age: a 6-year follow-up study. J Am Acad Child Adolesc Psychiatry. 2000; 39(6):694-702. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10846303.

69. Kamio Y, Inada N, Koyama T, Inokuchi E, Tsuchiya K, Kuroda M. Effectiveness of using the Modified Checklist for Autism in Toddlers in two-stage screening of autism spectrum disorder at

the 18-month health check-up in Japan. J Autism Dev Disord. 2014; 44(1):194-203. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23740200.

70. Wiggins LD, Piazza V, Robins DL. Comparison of a broad-based screen versus disorder-specific screen in detecting young children with an autism spectrum disorder. Autism. 2014; 18(2):76-84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23262658.

71. Smith NJ, Sheldrick RC, Perrin EC. An abbreviated screening instrument for autism spectrum disorders. Inf Mental Hlth J. 2013; 34(2):149-55. Available from: http://onlinelibrary.wiley.com/doi/10.1002/imhj.21356/abstract.

72. Matson JL, Kozlowski AM, Fitzgerald ME, Sipes M. True versus false positives and negatives on the Modified Checklist for Autism in Toddlers. Res Autism Spect Dis. 2013; 7(1):17-22. Available from: http://www.sciencedirect.com/science/article/pii/S1750946712000700.

73. Wiggins LD, Bakeman R, Adamson LB, Robins DL. The utility of the social communication questionnaire in screening for autism in children referred for early intervention. Focus Autism Other Dev Disabl. 2007; 22(1):33-8. Available from: http://foa.sagepub.com/content/22/1/33.

74. Taylor CM, Vehorn A, Noble H, Weitlauf AS, Warren ZE. Brief report: Can metrics of reporting bias enhance early autism screening measures? J Autism Dev Disord. 2014; 44(9):2375-80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24682706.

75. Stenberg N, Bresnahan M, Gunnes N, Hirtz D, Hornig M, Lie KK, et al. Identifying children with autism spectrum disorder at 18 months in a general population sample. Paediatr Perinat Epidemiol. 2014; 28(3):255-62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24547686.

76. Steenis LJP, Verhoeven M, Hessen DJ, van Baar AL. Parental and professional assessment of early child development: The ASQ-3 and the Bayley-III-NL. Early Hum Dev. 2015; 91(3):217-25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25703316.

77. Drew A, Baird G, Baron-Cohen S, Cox A, Slonims V, Wheelwright S, et al. A pilot randomised control trial of a parent training intervention for pre-school children with autism. Preliminary findings and methodological challenges. Eur Child Adolesc Psychiatry. 2002; 11(6):266-72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12541005.

78. Cohen H, Amerine-Dickens M, Smith T. Early intensive behavioral treatment: replication of the UCLA model in a community setting. Journal of developmental and behavioral pediatrics : JDBP. 2006; 27(2 Suppl):S145-55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16685181.

79. Magiati I, Charman T, Howlin P. A two-year prospective follow-up study of community-based early intensive behavioural intervention and specialist nursery provision for children with autism spectrum disorders. Journal of child psychology and psychiatry, and allied disciplines. 2007; 48(8):803-12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17683452.

80. Remington B, Hastings RP, Kovshoff H, degli Espinosa F, Jahr E, Brown T, et al. Early intensive behavioral intervention: outcomes for children with autism and their parents after two years. American journal of mental retardation : AJMR. 2007; 112(6):418-38. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17963434.

81. Howard JS, Sparkman CR, Cohen HG, Green G, Stanislaw H. A comparison of intensive behavior analytic and eclectic treatments for young children with autism. Res Dev Disabil. 2005; 26(4):359-83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15766629.

82. Smith T, Groen AD, Wynn JW. Randomized trial of intensive early intervention for children with pervasive developmental disorder. American journal of mental retardation : AJMR. 2000; 105(4):269-85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10934569.

83. Eikeseth S, Smith T, Jahr E, Eldevik S. Intensive behavioral treatment at school for 4- to 7-year-old children with autism. A 1-year comparison controlled study. Behav Modif. 2002; 26(1):49-68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11799654.

84. Eldevik S, Eikeseth S, Jahr E, Smith T. Effects of low-intensity behavioral treatment for children with autism and mental retardation. J Autism Dev Disord. 2006; 36(2):211-24. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16477514.

85. Ben Itzchak E, Lahat E, Burgin R, Zachor AD. Cognitive, behavior and intervention outcome in young children with autism. Res Dev Disabil. 2008; 29(5):447-58. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17923388.

86. Wong VC, Sun JG. Randomized controlled trial of acupuncture versus sham acupuncture in autism spectrum disorder. J Altern Complement Med. 2010; 16(5):545-53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20804366.

87. Wong VC, Chen WX. Randomized controlled trial of electro-acupuncture for autism spectrum disorder. Alternative medicine review : a journal of clinical therapeutic. 2010; 15(2):136-46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20806998.

88. Yan YF, Wei YY, Chen YH, Chen MM. [Effect of acupuncture on rehabilitation training of child's autism]. Zhongguo zhen jiu = Chinese acupuncture & moxibustion. 2007; 27(7):503-5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17722829.

89. Zhou H ZP. The effect of language therapy combined with point massage on communication disability in autism children. 2008; 3(23):24-6. Available from:

90. Allam H, ElDine NG, Helmy G. Scalp acupuncture effect on language development in children with autism: a pilot study. J Altern Complement Med. 2008; 14(2):109-14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18315511.

91. Chan AS, Cheung MC, Sze SL, Leung WW. Seven-star needle stimulation improves language and social interaction of children with autistic spectrum disorders. The American journal of Chinese medicine. 2009; 37(3):495-504. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19606510.

92. Wang CN, Liu Y, Wei XH, Li LX. [Effects of electroacupuncture combined with behavior therapy on intelligence and behavior of children of autism]. Zhongguo zhen jiu = Chinese acupuncture & moxibustion. 2007; 27(9):660-2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17926617.

93. Wong V SJ. Research on tongue acupuncture in children with autism. 2002. Available from:

94. CL W. Acupuncture and autism spectrum disorders - an assessor-blinded randomized controlled trial (M Phil). 2008. Available from:

95. Wallace IF, Berkman ND, Watson LR, Coyne-Beasley T, Wood CT, Cullen K, et al. Screening for Speech and Language Delay in Children 5 Years Old and Younger: A Systematic Review. Pediatrics. 2015. Available from: http://pediatrics.aappublications.org/content/early/2015/07/02/peds.2014-3889.abstract.

96. Dionne C, McKinnon S, Squires J, Clifford J. Developmental screening in a Canadian First Nation (Mohawk): psychometric properties and adaptations of ages & stages questionnaires (2nd edition). BMC Pediatr. 2014; 14:23. Available from: http://www.biomedcentral.com/1471-2431/14/23.

97. Paiva GS, Lima ACVMS, Lima MC, Eickmann SH. The effect of poverty on developmental screening scores among infants. Sao Paulo Med J. 2010; 128(5):276-83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21181068.

98. Simon AE, Pastor PN, Avila RM, Blumberg SJ. Socioeconomic disadvantage and developmental delay among US children aged 18 months to 5 years. J Epidemiol Community Health. 2013; 67(8):689-95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23709661.

99. Durkin MS, Maenner MJ, Meaney FJ, Levy SE, di Guiseppi C, Nicholas JS, et al. Socioeconomic inequality in the prevalence of autism spectrum disorder: Evidence from a U.S. cross-sectional study. PLoS One. 2010; 5(7):e11551. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20634960.

100. Pungello EP, Iruka IU, Dotterer AM, Mills-Koonce R, Reznick JS. The effects of socioeconomic status, race, and parenting on language development in early childhood. Dev Psychol. 2009; 45(2):544-57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19271838.

101. Guerrero AD, Garro N, Chang JT, Kuo AA. An update on assessing development in the pediatric office: has anything changed after two policy statements? Acad Pediatr. 2010; 10(6):400-4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21075321.

102. de Koning HJ, de Ridder-Sluiter JG, van Agt HM, Reep-van den Bergh CM, van der Stege HA, Korfage IJ, et al. A cluster-randomised trial of screening for language disorders in toddlers. J Med Screen. 2004; 11(3):109-16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/?term=15333268.

103. Eikeseth S, Smith T, Jahr E, Eldevik S. Outcome for children with autism who began intensive behavioral treatment between ages 4 and 7: a comparison controlled study. Behav Modif. 2007; 31(3):264-78. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17438342.

104. Rickards AL, Walstab JE, Wright-Rossi RA, Simpson J, Reddihough DS. A randomized, controlled trial of a home-based intervention program for children with autism and developmental delay. Journal of developmental and behavioral pediatrics : JDBP. 2007; 28(4):308-16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17700083.

105. Dawson G, Rogers S, Munson J, Smith M, Winter J, Greenson J, et al. Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. Pediatrics. 2010; 125(1):e17-23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19948568.

106. Tonge B, Brereton A, Kiomall M, Mackinnon A, King N, Rinehart N. Effects on parental mental health of an education and skills training program for parents of young children with autism: a randomized controlled trial. J Am Acad Child Adolesc Psychiatry. 2006; 45(5):561-9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16670650.

107. Tonge B, Brereton A, Kiomall M, Mackinnon A, Rinehart NJ. A randomised group comparison controlled trial of 'preschoolers with autism': a parent education and skills training intervention for young children with autistic disorder. Autism : the international journal of research and practice. 2014; 18(2):166-77. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22987897.

108. Ben-Itzchak E, Zachor DA. The effects of intellectual functioning and autism severity on outcome of early behavioral intervention for children with autism. Res Dev Disabil. 2007; 28(3):287-303. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16730944.

109. Birnbrauer JLD. The Murdoch Early Intervention Program after 2 years. 1993; 10:63-74. Available from:

110. Harris SL, Handleman JS. Age and IQ at intake as predictors of placement for young children with autism: a four- to six-year follow-up. J Autism Dev Disord. 2000; 30(2):137-42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10832778.

111. Lovaas OI. Behavioral treatment and normal educational and intellectual functioning in young autistic children. J Consult Clin Psychol. 1987; 55(1):3-9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3571656.

112. Matos MM, AE. Análisis Comportamental Aplicado (ACA) y Trastornos Generalizados del Desarrollo (TGD): Su evaluación en Argentina [Applied behavior analysis and pervasive developmental disabilities: Assessment in Argentina]. 2005; 22:59-76. Available from:

113. Sallows GO, Graupner TD. Intensive behavioral treatment for children with autism: four-year outcome and predictors. American journal of mental retardation : AJMR. 2005; 110(6):417-38. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16212446.

114. Anderson SADDEE, GL; Christian WP. Intensive home-based early intervention with autistic children. 1987; 10:352-66. Available from:

115. Bibby P, Eikeseth S, Martin NT, Mudford OC, Reeves D. Progress and outcomes for children with autism receiving parent-managed intensive interventions. Res Dev Disabil. 2002; 23(1):81-104. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12071397.

116. Smith T, Eikeseth S, Klevstrand M, Lovaas OI. Intensive behavioral treatment for preschoolers with severe mental retardation and pervasive developmental disorder. American journal of mental retardation : AJMR. 1997; 102(3):238-49. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9394133.

117. Harris SL, Handleman JS, Gordon R, Kristoff B, Fuentes F. Changes in cognitive and language functioning of preschool children with autism. J Autism Dev Disord. 1991; 21(3):281-90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1938774.

118. Sheinkopf SJ, Siegel B. Home-based behavioral treatment of young children with autism. J Autism Dev Disord. 1998; 28(1):15-23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9546298.

Figures

Figure 1. Analytic Framework

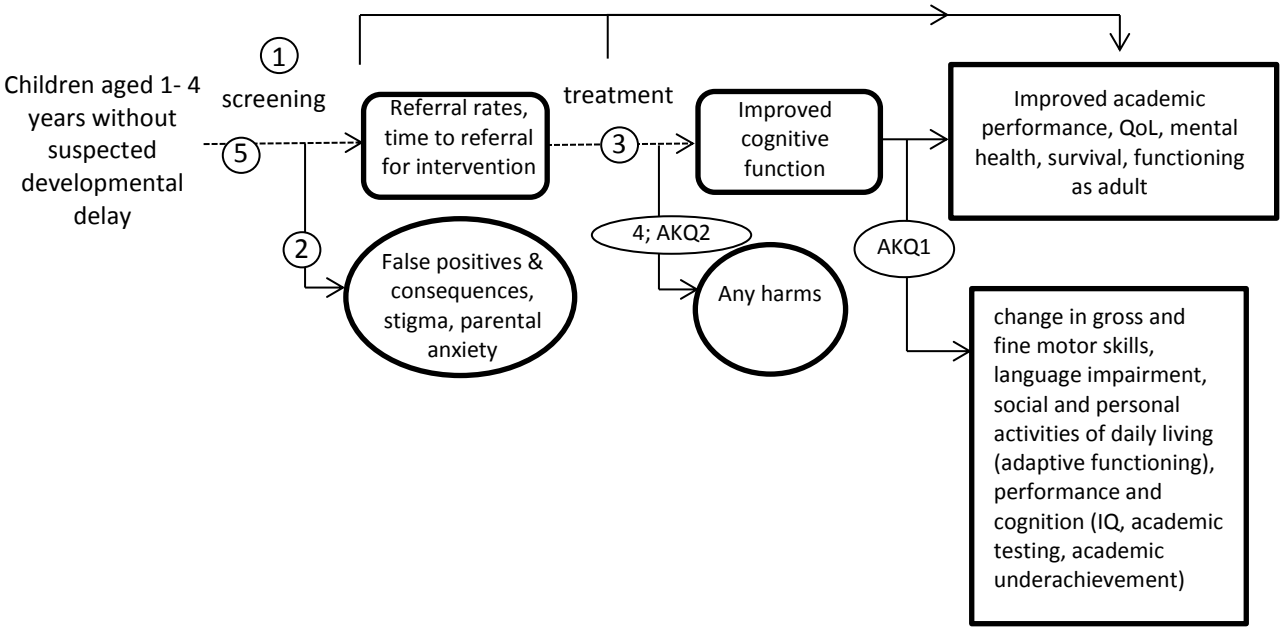


Figure 2a. Search Results for Stage I: Screening for Developmental Delay

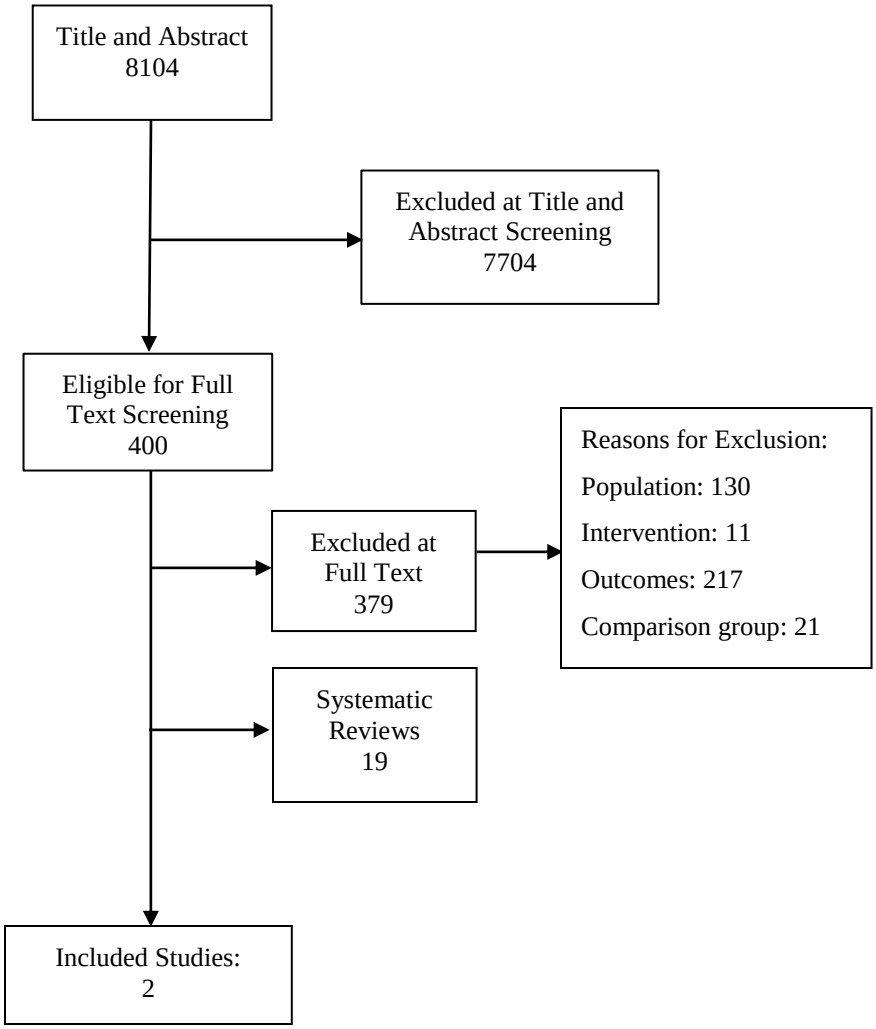


Figure 2b. Search Results for Stage II: Treatment of Developmental Delay Systematic Reviews

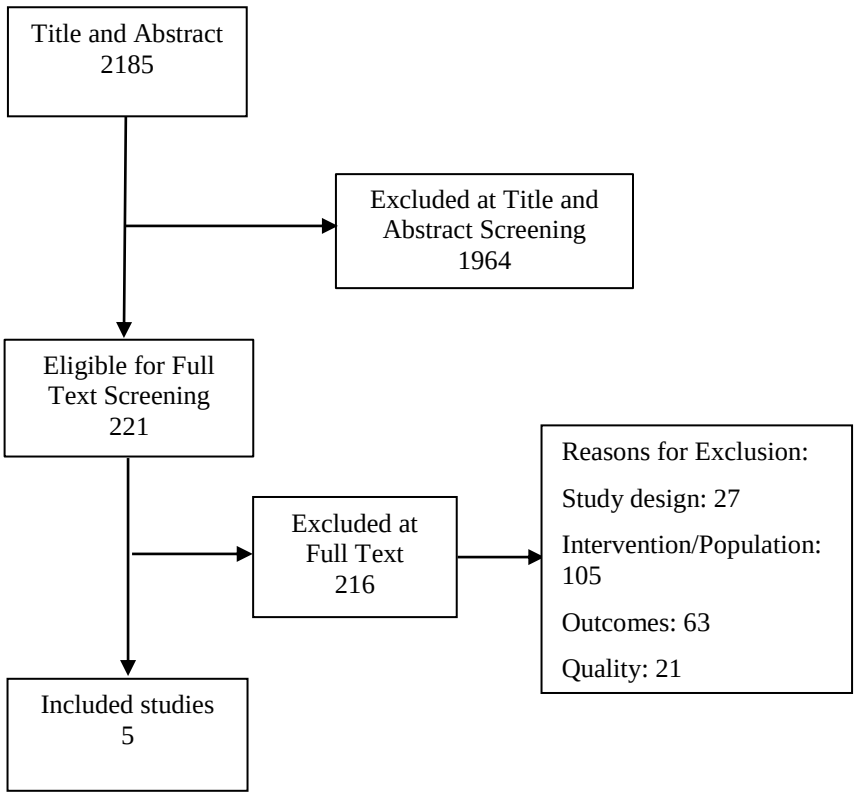


Figure 2c. Search Results for Stage II: Treatment of Developmental Delay RCTs

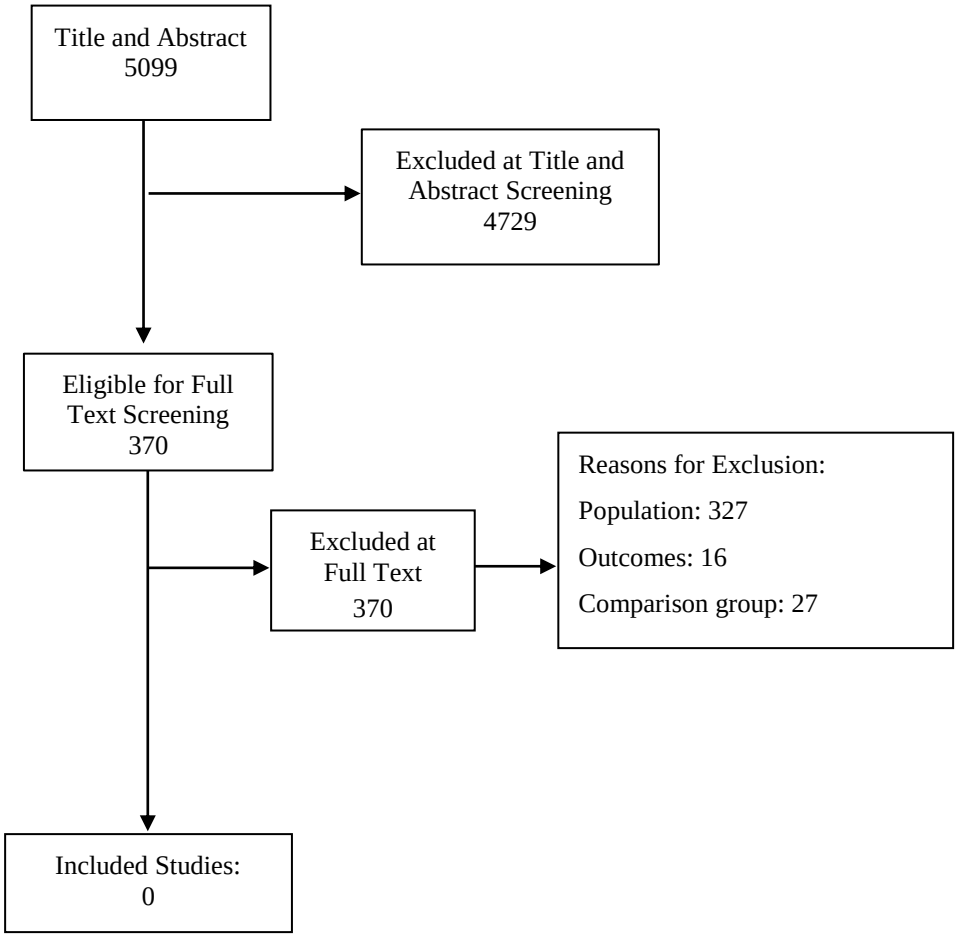


Figure 2d: Search Results for Stage II: Treatment for Developmental Delay (Addendum)

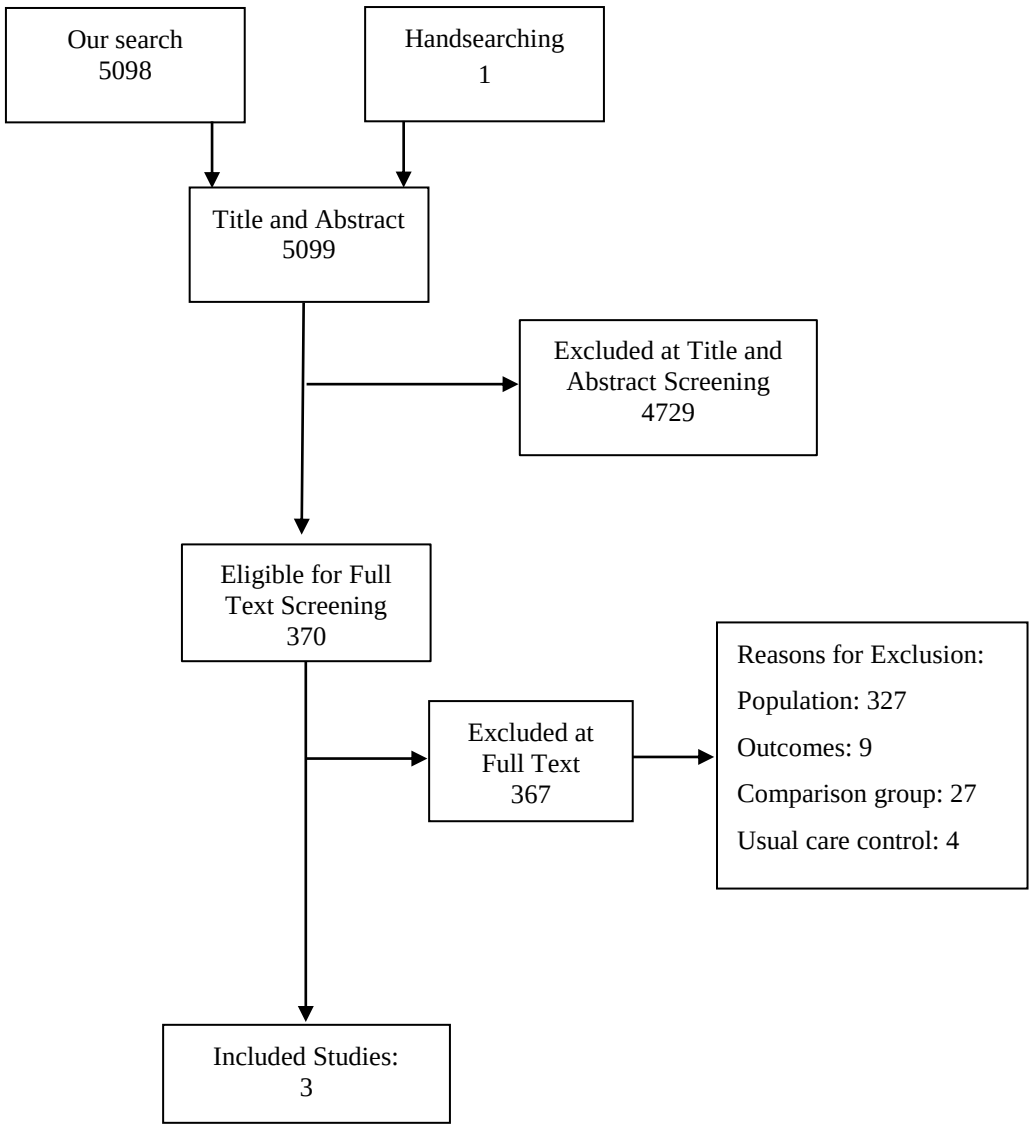
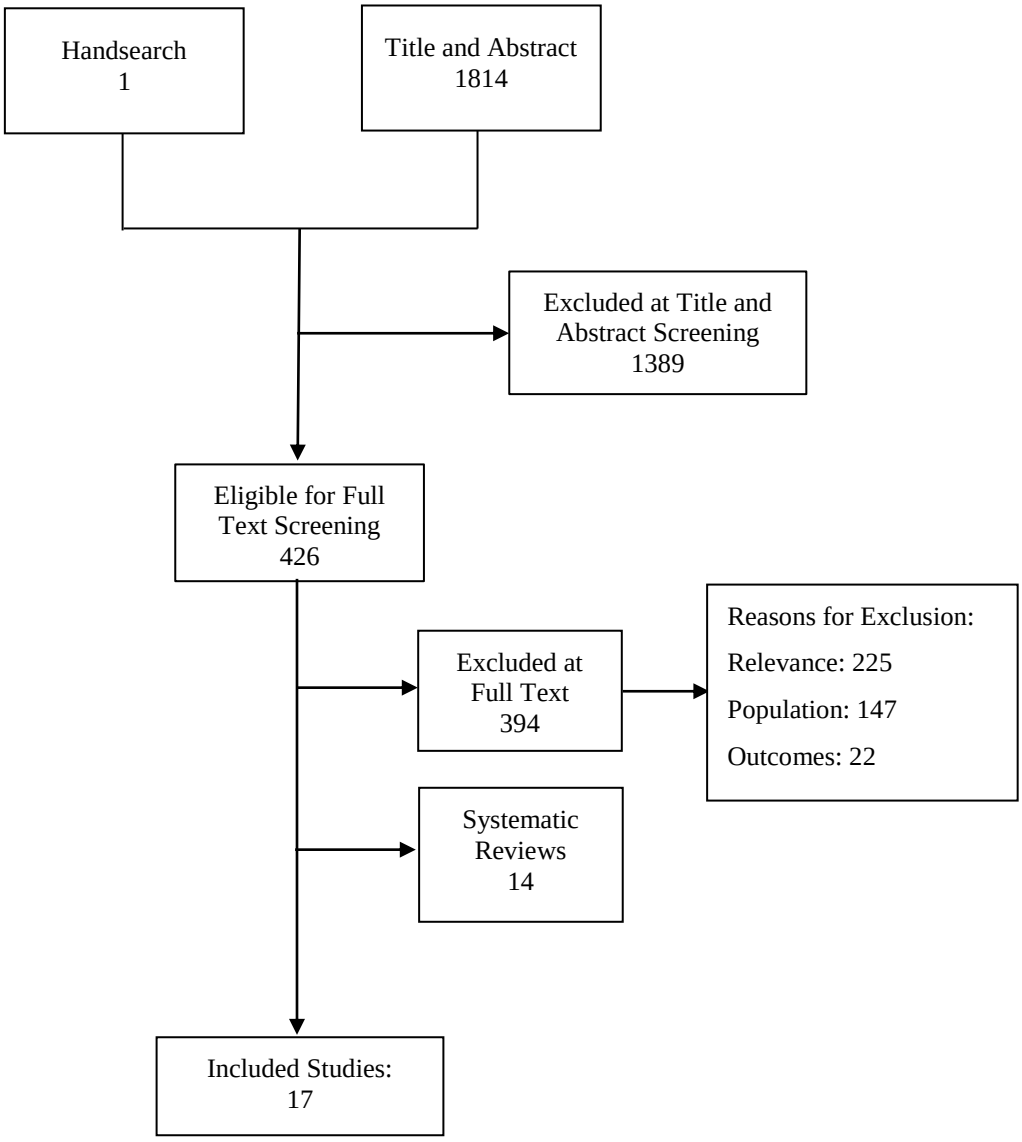


Figure 2e. Search Results for Stage III: Test Properties for Diagnostic Tools



Tables

Table 1a. Stage I (Screening): Characteristics of Included Studies

Study/Location	Guevara [46] U.S.
Objective	To determine the effectiveness of developmental screening on the identification of DDs, early intervention (EI) referrals, and EI eligibility
Methods	Design: RCT Recruitment: all physicians, nurse practitioners, and paediatric residents at four participating urban primary care practices affiliated with Children's Hospital of Philadelphia were eligible Inclusion Criteria: Children <30 months old, >36 weeks estimated gestational age; no major congenital anomalies or genetic syndromes; not living in foster care and not currently receiving EI services
Participants	Sample: 2,103 Intervention 1 n= 707; Intervention 2 n= 698; Control n= 698 Mean age (SD): Intervention 1= 10.5 (8.2) months; Intervention 2= 10.5 (8.1) months; Control= 10.4 (8.6) months Gender [Female n(%)]: Intervention 1= 342 (48.4); Intervention 2= 354 (50.9); Control= 351 (50.4) Race/Ethnicity [Black n (%)]: Intervention 1= 553 (78.2); Intervention 2= 521 (74.9); Control= 549 (78.9) Loss to follow-up: Intervention n= Not reported (NR); Control n= NR
Intervention	Description of Intervention 1: caregivers met with trained office staff to complete ASQ-II at 9, 18 and 30 month well child visit and M-CHAT at 18 and 24 month visit with the use of props Description of Intervention 2: caregivers completed ASQ-II at 9, 18 and 30 month well child visit and M-CHAT at 18 and 24 month without office support Description of Control: usual care developmental screen (milestones consisting of 8-10 questions from 4 domains) Duration of Intervention: 21 months (first screen at 9 months; last screen at 30 months) Length of follow-up: 18 months

Study/Location	van Agt 2007 [47] Netherlands *Companion paper:* de Koning [102]
Objective	To assess the effects of screening and early treatment of preschool children for language delay on language development and school performance at age 8
Methods	Design: cluster RCT Recruitment Setting: child health care physicians identified In 6 regions in the Netherlands, 4 regions in the south, 1 in the mid-south, and in 1 large city in the west, 55 physicians of child health centers were randomly assigned Inclusion/Exclusion criteria: The participating children were those who were between the age of 15 to 24 months in the given inclusion period and were living within the area of the intervention physicians' health care location and those who were living within the area of the control physician
Participants	Sample: 55 clusters Intervention n= 28 clusters; 6,485 children; Control n= 27 clusters, 4,955 children Mean age (SD): NR Gender [Female n(%)]: Overall: 50%; Intervention: 50.1%; Control: 49.9% Race/Ethnicity n (%): NR SES [Paternal education %]: Overall: Low: 18.8%, Intermediate: 45.2%, High: 36%; Intervention: Low: 18.7%, Intermediate: 45.5%, High: 35.8%; Control: Low: 18.8%, Intermediate: 44.8%, High: 36.3% Loss to follow-up: I n= 1,161; C n=860
Intervention	Description of Intervention: children were screened with VTO Language Screening instrument at ages 15/18 months and 24 months) Description of Control: usual monitoring system Duration of Intervention: 6-9 months (two screens) Length of follow-up: 81 months

Table 1b. Stage II (Treatment RCTs Addendum): Characteristics of Included Studies

STUDY/LOCATION	Buschmann 2009,[54] Germany
OBJECTIVE	To evaluate the effectiveness of a short, highly structured parent based language intervention group programme for 2-year-old children with specific expressive language delay (SELD, without deficits in receptive language)
METHODS	Design: RCT Recruitment: children were selected from a sample with language delay identified through developmental check-ups conducted in general paediatric practices Inclusion Criteria: singletons born at term without pre-, peri- or postnatal complications and a German speaking family background Exclusion Criteria: chronic heart deficits, persistent middle ear effusion accompanied by a significant hearing loss of >20 dB, visual impairments, genetic syndromes, pervasive developmental disorders or other diseases with a known influence on language development, deficits in receptive language and/or in non-verbal cognitive abilities, and previous language intervention
PARTICIPANTS	Sample: Eligible n=61; Randomized n=58 Intervention n= 29, Control n= 29 Mean age (SD): Overall: 24.7 (0.9) months Gender [Female n(%)]: Overall: n= 23; Intervention: n= 11; Control: n= 12 SES (Maternal education): No/low graduation (8-9): Intervention 12.5%, Control 8.7%; Middle School education: Intervention 37.5%, Control 56.5%; High School graduation: Intervention: 50.0%; Control: 34.8% Developmental Delay: specific expressive language delay Loss to follow-up: Intervention n=5; Control n=6
INTERVENTION	Description of Intervention: Heidelberg Parent-based Language Intervention (HPLI), a highly structured and interactive programme delivered by a paediatric neurologist; based on an interactive model of language intervention, which presumes that optimised parental input will provide better language learning opportunities for children

	Description of Control: Waiting list Duration of Intervention: 9 months; 8 sessions Length of follow-up: immediate post

STUDY/LOCATION	Glogowska 2000 [55] UK
OBJECTIVE	To compare routine speech and language therapy in preschool children with delayed speech and language against 12 months of watchful waiting
METHODS	Design: RCT Recruitment: children eligible for the trial were identified by 21 speech and language therapists working in 16 NHS community clinics Inclusion Criteria: newly referred singleton children acquiring English in a monolingual home; aged under 3½ years at initial attendance for speech and language therapy assessment; no diagnosis of severe learning difficulties or autism; no motor deficits; no primary diagnosis of dysfluency (stammering) or dysphonia (voice disorders); no siblings currently receiving speech and language therapy; children had to satisfy one of the clinical criteria (box 2); be considered to have significant clinical difficulties by the speech and language therapist; a "carer" had to attend sessions; parents had to give consent Funding: Research and development directorate of the South and West regional office of the NHS Executive
PARTICIPANTS	Sample: Eligible: n=159; Randomized n=159 Intervention n=71, Control n=88 (eligible and randomized) Mean age (SD): Intervention: 34.2 months, Control: 34.2 months Gender [Female n(%)]: Intervention n= 16 (23%); Control n= 23 (26.1%) SES (Maternal education): No qualifications: Intervention 11%, Control 18%; O level or similar (CSE or technical qualification): Intervention 80%, Control 72%; A level and higher: Intervention 9%, Control 9% Developmental Delay: delayed speech and language Loss to Follow-up: Intervention n=0; Control n=5
INTERVENTION	Description of Intervention: one-to-one speech and language therapy typically offered by a speech and language therapist and tailored to the

	child's needs Description of Control: watchful waiting Duration of Intervention: 8.4 months Length of follow-up: immediate post

STUDY/LOCATION	Hund-Reid 2013 [53] Canada
OBJECTIVE	To investigate the effectiveness of phonological awareness intervention in improving the PA skills of kindergarten children with moderate to severe language impairment
METHODS	Design: RCT Recruitment: speech-language pathologists based in classrooms recommended potential participants for the study Inclusion Criteria: receptive or expressive language percentile rank score cut-offs at or below the 6th percentile as measured by the CELF P-2; children were attending a school in the participating school district and were referred; developmental language scores of 1.5 SDs or greater below the mean; hearing within normal limits (Hearing Identification Procedures); nonverbal performance score on the KBIT-2 no lower than 70; PA scores at or below the 25th percentile (measured by the TOPEL), or presence of at-risk indicators in two PA measures from the DIBELS; not yet reading words; English spoken in the home
PARTICIPANTS	Sample: Eligible: n=50; Randomized: n=39 Intervention n= 22, Control n= 17 Mean age (SD) Intervention: 66.68 (4.81) months, Control: 64.13 (4.22) months Gender [Female n(%)]: Overall n= 10, Intervention n= 5, Control n= 5 SES (Maternal education): Intervention grade mean: 11.45 (1.41 SD), Control grade mean: 11.47 (1.25 SD) Developmental Delay: language impairment Loss to Follow-up: Intervention n= 3, Control n= 12
INTERVENTION	Description of Intervention: phonological awareness training program delivered by education assistants; teaching of one of two types of phoneme manipulations, blending and segmenting in each lesson, sound-symbol awareness activities

	Description of Control: no intervention Duration of Intervention: 14 weeks Length of follow-up: immediate post

Table 2a. Stage I (Screening): Summary of Risk of Bias Assessment of Included RCTs Using Cochrane's Risk of Bias Tool

Study	Sequence Generation	Allocation Concealment	Blinding of Participants and Personnel	Blinding of Outcome Assessors	Incomplete Reporting	Selective Reporting	Other Bias*
Guevera 2013[46]	L	L	H	H	L	L	L
van Agt 2007[47]	L	U	U	L	L	L	L

L (green) = Low Risk; U (yellow) = Unclear Risk; H (red) = High Risk

Table 2b. Stage II (Treatment RCTs Addendum): Summary of Risk of Bias Assessment of Included RCTs using Cochrane's Risk of Bias Tool

Study	Sequence Generation	Allocation Concealment	Blinding of Participants and Personnel	Blinding of Outcome Assessors	Incomplete Reporting	Selective Reporting	Other Bias*
Buschmann 2009 [54]	U	L	H	L	L	L	L
Glogowska 2000 [55]	U	L	H	L	L	L	L
Hund-Reid 2013 [53]	U	U	H	L	H	L	H

L (green) = Low Risk; U (yellow) = Unclear Risk; H (red) = High Risk

Table 3. Stage II (Treatment Systematic Reviews): AMSTAR Ratings Summary

Study ID	Q1: A priori design	Q2. Duplicate study selection and data extraction	Q3. Comprehensive literature search	Q4.Search regardless of publication type	Q5. Included and Excluded Studies list	Q6. Characteristics of Included Studies	Q7. Assessment and documentation of scientific quality	Q8. Use of scientific quality in formulation of conclusions	Q9. Appropriate methods used to combine study	Q10. Likelihood of publication bias	Q11. Conflict of interest stated	Overall rating:
Cheuk 2011 [50]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not applicable	Yes	10
Oono 2013 [48]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11
Reichow 2012 [49]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not applicable	Yes	10
Spreckley 2009 [52]	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Can't answer	No	8
Virues-Ortega 2010 [51]	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Can't answer	8

Table 4. Stage II (Treatment Systematic Reviews): Outcomes Summary

Study	Type of Intervention	Outcomes*	Included Studies meeting our Inclusion Criteria	Study Design of Included Studies	Number of participants	Measurement	Overall Effect Size and Calculation Method
Behavioural interventions							
Oono 2013 [48]	Parent-mediated early intervention	Cognitive Function 1(1)[1]	Drew 2002[77]	RCT	N=24	NR	NR. Authors state study did not report any difference between groups in developmental/intellectual gains
Reichow 2012 [49]	EIBI	Cognitive Function 5(5)[2]	Cohen 2006[78] Magiati 2007[79] Remington 2007[80] Howard 2005[81] Smith 2000[82]	CCTs RCT	N=200	Composite IQ	Four studies were pooled (Cohen 2006; Magiati 2007; Remington 2007; Howard 2005) SMD: 0.76 (95% CI 0.040 to 1.11); I^2=21% One study was not pooled (Smith 2000): g=0.74
		Quality of Life 5(5)[3]	Cohen 2006[78] Magiati 2007[79] Remington 2007[80] Howard 2005[81] Smith 2000[82]	CCTs	N=199	Vineland	Four studies were pooled (Cohen 2006; Magiati 2007; Remington 2007; Howard 2005) SMD 0.55 (95% CI 0.24 to 0.87); I^2=0% One study was not pooled (Smith 2000): g=- 0.03
		Harms 5(5)[4]	Cohen 2006[78] Magiati 2007[79] Remington 2007[80] Howard 2005[81] Smith 2000[82]	CCTs RCT	N=200	NA	No deterioration on primary outcomes or adverse events were reported as a result of treatment in any of the studies
Virues-Ortega 2010 [51]	ABA	Cognitive Function 3(3)[5]	Eikseth 2002[83]/2007[103] Eldevik 2006[84] Ben-Itzchak 2008[85]	RCTs CCTs	N=129	IQ	Median effect size: 1.34 (0.60, 2.08)**
Spreckley 2009 [52]	ABI	Cognitive Function 0 (0)[6]	No included studies[6]	RCT	NA	IQ	No studies[6]
Alternative interventions							
Cheuk 2011 [50]	Acupuncture plus conventional treatment	Cognitive Function 4(2)	Wong 2010a[86] Wong 2010b[87] Yan 2007[88] Zhou 2008[89]	RCTs CCTs	N=179	Wong 2010a: GMDS Wong 2010b:	*Acupuncture*: Wong 2010a: GMDS – general quotient MD 3.46 (95% CI -2.0 to 8.92)

Study	Type of Intervention	Outcomes*	Included Studies meeting our Inclusion Criteria	Study Design of Included Studies	Number of participants	Measurement	Overall Effect Size and Calculation Method
	Acupressure plus conventional treatment					Leiter-R Yan 2007: CPEP Zhou 2008: basic developmental assessment	Wong 2010b: No significant difference between the intervention and control groups in any of the domain scores Leiter-R – overall not provided Yan 2007: intervention group showed greater improvement in total score (MD 10.75, 95% CI 3.82 to 17.68, P = 0.002 *Acupressure*: Zhou 2008: no significant differences in developmental aspects
		Quality of Life 0(0)[7]	No trial reported on this outcome	RCTs CCTs	NA	NA	No trial reported this outcome
		Harms 9(4)[8]	Allam 2008[90] Chan 2009[91] Wang 2007[92] Wong 2002[93] Wong 2008[94] Wong 2010a[86] Wong 2010b[87] Yan 2007[88] Zhou 2008[89]	RCTs CCTs	N=357	Measurement not reported	*Acupuncture:* Two studies (Wong 2010a and Wong 2002) reported initial crying for fear and pain; one study (Wong 2010b) reported superficial bleeding or crying and irritability; one study (Wong 2008) reported worsening of sleeping patterns. Three studies (Wang 2007, Yan 2007, Allam 2008) did not report any adverse events *Acupressure:* Two studies (Chan 2009 and Zhou 2008) did not report adverse effects

*Outcomes are reported with the number of included studies on children (the number of included studies on children with a mean age between 1-6 years); ABA – Applied Behavioural Analysis; ABI – Applied Behavioural Intervention; BSID - Bayley Infant Development Measurement Scale; CPEP - Chinese version of Psychoeducational Profile; EIBI – Early Intensive Behavioural Intervention; NA - not available; g= Hedge's g; GMDS - Griffiths Mental Developmental Scale; IQ – Intelligence Quotient; Leiter-R - Leiter International Performance Scale – Revised; SMD – standardized mean difference; WISC – Weschler Intelligence Scale for Children

**Meta-analysis in this systematic review included studies that did not meet our inclusion criteria. Median effect size was calculated using the effect sizes provided for the three studies meeting our criteria.

[1]Five studies reported on developmental/intellectual gains. The following studies were not included in our analysis based on the descriptions provided by the systematic review: Smith 2000[82] (control group received less intensive version of the intervention); Rickards 2007[104] (control group received alternative intervention); Dawson 2010[105] (control group received alternative intervention); Tonge 2006[106]/2012[107] (as per characteristics of included the PEBM group was treated as the intervention with the PEAC group considered the control for the purpose of this review)

[2]Five studies measured intelligence (IQ). All studies met our inclusion criteria (control groups defined as treatment as usual) and were included in our analysis.

[3]Five studies measured quality of life. All studies met our inclusion criteria (control groups defined as treatment as usual) and were included in our analysis.
[4]Five studies reported no adverse events. All studies met our inclusion criteria (control groups defined as treatment as usual) and were included in our analysis.
[5]Seventeen studies provided composite IQ scores; we have included three studies. Four studies were excluded because they had already been considered as part of another systematic review: Cohen 2006[78], Howard 2005[81], Remington 2007[80], Smith 2000[82]. Eight studies were excluded because they did not include a control group: Ben-Itzchak 2007[108]; Birnbrauer 1993[109]; Harris 2000[110]; Lovaas 1987[111], Matos 2005[112]; Sallows 2005[113]; Anderson 1987[114]; Bibby 2001[115]. Two studies were excluded because the control group did not meet our inclusion criteria: Smith 1997[116], control group received low intensity <10 weekly hours of intervention and Harris 1991[117], control group included typically developing children. Sheinkoff 1998[118] and Magiati 2007[79]
[6]Although this systematic review met all our inclusion criteria, no studies were included. Three studies reported on cognitive outcomes. Two of these studies were already considered with other systematic reviews: Smith 2000[82] and Eikseth 2007[103]. One study (Sallows 2005[113]) had a control group that received a parent-directed ABI.
[7]Wong 2007[92] was not included as mean age including SD was not between 1 and 6 years.
[8]Wong 2007[92] was not included as mean age including SD was not between 1 and 6 years.

Table 5. Stage III (Test Properties): QUADAS-II Ratings Summary

Study	DD test assessed	Frequency of screening	Reference Standard	Domain 1: ROB Patient Selection	Domain 1: Applicability Patient Selection	Domain 2: ROB Index Test	Domain 2: Applicability Index Test	Domain 3: ROB Ref. Standard	Domain 3: Applicability Ref. Standard	Domain 4: ROB Flow and Timing
Corsello 2013[57]	SCQ and M-CHAT	One each	Case reviewer diagnosis	H	H	L	L	L	L	L
Eaves 2006[63]	M-CHAT	Twice	Diagnostic evaluation	H	L	L	L	U	L	U
Gollenberg 2009[61]	ASQ	Single	BSID II	U	L	U	L	U	L	L
Kamio 2014[69]	M-CHAT-JV	Twice	Diagnostic evaluation	L	L	L	L	H	L	H
Limbos 2011[59]	ASQ and PEDS	One each	Bayley Scales of Infant Development; Wechsler Preschool and Primary Scale of Intelligence; Vineland Adaptive Behavior Scales; Preschool Language Scale	H	L	L	L	L	L	L
Maljaars 2012[58]	SCQ	Single	Clinical diagnosis	H	L	U	L	L	L	H
Matson 2013[72]	M-CHAT	Single	Two clinical psychologist diagnoses	L	L	U	L	U	L	L
Robins 2014[56]	M-CHAT	Twice	Diagnostic evaluation	U	L	H	L	U	L	H
Rydz 2006[62]	ASQ	Single	BDI	L	L	L	L	L	L	H
Scambler 2001[67]	CHAT	Single	Various tools - U what combination informed the diagnosis	H	L	U	L	U	U	U
Smith 2013[71]	M-CHAT	Single	Diagnosis	H	L	U	L	U	L	L
Steenis 2015[76]	ASQ	Single	Bayley-III	U	U	L	L	L	L	H

Study	DD test assessed	Frequency of screening	Reference Standard	Domain 1: ROB Patient Selection	Domain 1: Applicability Patient Selection	Domain 2: ROB Index Test	Domain 2: Applicability Index Test	Domain 3: ROB Ref. Standard	Domain 3: Applicability Ref. Standard	Domain 4: ROB Flow and Timing
Stenberg 2014[75]	M-CHAT	Single	Clinical diagnosis	U	L	U	L	U	L	L
Stone 2004[65]	STAT	Single	Diagnostic evaluation	H	H	L	L	L	L	H
Taylor 2014[74]	M-CHAT	Single	Clinical diagnosis	U	L	U	L	U	L	L
Wiggins 2014[70]	M-CHAT and PEDS	One each	Diagnosis	U	L	L	L	L	L	L
Wiggins 2007[73]	SCQ	Single	Diagnosis using ADOS, clinical interview, and clinical judgment	H	H	U	L	H	L	U

L (green) = Low Risk; U (yellow) = Unclear Risk; H (red) = High Risk

Evidence Set 1 Stage 1 KQ1: Screening

- Summary of Screening Outcomes Evidence
- GRADE Evidence Profile Table 1.1: Effect of Screening for Developmental Delay
- GRADE Summary of Findings Table 1.2: Effect of Screening for Developmental Delay
- Forest Plots 1.1 to 1.3
  - 1.1: Effectiveness of Developmental Delay Screening - Referral Rates to Intervention (Screening Tool - ASQ-II)
  - 1.2: Effectiveness of Developmental delay screening – Time to intervention referral (Screening tool - ASQ-II)
  - 1.3: Effectiveness of Developmental delay screening: Academic performance - By outcome measures (VTO screening)

Summary of Screening Outcomes Evidence

1.1.a Referral Rates for Screening (with Office Support)
- 1 study; 1,399 participants
- Statistically significant ($p<0.00001$) higher referral rates for the intervention group as compared to the control group (RR 1.95 95% CI 1.49 to 2.54)

1.1.b Referral Rates for Screening (without Office Support)
- 1 study; 1,388 participants
- Statistically significant ($p=0.0001$) higher referral rates for the intervention group as compared to the control group (RR 1.71 95% CI 1.30 to 2.25)

1.2.a Time to Intervention Referral (with Office Support)
- 1 study; 1,399 participants
- Statistically significant ($p<0.00001$) shorter time to intervention referral for the intervention group as compared to the control group (RR 0.30 95% CI 0.19 to 0.48)

1.2.b Time to Intervention Referral (without Office Support)
- 1 study; 1,388 participants
- Statistically significant ($p<0.0001$) shorter time to intervention referral for the intervention group as compared to the control group (RR 0.36 95% CI 0.23 to 0.59)

1.3.a Academic Performance – Special School Attendance
- 1 study; 5,406 participants
- No significant differences ($p=0.08$) in special school attendance between the intervention and control groups (RR 0.71 95% CI 0.48 to 1.04)

1.3.b Academic Performance – Repeating a Grade
- 1 study; 5,334 participants
- No significant differences ($p=0.92$) in repeating a grade between the intervention and control groups (RR 0.99 95% CI 0.81 to 1.21)

1.3.c Academic Performance – Repeating a Grade (Language Problem)
- 1 study; 4,122 participants

- No significant differences (p=0.20) in repeating grade (for a language problem) between the intervention and control groups (RR 1.26 95% CI 0.89 to 1.80)

1.3.d Academic Performance – Below 10th percentile of oral test

- 1 study; 2,195 participants
- No significant differences (p=0.45) between the intervention and control groups in number of participants below the 10th percentile of the oral test (RR 0.88 95% CI 0.63 to 1.23)

1.3.e Academic Performance – Below 10th percentile of reading test

- 1 study; 3,172 participants
- No significant differences (p=1.00) between the intervention and control groups in number of participants below the 10th percentile of the reading test (RR 1.00 95% CI 0.72 to 1.40)

1.3.f Academic Performance – Below 10th percentile of spelling test

- 1 study; 2,953 participants
- No significant differences (p=0.14) between the intervention and control groups in number of participants below the 10th percentile of the spelling test (RR 0.67 95% CI 0.41 to 1.13)

GRADE Evidence Profile Table 1.1: Effect of Screening for Developmental Delay (ages 1 to 4 years old)

Quality Assessment							No. of Participants		Effect				Quality	Importance
No. of Studies	Design	Risk of Bias	Inconsistency	Indirectness	Imprecision	Other	Treatment	Control	Relative (95% CI)	Absolute per Million (Range)	ARR/ARI	NNS (95% CI)		
Referral rates to intervention (Screening tool - ASQ-II) - Screening with Office support (follow-up 18 months)														
1	randomized trial[1]	no serious risk[2]	no serious inconsistency[3]	serious[4]	no serious imprecision[5]	none[6]	140/704 (19.8863%)	71/695 (10.2158%)	RR 1.9466 (1.4925 to 2.5389)	96,703 more (from 50,313 more to 157,211 more)	9.67%	10 (6,20)	⊕⊕⊕O MODERATE	CRITICAL
Referral rates to intervention (Screening tool - ASQ-II) - Screening without Office support (follow-up 18 months)														
1	randomized trial[1]	no serious risk[2]	no serious inconsistency[3]	serious[4]	no serious imprecision[7]	none[6]	121/693 (17.4603%)	71/695 (10.2158%)	RR 1.7091 (1.3002 to 2.2467)	72,440 more (from 30,668 more to 127,361 more)	7.24%	14 (8,33)	⊕⊕⊕O MODERATE	CRITICAL
Time to referral (Screening tool - ASQ-II) - Screening with Office support (follow-up 18 months)														
1	randomized trial[1]	no serious risk[2]	no serious inconsistency[3]	serious[4]	no serious imprecision[8]	none[6]	-/704	-/695	RR 0.3000 (0.1871 to 0.4811)	-	-	-	⊕⊕⊕O MODERATE	CRITICAL
Time to referral (Screening tool - ASQ-II) - Screening without Office support (follow-up 18 months)														
1	randomized trial[1]	no serious risk[2]	no serious inconsistency[3]	serious[4]	no serious imprecision[9]	none[6]	-/693	-/695	RR 0.3649 (0.2276 to 0.5853)	-	-	-	⊕⊕⊕O MODERATE	CRITICAL
Academic performance - By outcome measures (VTO screening) - Special School attendance (follow-up 81 months)														
1	randomized trial[10]	serious risk[11]	no serious inconsistency[3]	no serious indirectness[12]	serious[13]	none[6]	83/3,118 (2.6620%)	85/2,288 (3.7150%)	RR 0.7103 (0.4847 to 1.0410)	10,762 fewer (from 19,144 fewer to 1,523 more)	-	-	⊕⊕OO LOW	CRITICAL
Academic performance - By outcome measures (VTO screening) - Repeating a grade (follow-up 81 months)														
1	randomized trial[10]	serious risk[11]	no serious inconsistency[3]	no serious indirectness[12]	serious[14]	none[6]	443/3,084 (14.3645%)	318/2,250 (14.1333%)	RR 0.9900 (0.8107 to 1.2091)	1,413 fewer (from 26,754 fewer to 29,553 more)	-	-	⊕⊕OO LOW	CRITICAL
Academic performance - By outcome measures (VTO screening) - Repeating a grade (language problems) (follow-up 81 months)														
1	randomized trial[10]	serious risk[11]	no serious inconsistency[3]	no serious indirectness[12]	serious[15]	none[6]	146/2,401 (6.0808%)	84/1,721 (4.8809%)	RR 1.2624 (0.8871 to 1.7964)	12,807 more (from 5,511 fewer to 38,871 more)	-	-	⊕⊕OO LOW	CRITICAL
Academic performance - By outcome measures (VTO screening) - Below 10 percentile of oral test (follow-up 81 months)														

Quality Assessment							No. of Participants		Effect				Quality	Importance
No. of Studies	Design	Risk of Bias	Inconsistency	Indirectness	Imprecision	Other	Treatment	Control	Relative (95% CI)	Absolute per Million (Range)	ARR/ARI	NNS (95% CI)		
1	randomized trial[10]	serious risk[11]	no serious inconsistency[3]	no serious indirectness[12]	serious[16]	none[6]	112/1,270 (8.8189%)	90/925 (9.7297%)	RR 0.8799 (0.6293 to 1.2302)	11,685 fewer (from 36,068 fewer to 22,398 more)	-	-	⊕⊕OO LOW	CRITICAL
Academic performance - By outcome measures (VTO screening) - Below 10 percentile of reading test (follow-up 81 months)														
1	randomized trial[10]	serious risk[11]	no serious inconsistency[3]	no serious indirectness[12]	serious[17]	none[6]	86/1,844 (4.6638%)	62/1,328 (4.6687%)	RR 1.0000 (0.7166 to 1.3954)	0 fewer (from 13,231 fewer to 18,460 more)	-	-	⊕⊕OO LOW	CRITICAL
Academic performance - By outcome measures (VTO screening) - Below 10 percentile of spelling test (follow-up 81 months)														
1	randomized trial[10]	serious risk[11]	no serious inconsistency[3]	no serious indirectness[12]	serious[18]	none[6]	48/1,728 (2.7778%)	52/1,225 (4.2449%)	RR 0.6798 (0.4092 to 1.1293)	13,592 fewer (from 25,079 fewer to 5,489 more)	-	-	⊕⊕OO LOW	CRITICAL
Cognitive Function														
0[19]	N/A[20]	N/A[21]	N/A[22]	N/A[23]	N/A[24]	N/A[25]	-	-	0[26]	-	-	-	N/A[27]	CRITICAL
Incidence of Mental Health														
0[19]	N/A[20]	N/A[21]	N/A[22]	N/A[23]	N/A[24]	N/A[25]	-	-	0[26]	-	-	-	N/A[27]	CRITICAL
Overall Quality of Life														
0[19]	N/A[20]	N/A[21]	N/A[22]	N/A[23]	N/A[24]	N/A[25]	-	-	0[26]	-	-	-	N/A[27]	CRITICAL
Incidence of Mental Health														
0[19]	N/A[20]	N/A[21]	N/A[22]	N/A[23]	N/A[24]	N/A[25]	-	-	0[26]	-	-	-	N/A[27]	CRITICAL

- Footnotes appear after the Summary of Findings Table

GRADE Summary of Findings Table 1.1: Effect of Screening for Developmental Delay (ages 1 to 4 years old)

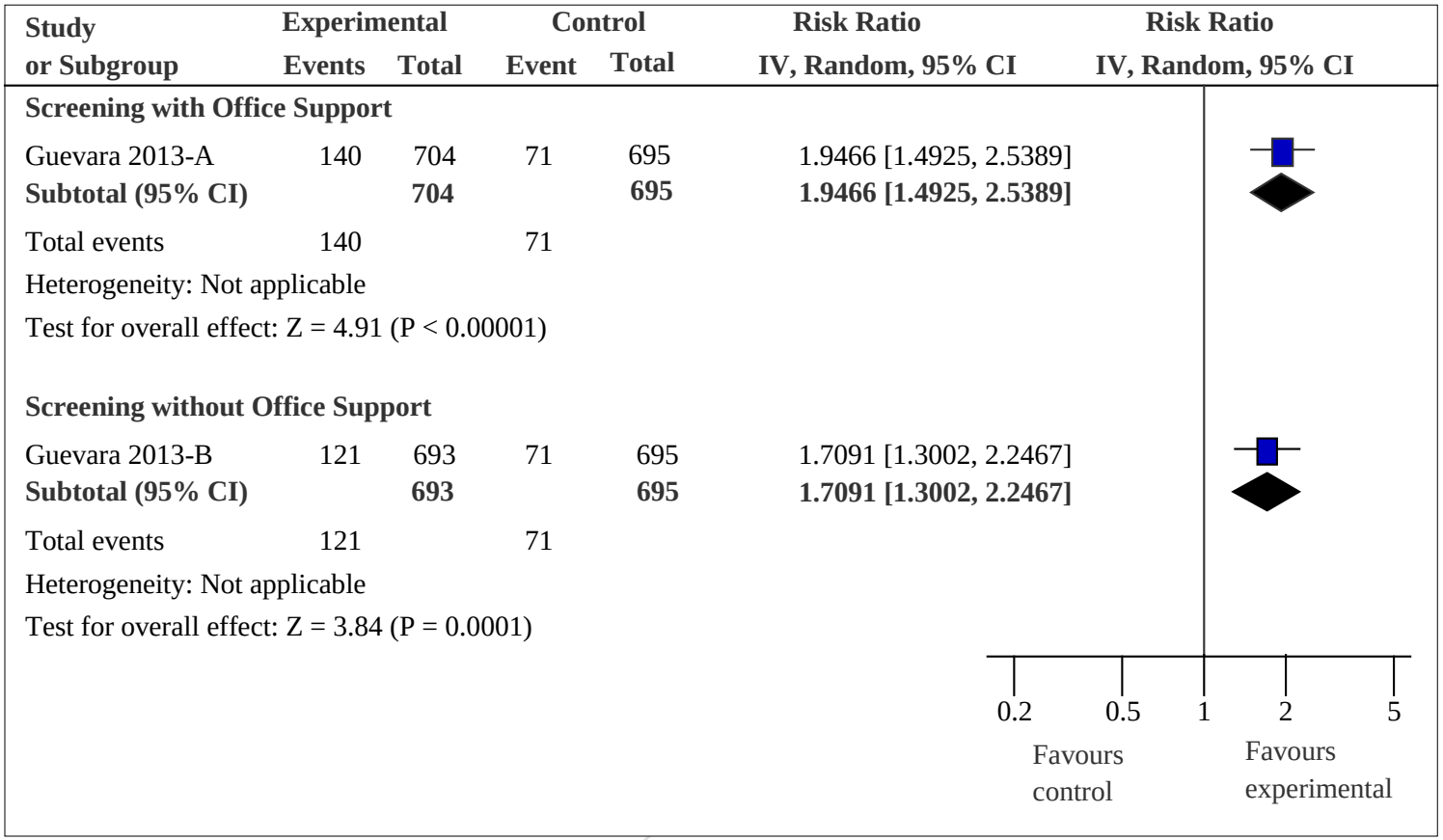
Outcomes	Illustrative Comparative Risks* (95% CI)		Relative Effect (95% CI)	No. of Participants (Studies)	Quality of the Evidence (GRADE)
	Assumed Risk Number per Million Control	Corresponding Risk Number per Million Treatment			
Referral rates to intervention (Screening tool - ASQ-II) - Screening with Office support Follow-up: 18 months	**102,158**	**198,861** (152,471 to 259,370)	**RR 1.9466** (1.4925 to 2.5389)	1,399 (1 study[1])	⊕⊕⊕⊖ **moderate**[2,3,4,5,6]
Referral rates to intervention (Screening tool - ASQ-II) - Screening without Office support Follow-up: 18 months	**102,158**	**174,599** (132,826 to 229,519)	**RR 1.7091** (1.3002 to 2.2467)	1,388 (1 study[1])	⊕⊕⊕⊖ **moderate**[2,3,4,6,7]
Time to intervention referral (Screening tool - ASQ-II) - Screening with Office support Follow-up: 18 months	**Not applicable**		**RR 0.3000** (0.1871 to 0.4811)	1,399 (1 study[1])	⊕⊕⊕⊖ **moderate**[2,3,4,6,8]
Time to intervention referral (Screening tool - ASQ-II) - Screening without Office support Follow-up: 18 months	**Not applicable**		**RR 0.3649** (0.2276 to 0.5853)	1,388 (1 study[1])	⊕⊕⊕⊖ **moderate**[2,3,4,6,9]
Academic performance - By outcome measures (VTO screening) - Special School attendance Follow-up: 81 months	**371,50**	**263,88** (18,007 to 38,674)	**RR 0.7103** (0.4847 to 1.0410)	5,406 (1 study[10])	⊕⊕⊖⊖ **low**[3,6,11,12,13]
Academic performance - By outcome measures (VTO screening) - Repeating a grade Follow-up: 81 months	**141,333**	**139,920** (114,579 to 170,886)	**RR 0.9900** (0.8107 to 1.2091)	5,334 (1 study[10])	⊕⊕⊖⊖ **low**[3,6,11,12,14]
Academic performance - By outcome measures (VTO screening) - Repeating a grade (language problems) Follow-up: 81 months	**48,809**	**61,616** (43,298 to 87,680)	**RR 1.2624** (0.8871 to 1.7964)	4,122 (1 study[10])	⊕⊕⊖⊖ **low**[3,6,11,12,15]
Academic performance - By outcome measures (VTO screening) - Below 10 percentile of oral test Follow-up: 81 months	**97,297**	**85,612** (61,229 to 119,695)	**RR 0.8799** (0.6293 to 1.2302)	2,195 (1 study[10])	⊕⊕⊖⊖ **low**[3,6,11,12,16]
Academic performance - By outcome measures (VTO screening) - Below 10 percentile of reading test Follow-up: 81 months	**46,687**	**46,687** (33,456 to 65,147)	**RR 1.0000** (0.7166 to 1.3954)	3,172 (1 study[10])	⊕⊕⊖⊖ **low**[3,6,11,12,17]
Academic performance - By outcome measures (VTO screening) - Below 10 percentile of spelling test	**42,449**	**28,857** (17,370 to 47,938)	**RR 0.6798** (0.4092 to 1.1293)	2,953 (1 study[10])	⊕⊕⊖⊖ **low**[3,6,11,12,18]

Outcomes	Illustrative Comparative Risks* (95% CI)	Relative Effect (95% CI)	No. of Participants	Quality of the Evidence
Follow-up: 81 months				

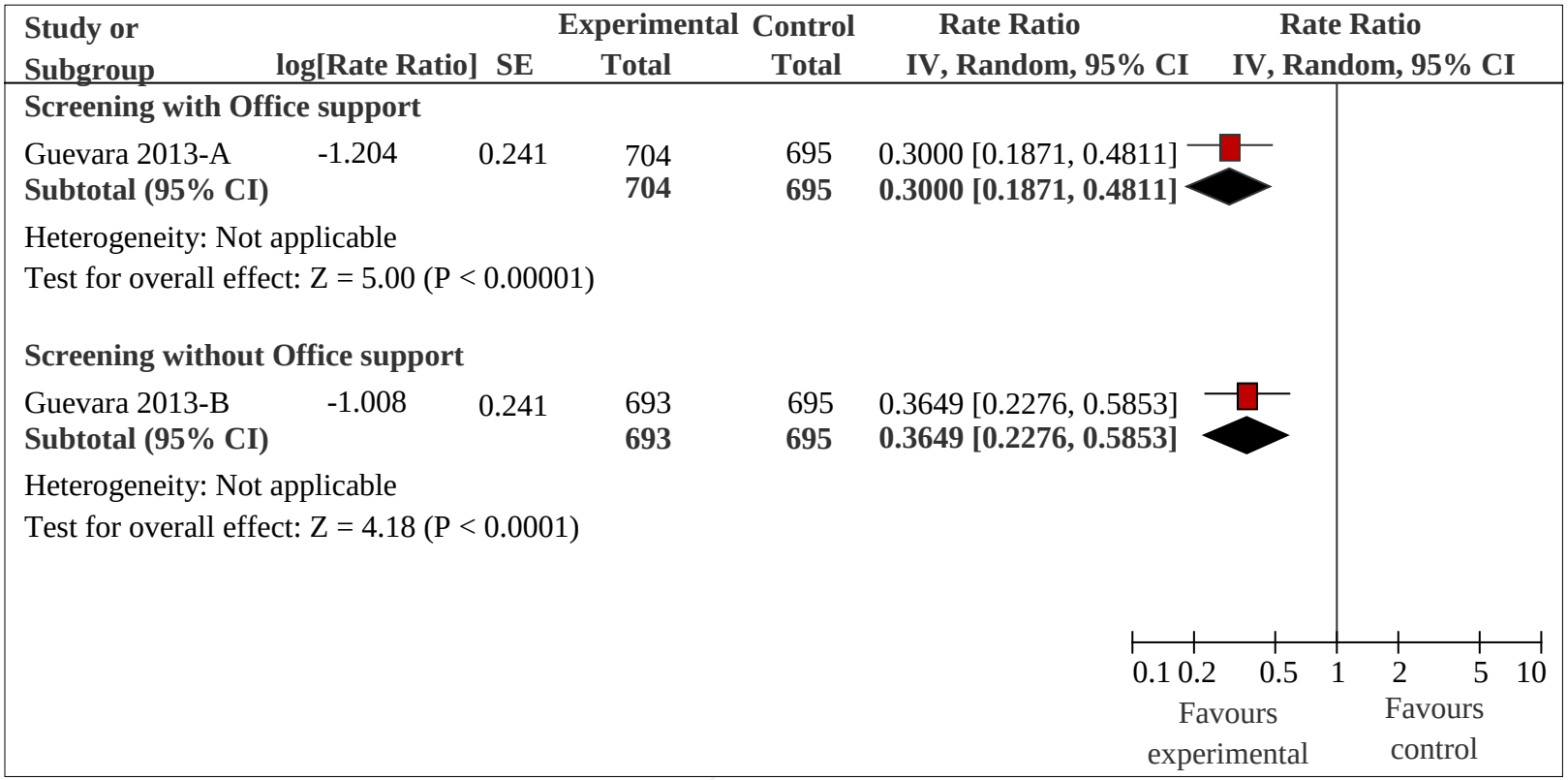
[1] The single study is Guevera et al. 2013 [46]
[2] Using Cochrane's Risk of Bias tool, for this outcome the study was rated as having a low risk of bias. There was low risk of bias for all domains except blinding, which was assessed as being high risk because parents and clinicians were aware of their screening status. As the control participants received usual care (developmental milestone screening) in this study, lack of blinding was not considered as having a large impact on outcomes of interest. Given that all of the information for this outcome is from a study with low risk of bias, this body of evidence was not downgraded for serious study limitations.
[3] A single study therefore cannot assess for inconsistency.
[4] This study included mixed gender children <12 months [mean age Intervention group A: 10.5 (8.2) months; Intervention group B: 10.5 (8.1) months; Control group: 10.4 (8.6) months] with and average risk for developmental delay. The intervention groups were screened using ASQ-II [one group with office support (A), one group without (B)] and the control group received usual care. The study took place in a primary care setting in the US and was published 2013. This body of evidence was downgraded because the population was not restricted to children aged 1-4 years.
[5] The number of events (Intervention A n= 140; Control n=71) and sample size (Intervention A n=704; Control n=695) are adequate. The pooled effect estimate is precise with a narrow confidence interval [RR 1.9466 (95% CI 1.4925 to 2.5389)]. This body of evidence was not downgraded for imprecision.
[6] There were too few studies (n<10) to assess publication bias.
[7] The number of events (Intervention B n= 121; Control n=71) and sample size (Intervention B n=693; Control n=695) are adequate. The pooled effect estimate is precise with a narrow confidence interval [RR 1.7091 (95% CI 1.3002 to 2.2467)]. This body of evidence was not downgraded for imprecision.
[8] The sample size is adequate (Intervention A n=704; Control n=695). The pooled effect estimate is precise with a narrow confidence interval [RR 0.3000 (95% CI 0.1871 to 0.4811)]. This body of evidence was not downgraded for imprecision.
[9] The sample size is adequate (Intervention B n=693; Control n=695). The pooled effect estimate is precise with a narrow confidence interval [RR 0.3649 (95% CI 0.2276 to 0.5853)]. This body of evidence was not downgraded for imprecision.
[10] This single study is van Agt et al. 2007.[47]
[11] Using Cochrane's Risk of Bias tool, for this outcome the study was rated as having unclear risk of bias. There was low risk of bias for all domains except allocation concealment and blinding of participants/personnel, which were assessed as having unclear risk because there was insufficient information to evaluate these domains. Given that all of the information for this outcome is from a study with unclear risk of bias, this body of evidence was downgraded for serious study limitations.
[12] This study included mixed gender children aged 15 months at study entry (mean age not reported) with an average risk for developmental delay. The intervention group was screened using VTO and the control group received usual care. The study took place in a primary care setting in the Netherlands and was published in 2007. There were no serious concerns regarding directness of this evidence.
[13] The sample size is adequate (3,118 intervention arm, 2,288 control arm) but the number of events is fairly low (83 intervention arm, 85 control arm) and the pooled effect estimate is not precise with a confidence interval that includes the no effect value [RR 0.7103 (95% CI 0.4847 to 1.0410)]. This body of evidence was downgraded for imprecision.
[14] The sample size is adequate (3,084 intervention arm, 2,250 control arm) and the number of events is sufficient (443 intervention arm, 318 control arm) but the pooled effect estimate is not precise with a confidence interval that includes the no effect value [RR 0.9900 (95% CI 0.8107 to 1.2091)]. This body of evidence was downgraded for imprecision.
[15] The sample size is adequate (2,401 intervention arm, 1,721 control arm) and the number of events is sufficient (146 intervention arm, 84 control arm) but the pooled effect estimate is not precise with a confidence interval that includes the no effect value [RR 1.2624 (95% CI 0.8871 to 1.7964)]. This body of evidence was downgraded for imprecision.
[16] The sample size is adequate (1,270 intervention arm, 925 control arm) and the number of events is sufficient (112 intervention arm, 90 control arm) but the pooled effect estimate is not precise with a confidence interval that includes the no effect value [RR 0.8799 (95% CI 0.6293 to 1.2302)]. This body of evidence was downgraded for imprecision.
[17] The sample size is adequate (1,844 intervention arm, 1,328 control arm) but the number of events is fairly low (86 intervention arm, 62 control arm) and the pooled effect estimate is not precise with a confidence interval that includes the no effect value [RR 1.0000 (95% CI 0.7166 to 1.3954)]. This body of evidence was downgraded for imprecision.
[18] The sample size is adequate (1,728 intervention arm, 1,225 control arm) but the number of events is low (48 intervention arm, 52 control arm) and the pooled effect estimate is not precise with a confidence interval that includes the no effect value [RR 0.6798 (95% CI 0.4092 to 1.1293)]. This body of evidence was downgraded for imprecision.
[19]No studies were found that met the inclusion criteria of this review for this intervention or outcomes
[20]No studies were found that met the inclusion criteria of this review for this intervention or outcomes
[21]Risk of bias cannot be assessed
22Inconsistency cannot not be assessed
[23]Indirectness cannot be assessed
[24]Imprecision cannot be assessed
[25]Other considerations cannot be assessed
[26]There are not studies to provide data on the effect of this treatment for these outcomes

[27]Since there no studies the overall quality of the evidence cannot be determined

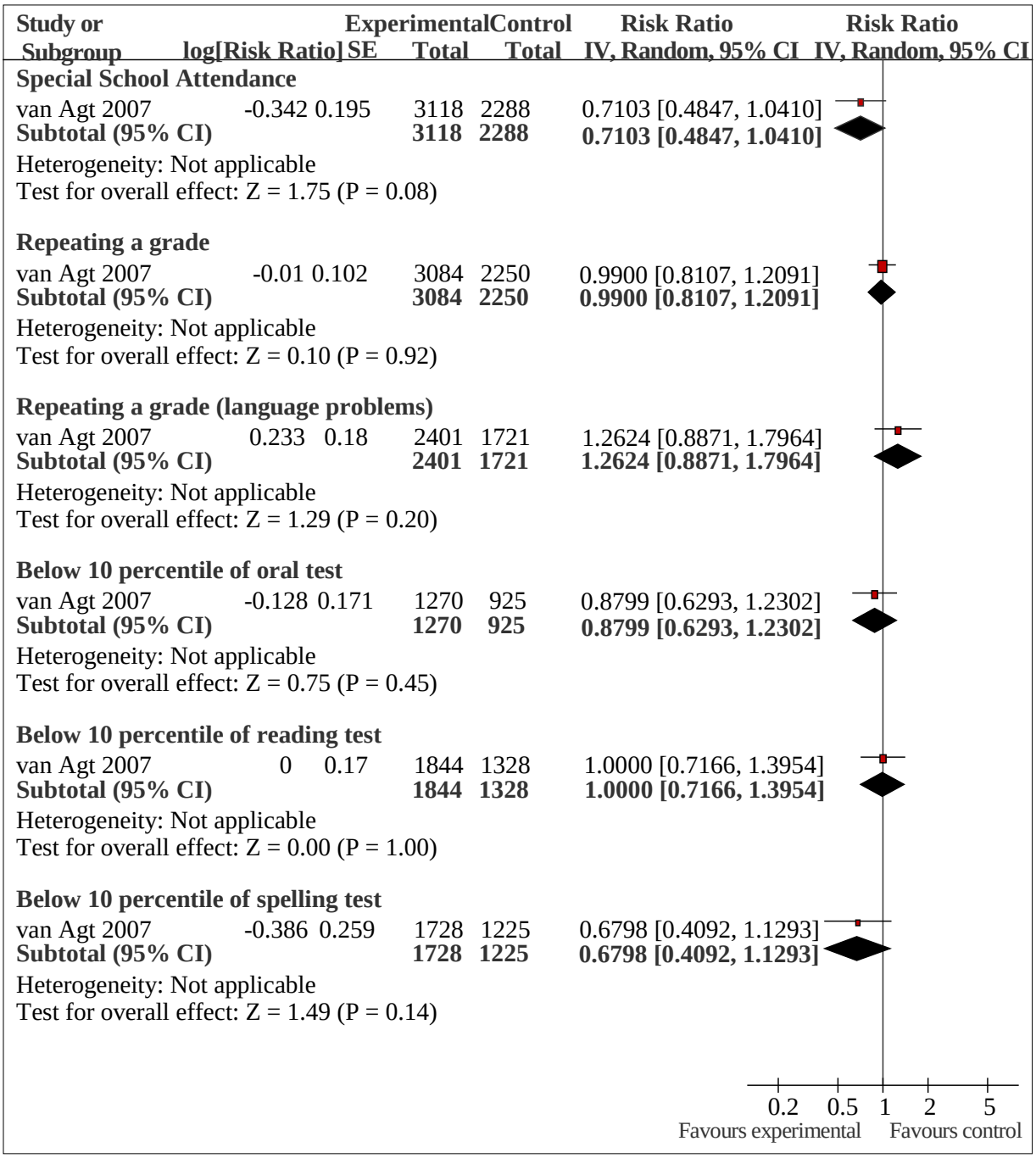
Forest Plot 1.1: Effectiveness of Developmental Delay Screening - Referral Rates to Intervention (Screening Tool - ASQ-II)



Forest plot 1.2: Effectiveness of Developmental delay screening – Time to intervention referral (Screening tool - ASQ-II)


Study or Subgroup
log[Rate Ratio]
SE
Experimental Total
Control Total
Rate Ratio IV, Random, 95% CI
Rate Ratio IV, Random, 95% CI
Screening with Office support
Guevara 2013-A -1.204 0.241 704 695 0.3000 [0.1871, 0.4811]
Subtotal (95% CI) 704 695 0.3000 [0.1871, 0.4811]
Heterogeneity: Not applicable
Test for overall effect: Z = 5.00 (P < 0.00001)
Screening without Office support
Guevara 2013-B -1.008 0.241 693 695 0.3649 [0.2276, 0.5853]
Subtotal (95% CI) 693 695 0.3649 [0.2276, 0.5853]
Heterogeneity: Not applicable
Test for overall effect: Z = 4.18 (P < 0.0001)
0.1 0.2 0.5 1 2 5 10
Favours experimental
Favours control

Forest plot 1.3: Effectiveness of Developmental delay screening: Academic Performance - By outcome measures (VTO screening)



Evidence Set 2 Stage II KQ 3

- GRADE Evidence Profile Table 2.1: Effect of treatment on developmental delay outcomes

ES Table 2.1 GRADE Evidence Profile: Effect of treatment on developmental delay outcomes

Quality Assessment							Results	Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Effect of Treatment on Cognitive Function (RCTs)									
0[1]	N/A[2]	N/A[3]	N/A[4]	N/A[5]	N/A[6]	N/A[7]	0[8]	N/A[9]	Critical
Effect of Treatment on Academic Performance (RCTs)									
0[1]	N/A[2]	N/A[3]	N/A[4]	N/A[5]	N/A[6]	N/A[7]	0[8]	N/A[9]	Critical
Effect of Treatment in Incidence of Mental Health Conditions (RCTs)									
0[1]	N/A[2]	N/A[3]	N/A[4]	N/A[5]	N/A[6]	N/A[7]	0[8]	N/A[9]	Critical
Effect of Treatment on Gross and Fine Motor Skills (RCTs)									
0[1]	N/A[2]	N/A[3]	N/A[4]	N/A[5]	N/A[6]	N/A[7]	0[8]	N/A[9]	Critical
Effect of Treatment on Quality of Life (RCTs)									
0[1]	N/A[2]	N/A[3]	N/A[4]	N/A[5]	N/A[6]	N/A[7]	0[8]	N/A[9]	Critical
Effect of Treatment on Functionality as an Adult (RCTs)									
0[1]	N/A[2]	N/A[3]	N/A[4]	N/A[5]	N/A[6]	N/A[7]	0[8]	N/A[9]	Critical

[1]No studies were found that met the inclusion criteria of this review for this treatment or outcomes
[2]No studies were found that met the inclusion criteria of this review for this treatment or outcomes
[3]Risk of bias cannot be assessed
[4]Inconsistency cannot not be assessed
[5]Indirectness cannot be assessed
[6]Imprecision cannot be assessed
[7]Other considerations cannot be assessed
[8]There are not studies to provide data on the effect of this treatment for these outcomes
[9]Since there no studies the overall quality of the evidence cannot be determined

Evidence Set 3 Stage II AKQ1

- GRADE Evidence Profile Table 3.1: Effect of treatment on developmental delay outcomes
- GRADE Summary of Findings Table 3.2: Effect of treatment on developmental delay outcomes
- Forest Plots 3.1 and 3.2
  - 3.1 Effect of treatment on language impairment
  - 3.2 Effect on social and personal activities of daily living (adaptive functioning - socialization)

ES Table 3.1 GRADE Evidence Profile: Effect of treatment on developmental delay outcomes

Quality assessment							No of patients		Effect	Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Control	SMD / MD (95% CI)		
Effect on language impairment (measured with: objectively; Better indicated by higher values)											
3[1]	randomised trials	serious[2]	no serious inconsistency[3]	no serious indirectness[4]	no serious imprecision[5]	none[6]	117	122	SMD 0.8114 higher (0.0241 to 1.5987 higher)	⊕⊕⊕O MODERATE	CRITICAL
Effect on Adaptive functioning (measured with: objectively; Better indicated by higher values)											
1[7]	randomised trials	serious[8]	no serious inconsistency[9]	no serious indirectness[10]	serious[11]	none[6]	71	84	MD 0.6000 higher (3.0495 lower to 4.2495 higher)	⊕⊕OO LOW	CRITICAL
Effect on Gross & Fine Motor Skills (measured with: objectively; Better indicated by higher values)											
0[12]	N/A[13]	NA[14]	N/A[15]	N/A[16]	N/A[17]	N/A[18]	-	-	-[19]	N/A[20]	IMPORTANT
Effect on Performance and Cognition											
0[12]	N/A[13]	N/A[14]	N/A[15]	N/A[16]	N/A[17]	N/A[18]	-	-	-[19]	N/A[20]	CRITICAL

ES Table 3.2 GRADE Summary of Findings Table: Effect of treatment on developmental delay outcomes

Outcomes	Illustrative comparative risks* (95% CI) Assumed risk Control	Corresponding risk Intervention	Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence Comments (GRADE)
Effect on language impairment objectively		The mean effect on language impairment in the intervention groups was **0.8114 standard deviations higher** (0.0241 to 1.5987 higher)		239 (3 studies[1])	⊕⊕⊕⊖ **moderate**[2,3,4,5,6]
Effect on Adaptive functioning objectively		The mean effect on adaptive functioning in the intervention groups was **0.6000 higher** (3.0495 lower to 4.2495 higher)		155 (1 study[7])	⊕⊕⊖⊖ **low**[8,9,10,11,6]

CI: Confidence interval;

GRADE Working Group grades of evidence

High quality: Further research is very unlikely to change our confidence in the estimate of effect.

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

Very low quality: We are very uncertain about the estimate.

[1] 1) Hund-Reid 2013; 2) Buschmann 2009; 3) Glogowska 2000

[2] Using Cochrane's Risk of Bias tool, for this outcome two studies were rated as unclear risk of bias, one study was rated as high risk of bias. Across studies, there was a lack of certainty (unclear ratings) regarding sequence generation (100%), and allocation concealment (33%) ; and high risk of bias associated with blinding of participants and personnel (100%); incomplete outcome reporting (33%) and other risk of bias (33%; i.e. baseline characteristics, pre-hoc power analysis, sample size <30 per arm). Given that most of the information is from studies at moderate risk, this body of evidence was downgraded for serious study limitations.

[3] The statistical heterogeneity is high [Chi2=12.60, df = 2 (P=0.002); I2=84%] but the direction of the effect is consistent and the confidence intervals overlap across most studies. The statistical heterogeneity is most likely due to small versus large treatment effects observed across studies. This body of evidence was not downgraded for inconsistency.

[4] Three RCTs provided data for this outcome. All studies included mixed gender children with ages ranging from 24.7 months to 66.68 months. In all studies the intervention group received a language intervention delivered in a primary care setting (two studies) or a school setting (one study) by either language therapists, a pediatric neurologist or education assistants. The control group received no intervention. The outcome of language impairment was accessed using DIBEL, SETK-2 and BLDS cross the three studies. Intervention lengths ranged from 14 weeks to 9 months; follow-up was immediate post in all studies. The studies were conducted in Germany, the UK and Canada. There were no serious concerns regarding indirectness for this body of evidence and it was not downgraded.

[5] The sample size is not adequate i.e. < 300 (117 intervention arm, 122 control arm) but the pooled effect estimate is precise and confidence interval does not include the null value "0" [SMD= 0.8114 (0.0241, 1.5987)]. This body of evidence was not downgraded for serious concerns regarding imprecision.

[6] There were too few studies (n<10) to assess publication bias.

[7] 1) Glogowska 2000

[8] Using Cochrane's Risk of Bias tool, for this outcome the study was rated as unclear risk of bias. In this study, there was a lack of certainty (unclear rating) regarding sequence generation; and high risk of bias associated with blinding of participants and personnel. Given that most of the information is from a study a moderate risk, this body of evidence was downgraded for serious study limitations

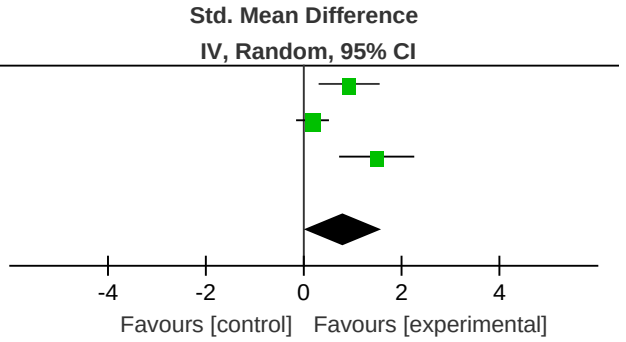
[9] The statistical heterogeneity across studies could not be assessed due to only one study providing data for this outcome.
[10] One study provided data for this outcome. The study included mixed gender children with a mean age of 34.2 months. The intervention consisted of one-on-one speech and language therapy with trained speech and language therapists, over 8.4 months. The outcome of Social and personal activities of daily living (Socialization, adaptive functioning) was accessed using VABS. Follow-up was immediate post. The study was conducted in the UK and was published in 2000. There were no serious concerns regarding indirectness for this body of evidence and it was not downgraded.
[11] The sample size is not adequate i.e. < 300 (71 intervention arm, 84 control arm) and the pooled effect estimate is not precise with confidence interval including the null value "0" [MD= 0.6000 (-3.0495, 4.2495)]. This body of evidence was downgraded for serious concerns regarding imprecision.
[12] No studies were found that met the inclusion criteria of this review for this treatment or outcomes
[13] No studies were found that met the inclusion criteria of this review for this treatment or outcomes
[14] Risk of bias cannot be assessed
[15] Inconsistency cannot not be assessed
[16] Inconsistency cannot not be assessed
[17] Inconsistency cannot not be assessed
[18] Inconsistency cannot not be assessed
[19] Inconsistency cannot not be assessed
[10] Since there no studies the overall quality of the evidence cannot be determined

Forest Plot 3.1 Effect of treatment on language impairment

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Buschmann, 2009	11.05	6.284	24	4.85	6.843	23	32.6%	0.9288 [0.3242, 1.5334]
Glogowska, 2000	2.6	1.075	71	2.4	1.138	84	37.9%	0.1794 [-0.1373, 0.4960]
Hund-Reid, 2013	12.16	6.687	22	2.36	6.02	15	29.6%	1.4916 [0.7433, 2.2399]
Total (95% CI)			**117**			**122**	**100.0%**	**0.8114 [0.0241, 1.5987]**

Heterogeneity: Tau² = 0.40; Chi² = 12.60, df = 2 (P = 0.002); I² = 84%

Test for overall effect: Z = 2.02 (P = 0.04)

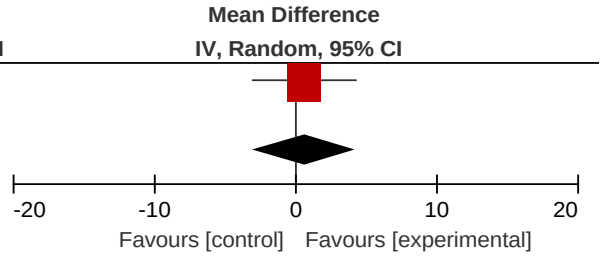


Forest Plot 3.2 Effect on social and personal activities of daily living (adaptive functioning - socialization)

Study or Subgroup	Mean Difference	SE	Experimental Total	Control Total	Weight	Mean Difference IV, Random, 95% CI
Glogowska, 2000	0.6	1.862	71	84	100.0%	0.6000 [-3.0495, 4.2495]
Total (95% CI)			**71**	**84**	**100.0%**	**0.6000 [-3.0495, 4.2495]**

Heterogeneity: Not applicable

Test for overall effect: Z = 0.32 (P = 0.75)



Evidence Set 4 Stage III KQ5: Test Properties

- Findings Summary Table 1: Test properties for diagnosis of ASD
- Findings Summary Table 2: Test Properties for diagnosis of AD
- Findings Summary Table 3: Test properties for diagnosis of DD
- Findings Summary Table 4: Test properties for diagnosis of DD – sub-domains
- Forest Plot 4.1 Diagnosis of ASD: Sensitivity and Specificity of MCHAT
- Forest Plot 4.2 Diagnosis of ASD: Positive and Negative Likelihood Ratio of MCHAT
- ROC Plot Diagnosis of ASD: Area under the curve (AUC) for MCAT

Findings Summary Table 1: Test Properties for diagnosis of Autism spectrum disorders (ASD)

Screening Test	Study Design	# of studies & Sample size	Cut-off value	Reference standard	Results
MCHAT alone	Cohort	9; 70,816	3 of total items failed or 2 of 6 critical items failed	Clinical-Diagnostic evaluation	**Sensitivity:** 78.0% (64.0% to 88.0%)
					Specificity: 69.0% (47.0% to 85.0%)
					Positive Likelihood Ratio: 2.5 (1.4 to 4.5)
					Negative Likelihood Ratio: 0.32 (0.19 to 0.52)
					Positive Predictive Value: 30.3% (17.3% to 50.4%)
					Negative Predictive Value: 94.8% (88.3% to 97.6%)
					False Positive Rate: 31.0% (15.0% to 53.0%)
					False Negative Rate: 22.0% (12.0% to 36.0%)
SCQ	Cohort	1; 67	12 score	Clinical-Diagnostic evaluation	**Sensitivity:** 69.2% (54.9% to 81.3%)
					Specificity: 56.3% (29.9% to 80.3%)
					Positive Likelihood Ratio: 1.6 (0.9 to 2.8)
					Negative Likelihood Ratio: 0.55 (0.30 to 0.99)
					Positive Predictive Value: 83.7% (69.3% to 93.2%)
					Negative Predictive Value: 36.0% (18.0% to 57.5%)
					False Positive Rate: 43.7% (19.7% to 70.1%)
					False Negative Rate: 30.8% (18.7% to 45.1%)
SCQ	Case-control	2; 121	15 score	Clinical-Diagnostic evaluation	**Sensitivity:** 57.3% (47.4% to 67.1%)
					Specificity: 94.5% (88.9% to 100%)
					Positive Likelihood Ratio: 4.3 (1.1 to 17.1)
					Negative Likelihood Ratio: 0.46 (0.33 to 0.59)
					Positive Predictive Value: 90.9% (81.8% to 100%)
					Negative Predictive Value: 63.5% (61.5% to 65.4%)
					False Positive Rate: 5.5% (0% to 11.1%)
					False Negative Rate: 42.7% (32.9% to 52.6%)
PEDS	Cohort	1; 52	>= 2 predictive concerns	Clinical-Diagnostic evaluation	**Sensitivity:** 56.7% (37.4% to 74.5%)
					Specificity: 40.9% (20.7% to 63.7%)
					Positive Likelihood Ratio: 0.96 (0.6 to 1.5)
					Negative Likelihood Ratio: 1.06 (0.55 to 2.02)
					Positive Predictive Value: 56.7% (37.4% to 74.5%)
					Negative Predictive Value: 40.9% (20.7% to 63.7%)
					False Positive Rate: 59.1% (36.3% to 79.3%)
					False Negative Rate: 43.3% (25.5% to 62.6%)

Findings Summary Table 2: Test properties for diagnosis of Autism disorder (AD)

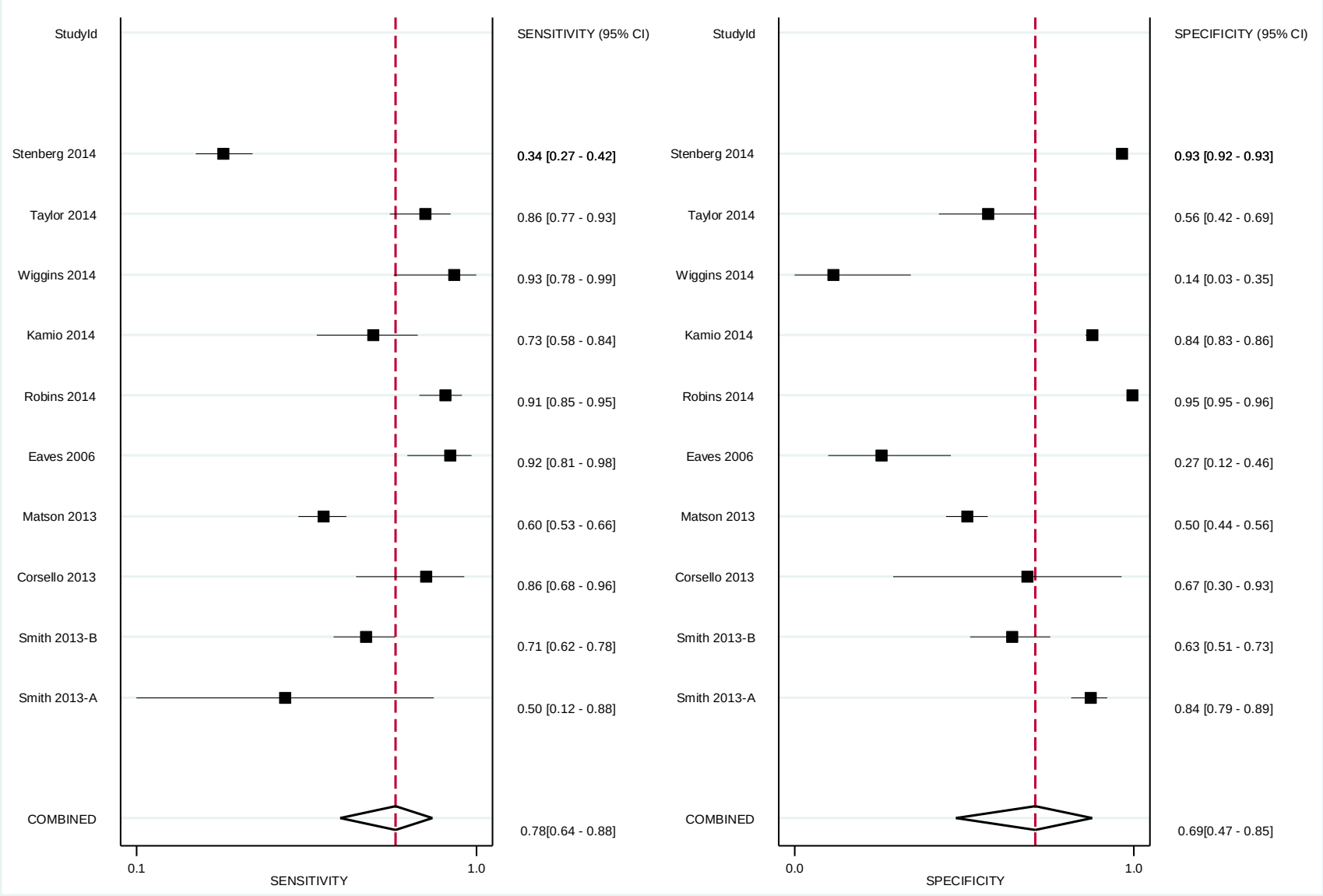
Screening Test	Study Design	# of studies & Sample size	Cut-off value	Reference standard	Results
MCHAT alone	Cohort	1; 1,851	3 of total items failed or 1 of 6 critical items failed	Clinical-Diagnostic evaluation	**Sensitivity:** 62.5% (35.4% to 84.8%)
					Specificity: 83.2% (81.4% to 84.9%)
					Positive Likelihood Ratio: 3.7 (2.5 to 5.5)
					Negative Likelihood Ratio: 0.45 (0.24 to 0.85)
					Positive Predictive Value: 3.1% (1.5% to 5.7%)
					Negative Predictive Value: 99.6% (99.2% to 99.9%)
					False Positive Rate: 16.8% (15.1% to 18.6%)
					False Negative Rate: 37.5% (15.2% to 64.6%)
CHAT	Case-control	1; 44	risk criteria (medium & high)	Clinical-Diagnostic evaluation	**Sensitivity:** 65.4% (44.3% to 82.8%)
					Specificity: 100% (81.5% to 100%)
					Positive Likelihood Ratio: NA
					Negative Likelihood Ratio: 0.35 (0.20 to 0.59)
					Positive Predictive Value: 100% (80.5% to 100%)
					Negative Predictive Value: 66.7% (46.0% to 83.5%)
					False Positive Rate: 0% (0% to 18.5%)
					False Negative Rate: 34.6% (17.2% to 55.7%)
STAT	Case-control	1; 52	optimal (2.13)	Clinical-Diagnostic evaluation	**Sensitivity:** 100% (86.8% to 100%)
					Specificity: 84.6% (65.1% to 95.6%)
					Positive Likelihood Ratio: 6.5 (2.6 to 16.0)
					Negative Likelihood Ratio: NA
					Positive Predictive Value: 86.7% (69.3 to 96.2%)
					Negative Predictive Value: 100% (84.6 to 100%)
					False Positive Rate: 15.4% (4.4% to 34.9%)
					False Negative Rate: 0% (0% to 13.2%)

Findings Summary Table 3: Test Properties for diagnosis of Developmental Delay (DD)

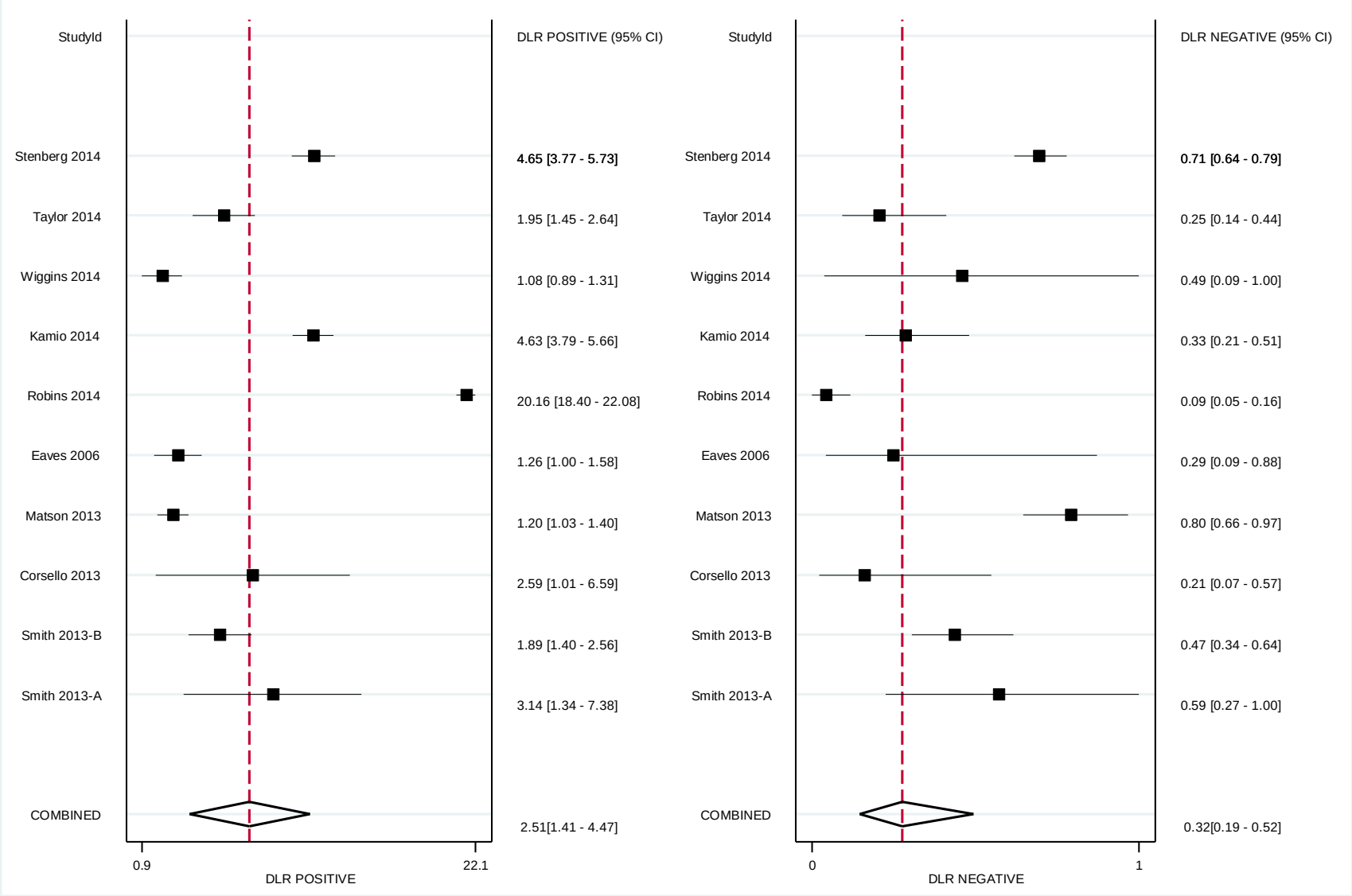
Screening Test	Study Design	# of studies & Sample size	Cut-off value	Reference standard	Results
ASQ*	Cohort	4; 1,001	< 2 SDs below mean on any domain or <1 SD on two domains	Diagnostic evaluation in one, BSID in two & BDI in one	**Sensitivity:** 55.0% (47.1% to 66.7%)
					Specificity: 86.0% (38.6% to 94.3%)
					Positive Likelihood Ratio 4.2 (1.1 to 8.2)
					Negative Likelihood Ratio 0.61 (0.47 to 0.86)
					Positive Predictive Value: 41.4% (23.2% to 71.4%)
					Negative Predictive Value: 84.9% (70.8% to 95.6%)
					False Positive Rate: 14.0% (5.7% to 61.4%)
					False Negative Rate: 45.0% (33.3% to 52.9%)
PEDS	Cohort	1; 331	>= 2 predictive concerns	Clinical-Diagnostic evaluation	**Sensitivity:** 41.1% (24.7% to 59.3%)
					Specificity: 89.3% (85.1% to 92.5%)
					Positive Likelihood Ratio 3.8 (2.3 to 6.4)
					Negative Likelihood Ratio 0.66 (0.5 to 0.88)
					Positive Predictive Value: 30.4% (17.7% to 45.8%)
					Negative Predictive Value: 92.9% (89.3% to 95.7%)
					False Positive Rate: 10.7% (7.5% to 14.9%)
					False Negative Rate: 58.9% (40.7% to 85.3%)

*Results reported for ASQ are medians.

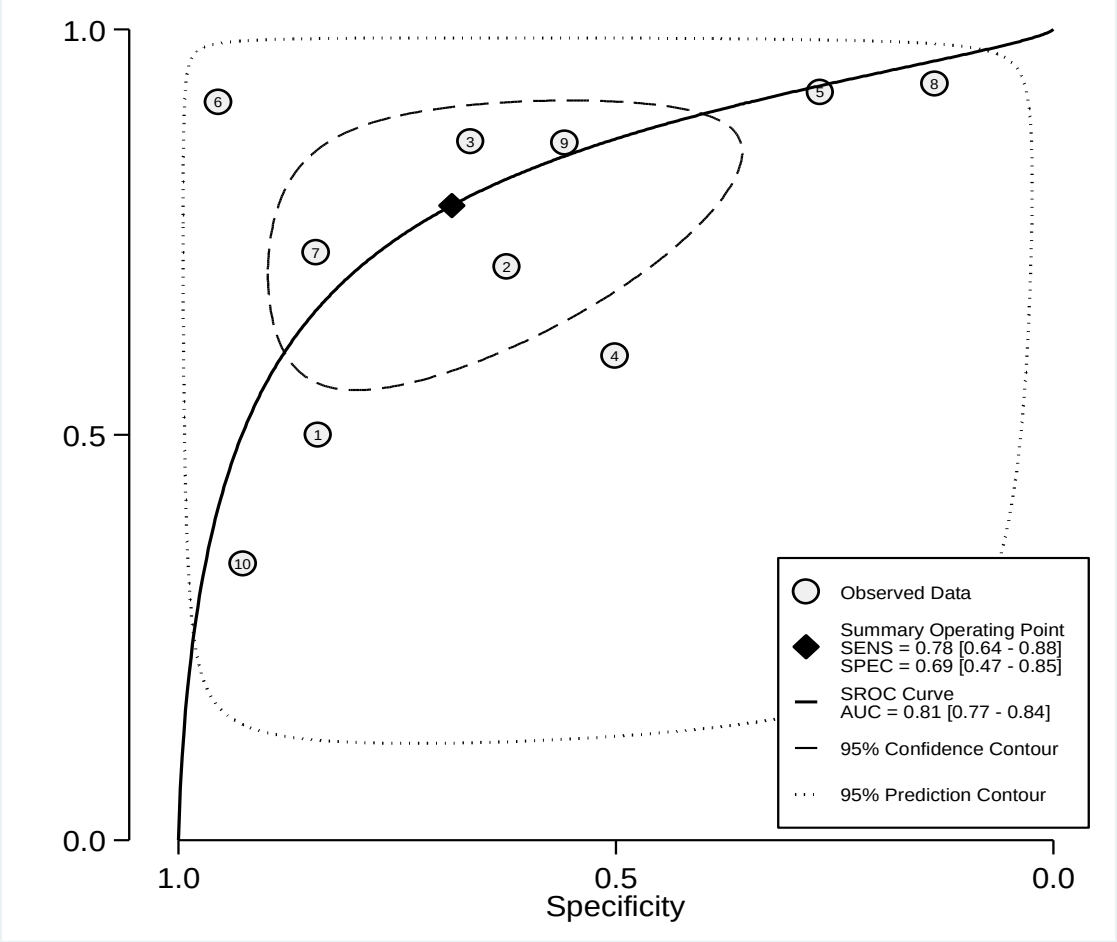
Forest Plot 4.1 Diagnosis of ASD: Sensitivity and Specificity for MCHAT



Forest Plot 4.2. Diagnosis of ASD: Positive and Negative Likelihood for MCHAT



ROC Plot. Diagnosis of ASD: Area under the curve (AUC) for MCHAT



Appendix 1: Stage I (Screening) Search Strategy

OVID-Medline

September 16th, 2015
1. exp Child Development Disorders, Pervasive/
2. Developmental Disabilities/
3. exp *Movement Disorders/
4. exp *Psychomotor Disorders/
5. exp Communication Disorders/
6. *Cognition Disorders/
7. or/1-6
8. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
9. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
10. ((cognitive or cognition) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
11. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
12. or/8-11
13. 7 or 12
14. well child.ti.
15. Mass Screening/
16. (screen or screened or screening).ti.
17. *early diagnosis/
18. or/15-17
19. 13 or 14
20. 18 and 19
21. limit 20 to "preschool child (2 to 5 years)"
22. exp *child, preschool/ or infant/
23. (pediatric* or paediatric* or child* or infant* or preschool or pre school).ti,ab,jn.
24. 22 or 23
25. 20 and 24
26. 21 or 25
27. mutation*.ti.
28. exp *genetics/
29. *Infant, Premature/
30. exp *Chromosomes/
31. or/27-30
32. 26 not 31
33. genetic screen*.ti,ab.
34. 32 not 33
35. limit 34 to (english or french)
36. limit 35 to (comment or editorial or letter or news or newspaper article)
37. 35 not 36

OVID-EMBASE
September 16th, 2015
1. "parents evaluation of developmental status"/
2. *developmental disorder/
3. developmental stage/ or exp postnatal development/
4. motor coordination/ or motor dysfunction/ or developmental coordination disorder/ or psychomotor disorder/ or motor performance/
5. cognitive development/ or language development/ or speech development/
6. social adaptation/ or social competence/
7. early intervention/
8. early childhood intervention/
9. developmental language disorder/ or language delay/ or speech delay/
10. or/1-9
11. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
12. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
13. ((cognitive or cognition or social) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
14. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
15. or/11-14
16. *early diagnosis/
17. mass screening/ or newborn screening/ or screening/ or screening test/
18. (screen or screened or screening).ti.
19. well child.ti.
20. or/16-19
21. 10 and 20
22. 15 and 20
23. developmental screening/
24. 21 or 22 or 23
25. (pediatric* or paediatric* or child* or infant* or preschool or pre school).ti,ab,jn.
26. limit 24 to (infant or child or preschool child <1 to 6 years>)
27. 24 and 25
28. 26 or 27
29. limit 28 to human
30. limit 29 to (english or french)
31. limit 30 to (book or book series or conference abstract or conference paper or conference proceeding or editorial or letter)
32. 30 not 31

OVID-PsycINFO
September 16th, 2015
1. delayed development/ or delayed speech/ or language delay/

2. developmental disabilities/ or aspergers syndrome/ or autism/ or exp communication disorders/ or exp learning disorders/ or exp pervasive developmental disorders/
3. developmental stages/
4. exp childhood development/
5. developmental age groups/ or exp motor development/ or physical development/ or exp precocious development/ or exp psychological development/
6. exp perceptual development/
7. communication skills/ or social cognition/ or emotional maturity/
8. or/1-7
9. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
10. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
11. ((cognitive or cognition or social) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
12. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
13. or/9-12
14. screening/ or health screening/
15. well child.ti.
16. (screen or screened or screening).ti.
17. exp screening tests/
18. *early intervention/
19. or/14-18
20. 8 and 19
21. 13 and 19
22. 20 or 21
23. limit 22 to (140 infancy or 160 preschool age)
24. (pediatric* or paediatric* or child* or infant* or preschool or pre school).ti,ab,jn.
25. 22 and 24
26. 23 or 25
27. limit 26 to human
28. limit 27 to (english or french)
29. limit 28 to (bibliography or "column/opinion" or "comment/reply" or dissertation or editorial or encyclopedia entry or letter or review-book or review-media or review-software & other)
30. 28 not 29

Appendix 2: Stage II (Treatment) Systematic Reviews Search Strategy

Medline-OVID

Last Searched September 16th, 2015

1. exp Child Development Disorders, Pervasive/
2. Developmental Disabilities/
3. exp *Movement Disorders/
4. exp *Psychomotor Disorders/
5. exp Communication Disorders/
6. *Cognition Disorders/
7. or/1-6
8. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
9. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
10. ((cognitive or cognition) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
11. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
12. or/8-11
13. 7 or 12
14. exp Child Development Disorders, Pervasive/dh, dt, mo, pc, rh, th [Diet Therapy, Drug Therapy, Mortality, Prevention & Control, Rehabilitation, Therapy]
15. Developmental Disabilities/dh, dt, mo, pc, rh, th
16. exp Movement Disorders/dh, dt, mo, pc, rh, th
17. exp Psychomotor Disorders/dh, dt, mo, pc, rh, th
18. exp Communication Disorders/dh, dt, mo, pc, rh, th
19. Cognition Disorders/dh, dt, mo, pc, rh, th
20. 14 or 15 or 16 or 17 or 18 or 19
21. exp *Therapeutics/
22. 13 and 21
23. limit 22 to (english or french)
24. limit 23 to yr="2009 -Current"
25. limit 24 to (meta analysis or systematic reviews)
26. (meta anal* or metaanal* or systematic).ti,ab.
27. 24 and 26
28. 25 or 27
29. limit 20 to (english or french)
30. limit 29 to yr="2009 -Current"
31. limit 30 to (meta analysis or systematic reviews)
32. 26 and 30
33. 31 or 32
34. 28 or 33

35. limit 34 to ("infant (1 to 23 months)" or "preschool child (2 to 5 years)" or "child (6 to 12 years)")
36. (preschool or young child or young children).tw.
37. 34 and 36
38. 35 or 37

EMBASE-OVID

Last Searched September 16th, 2015

1. mutation*.ti.
2. mutation*.ti.
3. exp *genetics/
4. exp *chromosome/
5. 2 or 3 or 4
6. developmental disorder/dm, dt, rh, su, th
7. exp thought disorder/dt, rh, su, th [Drug Therapy, Rehabilitation, Surgery, Therapy]
8. exp behavior disorder/dm, dt, pc, rt, rh, si, su, th [Disease Management, Drug Therapy, Prevention, Radiotherapy, Rehabilitation, Side Effect, Surgery, Therapy]
9. motor dysfunction/dm, dt, rh, su, th or developmental coordination disorder/dm, dt, rh, su, th or psychomotor disorder/dm, dt, rh, su, th or motor performance/dm, dt, rh, su, th
10. exp therapy/
11. autism/dm, dt, rh, su, th
12. developmental language disorder/dm, dt, rh, su, th or language delay/dm, dt, rh, su, th or speech delay/dm, dt, rh, su, th
13. 6 or 7 or 8 or 9 or 11 or 12
14. motor coordination/
15. developmental stage/
16. developmental screening/
17. social adaptation/ or social competence/
18. cognitive development/ or language development/ or speech development/
19. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti,ab.
20. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti,ab.
21. ((cognitive or cognition or social) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti,ab.
22. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti,ab.
23. 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22
24. exp therapy/ or early intervention/ or early childhood intervention/
25. (treatment or therapy).ti.
26. 24 or 25
27. 23 and 26
28. 13 or 27
29. limit 28 to (english or french)

30. limit 29 to (child or preschool child <1 to 6 years>)
31. limit 30 to (book or book series or conference abstract or conference paper or conference proceeding or editorial or letter)
32. 30 not 31
33. limit 32 to yr="2009 -Current"
34. limit 33 to (meta analysis or "systematic review")
35. (meta anal* or metaanal* or systematic).ti,ab.
36. 33 and 35
37. 34 or 36
38. 37 not 5

PsycINFO-OVID

Least Searched September 16th, 2015

1. delayed development/ or delayed speech/ or language delay/
2. developmental disabilities/ or aspergers syndrome/ or autism/ or exp communication disorders/ or exp learning disorders/ or exp pervasive developmental disorders/
3. developmental stages/
4. exp childhood development/
5. developmental age groups/ or exp motor development/ or physical development/ or exp precocious development/ or exp psychological development/
6. exp perceptual development/
7. communication skills/ or social cognition/ or emotional maturity/
8. or/1-7
9. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
10. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
11. ((cognitive or cognition or social) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
12. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
13. or/9-12
14. 8 or 13
15. exp Treatment/
16. treatment/ or exp counseling/ or early intervention/ or exp intervention/ or exp "prescribing (drugs)"/ or exp "side effects (treatment)"/ or exp treatment duration/ or treatment effectiveness evaluation/ or treatment outcomes/
17. (treatment or therapy).ti.
18. 15 or 16 or 17
19. 14 and 18
20. limit 19 to (english or french)
21. limit 20 to childhood

22. limit 21 to ("0200 book" or "0240 authored book" or "0280 edited book" or "0400 dissertation abstract" or (abstract collection or bibliography or chapter or "column/opinion" or "comment/reply" or dissertation or editorial or letter or review-book))
23. 21 not 22
24. limit 23 to yr="2009 -Current"
25. limit 24 to ("0830 systematic review" or 1200 meta analysis)
26. (meta anal* or metaanal* or systematic).ti,ab,pt.
27. 24 and 26
28. 25 or 27

Cochrane Database of Systematic Reviews-OVID

September 16th, 2015

1. (cerebral palsy or learning disabilit* or autism or asperger*).tw.
2. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
3. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
4. ((cognitive or cognition) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
5. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
6. or/1-5
7. (pediatric* or paediatric* or child* or infant* or preschool or pre school).ti,ab,jn.
8. 6 and 7
9. limit 8 to full systematic reviews
10. limit 9 to last 10 years
11. limit 9 to last 5 years

Appendix 3: Stage II (Treatment) RCT Search Strategy

Medline-OVID

Last Searched: September 16th, 2015

1. Developmental Disabilities/
2. exp Movement Disorders/
3. exp Psychomotor Disorders/
4. exp Communication Disorders/
5. Cognition Disorders/
6. or/1-5
7. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
8. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
9. ((cognitive or cognition) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
10. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
11. or/7-10
12. 6 or 11
13. exp *Therapeutics/
14. 12 and 13
15. (therap* or treatment?).ti.
16. 12 and 15
17. 14 or 16
18. Developmental Disabilities/dh, dt, mo, pc, rh, th
19. exp Movement Disorders/dh, dt, mo, pc, rh, th
20. exp Psychomotor Disorders/dh, dt, mo, pc, rh, th
21. exp Communication Disorders/dh, dt, mo, pc, rh, th
22. Cognition Disorders/dh, dt, mo, pc, rh, th
23. or/18-22
24. 17 or 23
25. limit 24 to (english or french)
26. limit 25 to yr="2000 -Current"
27. limit 26 to ("infant (1 to 23 months)" or "preschool child (2 to 5 years)" or "child (6 to 12 years)")
28. (preschool or young child or young children).tw.
29. 26 and 28
30. 27 or 29
31. random*.tw.
32. 30 and 31
33. limit 30 to (clinical trial or randomized controlled trial or controlled clinical trial)
34. 32 or 33

EMBASE-OVID

Last Searched: September 16th, 2015

1. developmental disorder/dm, dt, rh, su, th
2. exp thought disorder/dt, rh, su, th [Drug Therapy, Rehabilitation, Surgery, Therapy]
3. exp behavior disorder/dm, dt, pc, rt, rh, si, su, th [Disease Management, Drug Therapy, Prevention, Radiotherapy, Rehabilitation, Side Effect, Surgery, Therapy]
4. motor dysfunction/dm, dt, rh, su, th or developmental coordination disorder/dm, dt, rh, su, th or psychomotor disorder/dm, dt, rh, su, th or motor performance/dm, dt, rh, su, th
5. developmental language disorder/dm, dt, rh, su, th or language delay/dm, dt, rh, su, th or speech delay/dm, dt, rh, su, th
6. 1 or 2 or 3 or 4 or 5
7. motor coordination/
8. developmental stage/
9. developmental screening/
10. social adaptation/ or social competence/
11. cognitive development/ or language development/ or speech development/
12. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti,ab.
13. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti,ab.
14. ((cognitive or cognition or social) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti,ab.
15. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti,ab.
16. or/7-15
17. exp therapy/ or early intervention/ or early childhood intervention/
18. (treatment or therapy).ti.
19. 17 or 18
20. 16 and 19
21. 6 or 20
22. limit 21 to (english or french)
23. limit 22 to (child or preschool child <1 to 6 years>)
24. limit 23 to yr="2000 -Current"
25. limit 24 to randomized controlled trial
26. randomization/ or randomized controlled trial/ or "randomized controlled trial (topic)"/
27. 24 and 26
28. 25 or 27
29. limit 28 to (book or book series or conference abstract or conference paper or conference proceeding or editorial or letter)
30. 28 not 29

Cochrane Central-OVID

Last Searched: September 16th, 2015

1. Developmental Disabilities/
2. exp Movement Disorders/
3. exp Psychomotor Disorders/
4. exp Communication Disorders/
5. Cognition Disorders/
6. or/1-5
7. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
8. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
9. ((cognitive or cognition) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
10. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
11. or/7-10
12. 6 or 11
13. exp *Therapeutics/
14. 12 and 13
15. (therap* or treatment?).ti.
16. 12 and 15
17. 14 or 16
18. Developmental Disabilities/dh, dt, mo, pc, rh, th
19. exp Movement Disorders/dh, dt, mo, pc, rh, th
20. exp Psychomotor Disorders/dh, dt, mo, pc, rh, th
21. exp Communication Disorders/dh, dt, mo, pc, rh, th
22. Cognition Disorders/dh, dt, mo, pc, rh, th
23. or/18-22
24. 17 or 23
25. limit 24 to (english or french)
26. limit 25 to yr="2000 -Current"
27. autism.ti.
28. autism.ti,jn.
29. (autism or autistic or asperger*).ti.
30. 26 not 29
31. (child or children or adolescent or teen or pediatric or paediatric or school).mp.
32. adolescent/ or child/ or child, preschool/
33. 31 or 32
34. 30 and 33

PsycINFO-OVID

Last Searched: September 16th, 2015

1. delayed development/ or delayed speech/ or language delay/
2. developmental disabilities/ or exp communication disorders/ or exp learning disorders/
3. developmental stages/

4. exp childhood development/
5. developmental age groups/ or exp motor development/ or physical development/ or exp precocious development/ or exp psychological development/
6. communication skills/ or social cognition/ or emotional maturity/
7. exp perceptual development/
8. or/1-7
9. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
10. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
11. ((cognitive or cognition or social) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
12. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
13. or/9-12
14. 8 or 13
15. exp treatment/ or exp counseling/ or early intervention/ or exp intervention/ or exp "prescribing (drugs)"/ or exp "side effects (treatment)"/ or exp treatment duration/ or treatment effectiveness evaluation/ or treatment outcomes/
16. (treatment or therapy).ti.
17. 15 or 16
18. 14 and 17
19. limit 18 to (english or french)
20. limit 19 to childhood
21. limit 20 to yr="2000 -Current"
22. limit 21 to "2000 treatment outcome/clinical trial"
23. (random* not non-random*).tw.
24. 21 and 23
25. 22 or 24
26. limit 25 to ("0200 book" or "0240 authored book" or "0280 edited book" or "0400 dissertation abstract" or (abstract collection or bibliography or chapter or "column/opinion" or "comment/reply" or dissertation or editorial or letter or review-book))
27. 25 not 26

Appendix 4: Stage II (Treatment – Addendum) RCT Search Strategy

Medline-OVID

Last Searched: September 16th, 2015

1. Developmental Disabilities/
2. exp Movement Disorders/
3. exp Psychomotor Disorders/
4. exp Communication Disorders/
5. Cognition Disorders/
6. or/1-5
7. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
8. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
9. ((cognitive or cognition) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
10. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
11. or/7-10
12. 6 or 11
13. exp *Therapeutics/
14. 12 and 13
15. (therap* or treatment?).ti.
16. 12 and 15
17. 14 or 16
18. Developmental Disabilities/dh, dt, mo, pc, rh, th
19. exp Movement Disorders/dh, dt, mo, pc, rh, th
20. exp Psychomotor Disorders/dh, dt, mo, pc, rh, th
21. exp Communication Disorders/dh, dt, mo, pc, rh, th
22. Cognition Disorders/dh, dt, mo, pc, rh, th
23. or/18-22
24. 17 or 23
25. limit 24 to (english or french)
26. limit 25 to yr="2000 -Current"
27. limit 26 to ("infant (1 to 23 months)" or "preschool child (2 to 5 years)" or "child (6 to 12 years)")
28. (preschool or young child or young children).tw.
29. 26 and 28
30. 27 or 29
31. random*.tw.
32. 30 and 31
33. limit 30 to (clinical trial or randomized controlled trial or controlled clinical trial)
34. 32 or 33

EMBASE-OVID

Last Searched: September 16th, 2015

1. developmental disorder/dm, dt, rh, su, th
2. exp thought disorder/dt, rh, su, th [Drug Therapy, Rehabilitation, Surgery, Therapy]
3. exp behavior disorder/dm, dt, pc, rt, rh, si, su, th [Disease Management, Drug Therapy, Prevention, Radiotherapy, Rehabilitation, Side Effect, Surgery, Therapy]
4. motor dysfunction/dm, dt, rh, su, th or developmental coordination disorder/dm, dt, rh, su, th or psychomotor disorder/dm, dt, rh, su, th or motor performance/dm, dt, rh, su, th
5. developmental language disorder/dm, dt, rh, su, th or language delay/dm, dt, rh, su, th or speech delay/dm, dt, rh, su, th
6. 1 or 2 or 3 or 4 or 5
7. motor coordination/
8. developmental stage/
9. developmental screening/
10. social adaptation/ or social competence/
11. cognitive development/ or language development/ or speech development/
12. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti,ab.
13. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti,ab.
14. ((cognitive or cognition or social) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti,ab.
15. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti,ab.
16. or/7-15
17. exp therapy/ or early intervention/ or early childhood intervention/
18. (treatment or therapy).ti.
19. 17 or 18
20. 16 and 19
21. 6 or 20
22. limit 21 to (english or french)
23. limit 22 to (child or preschool child <1 to 6 years>)
24. limit 23 to yr="2000 -Current"
25. limit 24 to randomized controlled trial
26. randomization/ or randomized controlled trial/ or "randomized controlled trial (topic)"/
27. 24 and 26
28. 25 or 27
29. limit 28 to (book or book series or conference abstract or conference paper or conference proceeding or editorial or letter)
30. 28 not 29

Cochrane Central-OVID

Last Searched: September 16th, 2015

1. Developmental Disabilities/
2. exp Movement Disorders/
3. exp Psychomotor Disorders/
4. exp Communication Disorders/
5. Cognition Disorders/
6. or/1-5
7. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
8. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
9. ((cognitive or cognition) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
10. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
11. or/7-10
12. 6 or 11
13. exp *Therapeutics/
14. 12 and 13
15. (therap* or treatment?).ti.
16. 12 and 15
17. 14 or 16
18. Developmental Disabilities/dh, dt, mo, pc, rh, th
19. exp Movement Disorders/dh, dt, mo, pc, rh, th
20. exp Psychomotor Disorders/dh, dt, mo, pc, rh, th
21. exp Communication Disorders/dh, dt, mo, pc, rh, th
22. Cognition Disorders/dh, dt, mo, pc, rh, th
23. or/18-22
24. 17 or 23
25. limit 24 to (english or french)
26. limit 25 to yr="2000 -Current"
27. autism.ti.
28. autism.ti,jn.
29. (autism or autistic or asperger*).ti.
30. 26 not 29
31. (child or children or adolescent or teen or pediatric or paediatric or school).mp.
32. adolescent/ or child/ or child, preschool/
33. 31 or 32
34. 30 and 33

PsycINFO-OVID

Last Searched: September 16th, 2015

1. delayed development/ or delayed speech/ or language delay/

2. developmental disabilities/ or exp communication disorders/ or exp learning disorders/
3. developmental stages/
4. exp childhood development/
5. developmental age groups/ or exp motor development/ or physical development/ or exp precocious development/ or exp psychological development/
6. communication skills/ or social cognition/ or emotional maturity/
7. exp perceptual development/
8. or/1-7
9. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
10. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
11. ((cognitive or cognition or social) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
12. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).ti.
13. or/9-12
14. 8 or 13
15. exp treatment/ or exp counseling/ or early intervention/ or exp intervention/ or exp "prescribing (drugs)"/ or exp "side effects (treatment)"/ or exp treatment duration/ or treatment effectiveness evaluation/ or treatment outcomes/
16. (treatment or therapy).ti.
17. 15 or 16
18. 14 and 17
19. limit 18 to (english or french)
20. limit 19 to childhood
21. limit 20 to yr="2000 -Current"
22. limit 21 to "2000 treatment outcome/clinical trial"
23. (random* not non-random*).tw.
24. 21 and 23
25. 22 or 24
26. limit 25 to ("0200 book" or "0240 authored book" or "0280 edited book" or "0400 dissertation abstract" or (abstract collection or bibliography or chapter or "column/opinion" or "comment/reply" or dissertation or editorial or letter or review-book))
27. 25 not 26

Appendix 5: Stage III (Test Properties) Search Strategy

Test Properties Search Strategy

Medline-OVID

Last Searched September 16th, 2015

1. (The Communication and Symbolic Behavior Scales Developmental Profile Infant-Toddler).tw.
2. CSBS-DP.mp.
3. Get Ready to Read.mp.
4. GRTR-R.mp.
5. (Individual Growth and Development Indicators).mp.
6. IGDI.mp.
7. IGDIs.mp.
8. Parents' Evaluation of Developmental Status.mp.
9. evaluation of developmental status.mp.
10. Ages & Stages Questionnaires.mp.
11. Ages & Stages Questionnaire.mp.
12. (ASQ adj10 ages).mp.
13. YACHT-18.mp.
14. (Young Autism and other developmental disorders CHeckup Tool).mp.
15. (Young Autism and other developmental disorders Checkup Tool).mp.
16. (Screening Tool for Autism in Toddlers and Young Children).mp.
17. (STAT and autism).mp.
18. Early Screening of Autistic Traits Questionnaire.mp.
19. (ESAT and (Autism or autistic)).mp.
20. Checklist for Early Signs of Developmental Disorders.mp.
21. CESDD.mp.
22. Swedish MacArthur-Bates Communicative Development Inventory.mp.
23. Swedish Communication Screening.mp.
24. (Infant-Toddler Social and Emotional Assessment).mp.
25. (Bayley Scales of Infant and Toddler Development).mp.
26. Child Behavior Checklist.mp.
27. MacArthur Communication Developmental Inventory.mp.
28. Battelle Developmental Inventory.mp.
29. Modified Checklist for Autism in Toddlers.mp.
30. M-CHAT.mp.
31. Language Development Survey.mp.
32. Reynell Developmental Language Scales.mp.
33. Preschool language checklist.mp.
34. (preschool language adj3 checklist).mp.
35. Minnesota child development inventory.mp.
36. McCarthy Scales of Children* Abilities.mp.
37. Stanford-Binet.mp.
38. (AAPS-R or Templin-Darley).mp.

39. TALC-R.mp.
40. (Fluharty Preschool Speech and Language Screening Test).mp.
41. Early Screening Profiles.mp.
42. Lollipop test.mp.
43. Ten Question Screen.mp.
44. Wechsler Intelligence Test for Children.mp.
45. Wechsler Intelligence Test.mp.
46. Early Language Milestone Scale.mp.
47. Miller Assessment for Preschoolers.mp.
48. Northwestern Syntax Screening Test.mp.
49. Screening Kit of Language Development.mp.
50. skold.mp.
51. The Adelaide Psychomotor Screen.mp.
52. North Carolina Psychoeducational Screening test.mp.
53. (Wechsler Preschool and Primary Scale of Intelligence).mp.
54. Social Responsiveness Scale.mp.
55. Developmental Concerns Questionnaire.mp.
56. First Year Inventory.mp.
57. Language Evaluation Scale.mp.
58. Nipissing District Developmental Screen.mp.
59. Nipissing Developmental Screen.mp.
60. PPVT A.mp.
61. or/1-60
62. Social Communication Questionnaire.mp.
63. Screening Tool for Autism in Two-Year-Olds.mp.
64. Pervasive Developmental Disorders Screening Test.mp.
65. Checklist for Autism in Toddlers.mp.
66. Motor Quotient.mp.
67. Early Motor Pattern Profile.mp.
68. (Communication and Symbolic Behavior Scales).mp.
69. Capute Scales.mp.
70. Cognitive Adaptive Test Clinical Linguistic Auditory Milestone Scale.mp.
71. Parents' Evaluation of Developmental Status.mp.
72. Infant Development Inventory.mp.
73. Denver-II Developmental Screening Test.mp.
74. Child Development Review-Parent Questionnaire.mp.
75. Child Development Review.mp.
76. Child Development Inventory.mp.
77. Brigance Screens.mp.
78. Bayley Infant Neurodevelopmental Screen*.mp.
79. Ages & Stages Questionnaires.mp.
80. Battelle Developmental Inventory Screening Tool,.mp.
81. Battelle Developmental Inventory.mp.
82. (Rourke adj2 (baby or record*)).mp.
83. or/62-82
84. 61 or 83

85. ((development* delay or child* development) and (Screening tool* or checklist*)).mp.
86. 84 or 85
87. (sensitivity or specificity or predictive values or likelihood ratios).mp.
88. 86 and 87
89. exp "Sensitivity and Specificity"/
90. 86 and 89
91. 88 or 90
92. limit 91 to (comment or editorial or letter)
93. 91 not 92
94. limit 93 to (english or french)

EMBASE-OVID

Last Searched September 16th, 2015

1. (The Communication and Symbolic Behavior Scales Developmental Profile Infant-Toddler).tw.
2. CSBS-DP.mp.
3. Get Ready to Read.mp.
4. GRTR-R.mp.
5. (Individual Growth and Development Indicators).mp.
6. IGDI.mp.
7. IGDIs.mp.
8. Parents' Evaluation of Developmental Status.mp.
9. evaluation of developmental status.mp.
10. Ages & Stages Questionnaires.mp.
11. Ages & Stages Questionnaire.mp.
12. (ASQ adj10 ages).mp.
13. YACHT-18.mp.
14. (Young Autism and other developmental disorders CHeckup Tool).mp.
15. (Young Autism and other developmental disorders Checkup Tool).mp.
16. (Screening Tool for Autism in Toddlers and Young Children).mp.
17. (STAT and autism).mp.
18. Early Screening of Autistic Traits Questionnaire.mp.
19. (ESAT and (Autism or autistic)).mp.
20. Checklist for Early Signs of Developmental Disorders.mp.
21. CESDD.mp.
22. Swedish MacArthur-Bates Communicative Development Inventory.mp.
23. Swedish Communication Screening.mp.
24. (Infant-Toddler Social and Emotional Assessment).mp.
25. (Bayley Scales of Infant and Toddler Development).mp.
26. Child Behavior Checklist.mp.
27. MacArthur Communication Developmental Inventory.mp.
28. Battelle Developmental Inventory.mp.
29. Modified Checklist for Autism in Toddlers.mp.
30. M-CHAT.mp.

31. Language Development Survey.mp.
32. Reynell Developmental Language Scales.mp.
33. Preschool language checklist.mp.
34. (preschool language adj3 checklist).mp.
35. Minnesota child development inventory.mp.
36. McCarthy Scales of Children* Abilities.mp.
37. Stanford-Binet.mp.
38. (AAPS-R or Templin-Darley).mp.
39. TALC-R.mp.
40. (Fluharty Preschool Speech and Language Screening Test).mp.
41. Early Screening Profiles.mp.
42. Lollipop test.mp.
43. Ten Question Screen.mp.
44. Wechsler Intelligence Test for Children.mp.
45. Wechsler Intelligence Test.mp.
46. Early Language Milestone Scale.mp.
47. Miller Assessment for Preschoolers.mp.
48. Northwestern Syntax Screening Test.mp.
49. Screening Kit of Language Development.mp.
50. skold.mp.
51. The Adelaide Psychomotor Screen.mp.
52. North Carolina Psychoeducational Screening test.mp.
53. (Wechsler Preschool and Primary Scale of Intelligence).mp.
54. Social Responsiveness Scale.mp.
55. Developmental Concerns Questionnaire.mp.
56. First Year Inventory.mp.
57. Language Evaluation Scale.mp.
58. Nipissing District Developmental Screen.mp.
59. Nipissing Developmental Screen.mp.
60. PPVT A.mp.
61. or/1-60
62. Social Communication Questionnaire.mp.
63. Screening Tool for Autism in Two-Year-Olds.mp.
64. Pervasive Developmental Disorders Screening Test.mp.
65. Checklist for Autism in Toddlers.mp.
66. Motor Quotient.mp.
67. Early Motor Pattern Profile.mp.
68. (Communication and Symbolic Behavior Scales).mp.
69. Capute Scales.mp.
70. Cognitive Adaptive Test Clinical Linguistic Auditory Milestone Scale.mp.
71. Parents' Evaluation of Developmental Status.mp.
72. Infant Development Inventory.mp.
73. Denver-II Developmental Screening Test.mp.
74. Child Development Review-Parent Questionnaire.mp.
75. Child Development Review.mp.
76. Child Development Inventory.mp.

77. Brigance Screens.mp.
78. Bayley Infant Neurodevelopmental Screen*.mp.
79. Ages & Stages Questionnaires.mp.
80. Battelle Developmental Inventory Screening Tool,.mp.
81. Battelle Developmental Inventory.mp.
82. (Rourke adj2 (baby or record*)).mp.
83. or/62-82
84. 61 or 83
85. ((development* delay or child* development) and (Screening tool* or checklist*)).mp.
86. 84 or 85
87. (sensitivity or specificity or predictive values or likelihood ratios).mp.
88. 86 and 87
89. exp "Sensitivity and Specificity"/
90. 86 and 89
91. 88 or 90
92. screening test/
93. (development* delay or child* development or autism).mp.
94. 92 and 93
95. 91 or 94
96. limit 95 to (english or french)
97. limit 96 to (book or book series or conference abstract or conference paper or "conference review" or editorial or letter or note)
98. 96 not 97

PsycINFO-OVID

Last Searched September 16th, 2015

1. (The Communication and Symbolic Behavior Scales Developmental Profile Infant-Toddler).tw.
2. CSBS-DP.mp.
3. Get Ready to Read.mp.
4. GRTR-R.mp.
5. (Individual Growth and Development Indicators).mp.
6. IGDI.mp.
7. IGDIs.mp.
8. Parents' Evaluation of Developmental Status.mp.
9. evaluation of developmental status.mp.
10. Ages & Stages Questionnaires.mp.
11. Ages & Stages Questionnaire.mp.
12. (ASQ adj10 ages).mp.
13. YACHT-18.mp.
14. (Young Autism and other developmental disorders CHeckup Tool).mp.
15. (Young Autism and other developmental disorders Checkup Tool).mp.
16. (Screening Tool for Autism in Toddlers and Young Children).mp.
17. (STAT and autism).mp.

18. Early Screening of Autistic Traits Questionnaire.mp.
19. (ESAT and (Autism or autistic)).mp.
20. Checklist for Early Signs of Developmental Disorders.mp.
21. CESDD.mp.
22. Swedish MacArthur-Bates Communicative Development Inventory.mp.
23. Swedish Communication Screening.mp.
24. (Infant-Toddler Social and Emotional Assessment).mp.
25. (Bayley Scales of Infant and Toddler Development).mp.
26. Child Behavior Checklist.mp.
27. MacArthur Communication Developmental Inventory.mp.
28. Battelle Developmental Inventory.mp.
29. Modified Checklist for Autism in Toddlers.mp.
30. M-CHAT.mp.
31. Language Development Survey.mp.
32. Reynell Developmental Language Scales.mp.
33. Preschool language checklist.mp.
34. (preschool language adj3 checklist).mp.
35. Minnesota child development inventory.mp.
36. McCarthy Scales of Children* Abilities.mp.
37. Stanford-Binet.mp.
38. (AAPS-R or Templin-Darley).mp.
39. TALC-R.mp.
40. (Fluharty Preschool Speech and Language Screening Test).mp.
41. Early Screening Profiles.mp.
42. Lollipop test.mp.
43. Ten Question Screen.mp.
44. Wechsler Intelligence Test for Children.mp.
45. Wechsler Intelligence Test.mp.
46. Early Language Milestone Scale.mp.
47. Miller Assessment for Preschoolers.mp.
48. Northwestern Syntax Screening Test.mp.
49. Screening Kit of Language Development.mp.
50. skold.mp.
51. The Adelaide Psychomotor Screen.mp.
52. North Carolina Psychoeducational Screening test.mp.
53. (Wechsler Preschool and Primary Scale of Intelligence).mp.
54. Social Responsiveness Scale.mp.
55. Developmental Concerns Questionnaire.mp.
56. First Year Inventory.mp.
57. Language Evaluation Scale.mp.
58. Nipissing District Developmental Screen.mp.
59. Nipissing Developmental Screen.mp.
60. PPVT A.mp.
61. or/1-60
62. Social Communication Questionnaire.mp.
63. Screening Tool for Autism in Two-Year-Olds.mp.

64. Pervasive Developmental Disorders Screening Test.mp.
65. Checklist for Autism in Toddlers.mp.
66. Motor Quotient.mp.
67. Early Motor Pattern Profile.mp.
68. (Communication and Symbolic Behavior Scales).mp.
69. Capute Scales.mp.
70. Cognitive Adaptive Test Clinical Linguistic Auditory Milestone Scale.mp.
71. Parents' Evaluation of Developmental Status.mp.
72. Infant Development Inventory.mp.
73. Denver-II Developmental Screening Test.mp.
74. Child Development Review-Parent Questionnaire.mp.
75. Child Development Review.mp.
76. Child Development Inventory.mp.
77. Brigance Screens.mp.
78. Bayley Infant Neurodevelopmental Screen*.mp.
79. Ages & Stages Questionnaires.mp.
80. Battelle Developmental Inventory Screening Tool,.mp.
81. Battelle Developmental Inventory.mp.
82. (Rourke adj2 (baby or record*)).mp.
83. or/62-82
84. 61 or 83
85. ((development* delay or child* development) and (Screening tool* or checklist*)).mp.
86. 84 or 85
87. (sensitivity or specificity or predictive values or likelihood ratios).mp.
88. 86 and 87
89. exp "Sensitivity and Specificity"/
90. 86 and 89
91. 88 or 90
92. screening test/
93. (development* delay or child* development or autism).mp.
94. 92 and 93
95. 91 or 94
96. exp Test Validity/ or exp "Item Analysis (Test)"/ or exp Test Reliability/ or exp Psychometrics/
97. 86 and 96
98. 91 or 95 or 97
99. limit 98 to (english or french)
100. limit 99 to (chapter or "column/opinion" or "comment/reply" or dissertation or editorial or letter or review-book)
101. 99 not 100
102. limit 101 to childhood

Appendix 6: Contextual Questions Search Strategy

Medline-OVID
April 23 2014
1. exp Child Development Disorders, Pervasive/
2. Developmental Disabilities/
3. exp *Movement Disorders/
4. exp *Psychomotor Disorders/
5. exp Communication Disorders/
6. *Cognition Disorders/
7. or/1-6
8. ((motor or movement) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
9. ((speech or language) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
10. ((cognitive or cognition) adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
11. (development* adj2 (development or disabilit* or delay or impairment? or deficit? or disorder?)).tw.
12. or/8-11
13. 7 or 12
14. (screen or screening* or screened).tw.
15. well child.tw.
16. Mass Screening/
17. exp early diagnosis/
18. Public Health Surveillance/ or Population Surveillance/
19. surveillance.tw.
20. 14 or 15 or 16 or 17 or 18 or 19
21. 13 and 20
22. limit 21 to ("infant (1 to 23 months)" or "preschool child (2 to 5 years)")
23. exp child, preschool/ or infant/
24. 21 and 23
25. (pediatric* or paediatric* or child* or infant* or preschool or pre school).ti,ab,jn.
26. 21 and 25
27. 22 or 24 or 26
28. mutation*.ti.
29. exp *genetics/
30. *Infant, Premature/
31. exp *Chromosomes/
32. genetic screen*.ti,ab.
33. 28 or 29 or 30 or 31 or 32
34. 27 not 33
35. limit 34 to (english or french)
36. limit 35 to (comment or editorial or letter or news or newspaper article)
37. 35 not 36
38. limit 37 to yr="2009 - 2014"

39. exp continental population groups/
40. exp Ethnic Groups/
41. indians, north american/ or inuits/
42. first nations.tw.
43. (aboriginal? and canada).tw.
44. native canadians.tw.
45. (immigran* or new canadians).tw.
46. ((African or Asian or Indo or Columbian or Spanish or Chinese) adj2 Canadian?).mp.
47. Rural Population/
48. (rural adj (population? or area? or region?)).tw.
49. Rural Health/ or Rural Health Services/
50. Healthcare Disparities/
51. Social Class/
52. poverty/
53. socioeconomic.tw.
54. Socioeconomic Factors/
55. (poor or disadvantaged or poverty or social status).tw.
56. exp homeless persons/ or vulnerable populations/
57. exp "Costs and Cost Analysis"/
58. (cost or costs).tw.
59. *"patient acceptance of health care"/ or *patient compliance/ or *patient participation/ or patient satisfaction/ or patient preference/ or *treatment refusal/
60. ((parent? or guardian*) adj3 (acceptance or preference? or satisfaction or experience?)).tw.
61. (consumer? adj3 (acceptance or preference? or satisfaction or experience?)).tw.
62. (patient? adj3 (acceptance or perference? or satisfaction or experience?)).tw.
63. willingness to pay.tw.
64. ((conjoint or contingent) adj3 (valuation or analysis)).tw.
65. exp Canada/
66. (Canada or Canadian or Ontario or British Columbia or Alberta or Saskatchewan or Manitoba or Quebec or Nova Scotia or Prince Edward Island or Newfoundland or New Brunswick or Yukon or Northwest Territories or Nunavut).tw.
67. or/39-66
68. (Canada or Canadian or Ontario or British Columbia or Alberta or Saskatchewan or Manitoba or Quebec or Nova Scotia or Prince Edward Island or Newfoundland or New Brunswick or Yukon or Northwest Territories or Nunavut).ti.
69. 67 or 68
70. 38 and 69
71. screen time.tw.
72. 70 not 71
73. exp Infant, Newborn/
74. (prenatal or perinatal or newborn* or babies or premature).ti.
75. 73 or 74
76. 72 not 75

Appendix 7: Acknowledgements

Funding was provided by the Canadian Institutes of Health Research (www.cihr-irsc.gc.ca).

We are grateful to Sharon Peck-Reid for database management and formatting of the report. Dr. Jessie McGowan for being the peer reviewer of our search strategy. Dr. Alice Carter, Dr. Sharon Smile, Dr. Isabel Smith, Dr. Lisa Hartling, Dr. Brian Reichow and Dr. Leslie Anne Campbell who peer reviewed versions of this evidence synthesis. The Public Health Agency of Canada Scientific Research Manager, Dr. Lesley Dunfield contributed to the background chapter as part of original protocol developmental and review of drafts of the technical report. Similarly, the Developmental Delay Working Group of the Canadian Task Force for Preventive Health Care members, Dr. Paula Bauer, Dr. Denis Leduc, Dr. Patricia Parkin, Dr. Kevin Pottie, Dr. Brett Thombs and Dr. Marcello Tonelli provided comments on the protocol and initial analysis and technical report.