Screening for Developmental Delay in Early Childhood (ages 1-4)

Rachel Warren MA, Meghan Kenny MA, Teresa Bennett MD, PhD, Donna Fitzpatrick-Lewis MSW, Muhammad Usman Ali MD, Diana Sherifali PhD, Parminder Raina PhD

Affiliations: From the McMaster Evidence Review and Synthesis Centre (MERSC), Offord Centre for Child Studies, McMaster University, Hamilton, Ont.

Competing Interests: None declared by any author.

Contributors: All authors performed tasks involved in conducting the full systematic review. All authors reviewed, contributed revisions and approved the manuscript prior to submission

Acknowledgements: We are grateful to Sharon Peck-Reid for database management and formatting of the report. Maureen Rice for her screening contributions and for conducting the search. Dr. Jessie McGowan for being the peer reviewer of our search strategy. Dr. Alice Carter, Dr. Sharon Smile and Dr. Isabel Smith who peer reviewed the 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. 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.

Funding: Funding for this review is from the Canadian Institutes of Health Research. The views expressed herein are the opinions of the authors and do not necessarily represent the views of the Canadian Institutes of Health Research.

Corresponding Author: Donna Fitzpatrick-Lewis, McMaster Evidence Review and Synthesis Centre, 50 Main Street East, Hamilton ON L8N 1E9

Email: fitzd@mcmaster.ca

Abstract

Background: This systematic review synthesizes the effectiveness and harms of screening for developmental delay (DD) in children aged 1-4 years.

Methods: We searched Medline, Embase and PsychINFO answer the question of effectiveness of screening (no beginning date limitations to February 24th, 2014). See PROPSERO CRD42014009809.

Results: For effectiveness of screening, two studies met the inclusion criteria. One moderate quality study used ASQ-II as a screening tool and reported significantly more referrals to early intervention than the control group with a relative risk (RR) of 1.95 (95% CI 1.49, 2.54) in the intervention group with office support and an RR 1.71 (95% CI 1.30, 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 0.30 [95% CI 0.19, 0.48]), and a 64% shorter time to referral for the intervention group without office support (Rate Ratio 0.36 [95% CI 0.23, 0.59]), both compared to the control group. One low quality study using (VTO) Language Screening tool for mixed gender children aged 15 months at entry for language delay reported no differences between groups in academic performance outcomes at age eight.

Conclusion: The evidence on screening in primary care for DD in children aged 1 to 4 years of age without suspected DD to improve outcomes is inconclusive.

Background

The infancy-to-preschool period of child development between ages 0-5 years is widely recognized as a uniquely sensitive period for the foundation of cognitive ability and related functioning. Intensive change also typically occurs during this time across the domains of language, social and motor development. Intellectual disability and other developmental disabilities that often co-occur (e.g., autism spectrum disorders, or ASD) frequently entail lifelong challenges with respect to daily functioning and well-being for individuals and caregivers, and are considered to be detectable during the preschool period. Many caregivers, researchers and policymakers therefore argue that detection and intervention between the ages of 0-4 years is essential in order to optimize outcomes for children and families.[1-3]

Screening for children at risk of intellectual disability and related impairment is an important challenge for health care providers and policy-makers. Global developmental delay (DD) has been defined by the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) "as the failure of an individual under age 5 to meet expected developmental milestones across multiple areas of intellectual functioning."[4] Developmental delay may be understood as the failure to meet expected milestones across a given domain of development (e.g., cognitive, language, social or motor development). The DSM-5 emphasizes the difficulty in reliably assessing intellectual and related functioning among very young children during a period of intensive and variable developmental change.[4]

Existing guidelines and recommendations for screening children for DD vary. In 1994, the Canadian Task Force on the Periodic Health Examination 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. [5] In the same year, the Canadian Task Force on Preventive Health Care (CTFPHC) recommendation on Preschool Screening for Developmental Problems[6] found good evidence to recommend against the use of the Denver Developmental Screening Test (DDST)[7] in asymptomatic preschool children, as well as insufficient evidence for other screening tools. 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[8]; this guideline is currently being updated[9]. Conversely, the American Academy of Pediatrics (AAP) recommends screening for DD using a standardized tool at 9, 18 and 24 or 30 months of age [10] and screening for autism at 18 months and 24 months.[11] In Canada, Ontario has implemented an enhanced 18 month well-baby visit, which includes using the Nipissing District Developmental Screen (NDDS) as a surveillance tool to assess for global DD.[12, 13]

The systematic review on which this paper is based provided evidence for the CTFPHC to inform recommendations on screening for DD in children aged 1 to 4 years, who are not suspected of having DD or who are at risk, in a primary care setting. This systematic review synthesizes the effectiveness and harms of screening for DD in children with respect to improving cognitive, academic and functional outcomes.

Methods

We conducted a systematic literature search to address the effectiveness of screening for DD in children 1 to 4 years who were not suspected of having DD or at risk. For full details see PROSPERO CRD42014009809. Similar methods have been used by and are reported in other publications authored by our review team.[14-16]

Search Strategy

We searched Medline, Embase and PsychINFO with no beginning date limitations through February 24th, 2014. The published results of studies had to be available in either English or French. The effectiveness of the screening search was peer-reviewed using the PRESS format.[17]

Study Selection, Quality Assessment and Data Abstraction

Titles and abstracts of papers were reviewed in duplicate; any article marked for inclusion by either team member went on to full-text rating, which was performed independently by two people. All disagreements were resolved through discussions and consensus. The population of interest was children aged 1-4 years of age who were not 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). Screening with any tests, 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.[6]Settings were limited to primary care settings and public health clinics. Studies conducted in school settings were not included. To answer the question about the effectiveness of screening, only 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.

To answer the question of effectiveness of screening, 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[18] including anxiety; depression; oppositional defiant disorder (ODD); obsessive-compulsive disorder (OCD); overall quality of life; survival; and functionality as an adult (including employment; criminality; and independence). To answer the question on harms of screening outcomes included parental anxiety and stigma (labeling). There was no minimum follow-up time necessary for inclusion in our evidence summary.

One team member completed full data abstraction in a web-based systematic review software program[19] and a second team member verified this extraction; disagreements were resolved through discussion and/or third party consultation. All studies included to answer the effectiveness of screening question were assessed using the Cochrane Risk of Bias tool.[20] The strength of the evidence was determined based on the GRADE system of rating the quality of evidence using GRADEPro software.[21, 22]

A meta-analysis could not be performed due to a paucity of studies reporting on effectiveness of DD screening. For the effectiveness of DD screening that showed a 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 (GRADE estimates the absolute number per million using the control group event rate and risk ratio with the 95% confidence interval). For GRADE ratings see Tables 1 and 2.

Results

The search located 6,243 unique citations screened at title and abstract; 356 were screened at full text (Figure 1). We included two studies. The reference lists of sixteen identified systematic reviews were searched; no papers were added to our database as a result. Characteristics of included studies are provided in Table 3.

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. [23] This 2013 United States of America (US) study included 2,103 mixed-gender children who were randomly allocated to the office support group (mean age 10.5 months [SD 8.2]), no office support group (mean age 10.5 months [SD 8.1]) or usual care group (mean age 10.4 months[SD 8.6]). Those families allocated to the office support group met with trained office staff to complete the 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 of 8-10 questions from 4 domains) could also be referred to EI services. The screening arm with office support showed significantly more referrals to early intervention than the control group with a relative risk (RR) of 1.95 (95% CI 1.49, 2.54). The absolute risk increase was 9.67%. The number needed to screen for one child to be referred was 10 (95% CI 6, 20). The referral rates were also significantly more for the screening without office support group (RR 1.71; 95% CI 1.30, 2.25) as compared to the control group. The absolute risk increase was 7.24%. The number needed to screen for one child to be referred was 14 (95% CI 8, 33).

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, 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, 0.59]), both compared to the control group. 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.[24] This 2007 study, from the Netherlands, included 11,440 mixed gender children aged 15 months at study entry (mean age not reported). Intervention children were screened twice, once at 18 and once at 24 months using the VroegTijdige Onderkenning Ontwikkelingsstoornissen (VTO) Language Screening instrument and control children received usual monitoring. A final score ranging from 0 to 7 was assigned; 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. Assessment of academic performance at age eight showed no differences between groups with a relative risk (RR) of 0.71 (95% CI 0.48,1.04) of attending a special school; an RR of 0.99 (95% CI 0.81,1.21) of repeating a grade; an RR of 1.26 (95% CI 0.89,1.80) of repeating a grade because of language problems in regular primary school; an RR of 0.88 (95% CI 0.63,1.23) of being below the 10th percentile of oral tests; an RR of 1.00 (95% CI 0.72,1.40) of being below the 10th percentile of reading tests in grade 2; and an RR of 0.68 (95% CI 0.41,1.13) of being below the 10th percentile of spelling tests in grade 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.

Optimal Interval and Harms of Screening

We found no studies meeting our inclusion criteria that reported optimal intervals or harms of screening.

Discussion

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.[23] Referral rate is an intermediate outcome, therefore, conclusions about long-term outcomes related to screening and referral to early intervention programs cannot be drawn from this study. The second included study reported longer-term follow-up data (81 months) on academic performance outcomes in children screened at 15 months for speech and language delay. In this case, screening did not show a significant improvement in academic outcomes such as attending a special school; repeating a grade; repeating a grade because of language problems in regular primary school;

being below the 10th percentile of oral tests; being below the 10th percentile of reading tests in grade 2; or being below the 10th percentile of spelling tests in grade 2. Ideally, the intermediate outcome of early referral leads to early interventions which then improve long term outcomes. Unfortunately, our evidence does not confirm this. Furthermore, we found no evidence on which screening intervals are most effective and result in the least harm.

This review does not investigate treatment of DD and as such, we have no evidence on the effectiveness of early intervention programs: our first study investigates only to the point of referral and our second study does not indicate whether or not the children received an early intervention program between initial screen and 8 year assessment. The evidence on screening effectiveness is limited and inconclusive, but we cannot comment on the effectiveness of early intervention programs and their impact on cognitive, academic and adaptive functioning.

Currently, no guidelines exist on screening for global DD. The findings of this systematic review are in keeping with the most recent guideline (2006) from the USPSTF, which found insufficient evidence on screening for speech and language delay in children up to 5 years.[8] Despite the clear lack of evidence found in this review and the previous USPSTF review, screening of children is regularly implemented and endorsed. In the US, AAP recommends screening for DD at regular intervals up to 30 months.[10] In Canada, Ontario uses NDDS as a surveillance tool to monitor for DD in children at 18 month old visits. Despite common use, we found no peer-reviewed RCT evidence on the NDDS tool. In fact, a recent Canadian observational study evaluated the NDDS and found evidence that the tool should not be used on its own.[25] Further investigation into these commonly used tools is required in order to determine whether their continued use is clinically relevant and appropriate.

Limitations

The inclusion and exclusion criteria for this review limited our results. First, only publications in English and French were considered for inclusion. For the question of effectiveness of screening, only RCT data was included, thus excluding controlled clinical trials or observational studies that may have reported on our outcomes of interest. Though this limits the breadth of evidence available, it ensures a higher quality of evidence. For this review we selected high-level, long-term outcomes including cognitive, academic and adaptive functioning. This approach meant studies reporting on shorter-term, specific outcomes such as changes in expressive or receptive language or changes in social or motor functioning were excluded, as well as outcomes related to symptoms of ASD. Based on our exclusion, it is clear that there is research focused on these immediate outcomes, rather than the long-term outcomes this systematic review aimed to report on. Clearly, further trial research is needed to provide more conclusive results regarding the effectiveness of screening for DD in children 1 to 4 years old as they relate to improved cognitive functioning and related academic and adaptive functioning. Publication bias and methodological inconsistency could not be assessed due to lack of studies.

Conclusion

The evidence on screening for DD 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.

References

1. Messinger D, Young GS, Ozonoff S, Dobkins K, Carter A, Zwaigenbaum L, et al. Beyond autism: a baby siblings research consortium study of high-risk children at three years of age. J Am Acad Child Adolesc Psychiatry. 2013; 52(3):300-8. e1. Available from:

2. Campbell F, Conti G, Heckman JJ, Moon SH, Pinto R, Pungello E, et al. Early childhood investments substantially boost adult health. Science. 2014; 343(6178):1478-85. Available from:

3. Wallace KS, Rogers SJ. Intervening in infancy: implications for autism spectrum disorders. Journal of Child Psychology and Psychiatry. 2010; 51(12):1300-20. Available from:

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

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

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

7. Denver developmental materials, inc. Available from: http://denverii.com/denverii/.

8. 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/?term=16452370.

9. U.S. Preventive Services Task Force [Internet]. Screening for speech and language delay in preschool children, topic page. Rockville, MD: U.S. Preventive Services Task Force. Available from: http://www.uspreventiveservicestaskforce.org/uspstf/uspschdv.htm.

10. Council on Children With Disabilities, Section on Developmental Behavioral P, 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://pediatrics.uchicago.edu/chiefs/DBP/documents/reading%20pdf/Dev%20Screening.Council.pdf.

11. 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.

12. Ontario Children's Health Network, Ontario College of Family Physicians, Ontario Ministry of Children and Youth Services. 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.

13. 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.

14. Peirson L, Fitzpatrick-Lewis D, Morrison K, Ciliska D, Kenny M, Ali U, et al. Prevention of overweight/obesity in children and youth: a systematic review. CMAJ Open2014.

15. Peirson L, Douketis J, al. e. Prevention of overweight and obeisty in adult populations: a systematic review. CMAJ Open2014.

16. Peirson L, Douketis J, Ciliska D, Fitzpatrick-Lewis D, Ali MU, Raina P. Treatment for overweight and obesity in adult populations: a systematic reivew and meta-analysis. CMAJ Open2014.

17. Sampson M, McGowan J, Lefebvre C. An evidence based checklist for the peer review of Electronic Search Strategies (PRESS EBC). Evidence-Based Librarianship and Information Practice. 2010; 5(1):149-54. Available from: http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/7402/6436.

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

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

20. Cochrane handbook for systematic reviews of interventions. J. P. T. Higgins, S. Green, editors. New York, NY: John Wiley & Sons, Ltd. Publications; 2011. www.cochrane-handbook.org.

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

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

23. 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.

24. 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://pediatrics.aappublications.org/content/120/6/1317.abstract.

25. Cairney J, Clinton J, Veldhuizen S, Rodrigue AL. Evaluation of the Revised Nipissing District Developmental Screening (NDDS) Tool for Use in General Population Samples of Infants and Children. p. 23.

Figure 1: Screening for Developmental Delay Search Results

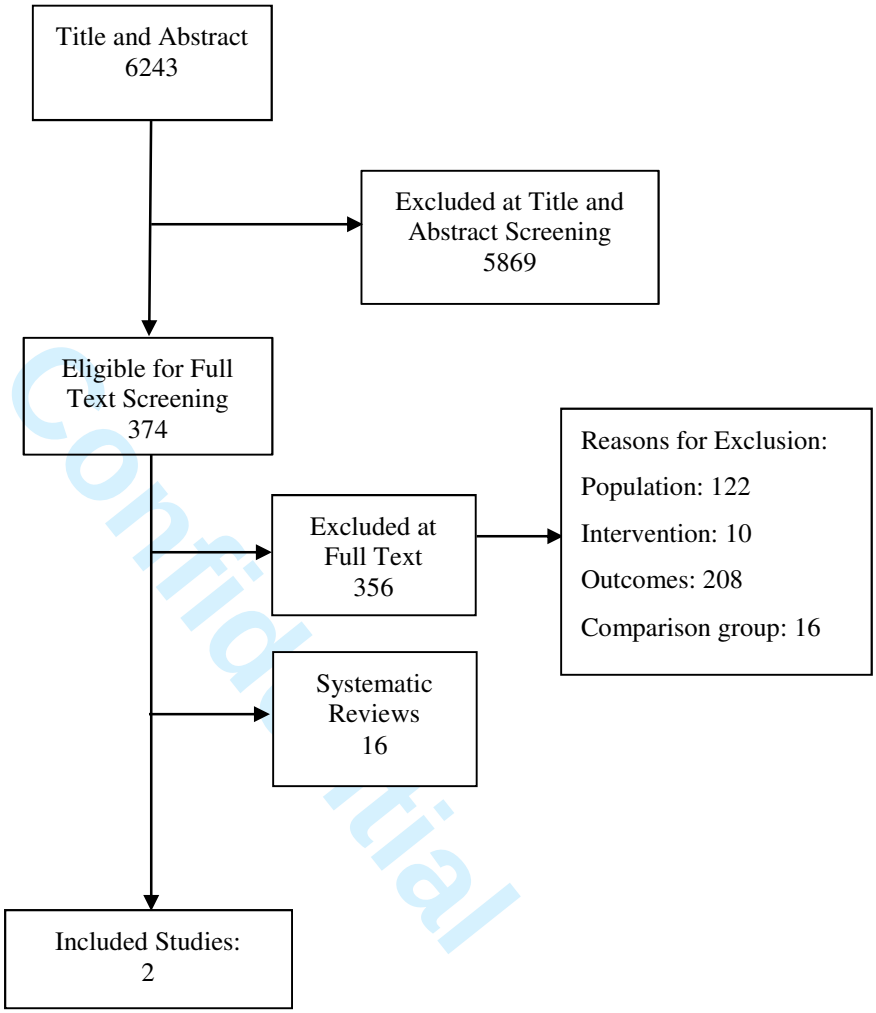

Table 1. GRADE Evidence Profile Table 1.1: Effect of Screening for Developmental Delay (ages 1 to 4 years old)

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)														

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

- Footnotes appear after the Summary of Findings Table

Table 2. 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	**0 per 1,000,000**	**0 per 1,000,000** (0 to 0)	**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	**0 per 1,000,000**	**0 per 1,000,000** (0 to 0)	**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 Follow-up: 81 months	**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]

[1] The single study is Guevera et al. 2013 [37]

[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, 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, 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, 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, 0.5853)]. This body of evidence was not downgraded for imprecision.

[10] This single study is van Agt et al. 2007.[38]

[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, 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, 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, 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, 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, 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, 1.1293)]. This body of evidence was downgraded for imprecision.

Table 3: Characteristics of Included Studies

Study Year Location	Risk of Bias	Participants	Intervention	Comparator	Length of follow-up	Exclusions
Guevera, 2013 US	Low	Sample: 2103 Intervention 1 n= 707; Intervention 2 n= 698; Control n= 698 Mean age (SD): Intervention 1= 10.5 (8.2); Intervention 2= 10.5 (8.1); Control= 10.4 (8.6) Gender [Female n(%)]: Intervention 1= 342 (48.4); Intervention 2= 354 (50.9); Control= 351 (50.4) Race/Ethnicity n (%): Intervention 1= 553 (78.2); Intervention 2= 521 (74.9); Control= 549 (78.9) Loss to follow-up: Intervention n= NR; Control n= NR	Caregivers completed Ages and Stages Questionnaire II at the child's 9, 18 and 30 month well child visit and the Modified Checklist for Autism in Toddlers at the 18 and 24 month visit	Caregivers completed the tools without the aid of standardized props either by mail before their visit or at the appointment check-in period	18 months	Inclusion: Children were eligible if they were <30 months old, >36 weeks' estimated gestational age, with no major congenital anomalies or genetic syndromes, not living in foster care and not currently receiving early intervention services
van Agt, 2007 Netherlands Companion paper: de Koning, 2004	Unclear	Sample: 55 clusters Intervention n= 28 clusters; 6,485 children; Control n= 27 clusters, 4,955 children Mean age (SD): not reported Gender [Female n(%)]: Overall: 50%; Intervention: 50.1%; Control: 49.9% Race/Ethnicity n (%): NR Loss to follow-up: I n= 1,161; C n=860	A structured screening instrument was conducted twice (at ages 15/18 months and 24 months) -the VTO Language Screening instrument consisted of a uniform set of questions for parents and test elements for the child	Usual care	Follow-up at age 8	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