Screening for developmental delay among children aged 1–4 years: a systematic review

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

Abstract

Background: Existing guidelines on screening children less than 5 years of age for developmental delay vary. In this systematic review, we synthesized the literature on the effectiveness and harms of screening for developmental delay in asymptomatic children aged 1–4 years.

Methods: We searched MEDLINE, Embase and PsychINFO for relevant articles published to June 16, 2015. We identified studies that included children aged 1–4 years who were not at high risk of developmental delay, screened in a primary care setting. Randomized trials and controlled cohort studies were considered for benefits (cognitive, academic and functional outcomes); no restrictions on study design were imposed for the review of harms.

Results: Two studies were included. One used the Ages and Stages Questionnaire II for screening and reported significantly more referrals to early intervention in the intervention groups than in the control group (relative risk [RR] 1.95, 95% confidence interval [CI] 1.49–2.54, in the intervention group with office support and RR 1.71, 95% CI 1.30–2.25, in the intervention group without office support). The time to referral was 70% shorter in the intervention group with office support (rate ratio 0.30, 95% CI 0.19–0.48) and 64% shorter in the intervention group without office support (rate ratio 0.36, 95% CI 0.23–0.59), compared with the control group. The other study used the VroegTijdige Onderkenning Ontwikkelingsstoornissen Language Screening instrument to screen children aged 15 months at enrolment for language delay. It reported no differences between groups in academic performance outcomes at age 8 years.

Interpretation: The evidence on screening for developmental delay in asymptomatic children aged 1–4 years is inconclusive. Further research with longer-term outcomes is needed to inform decisions about screening and screening intervals.

The preschool period of child development (age < 5 yr) is widely recognized as a uniquely sensitive period for the foundation of cognitive ability and related functioning. Intensive change 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) 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 policy-makers therefore argue that detection and intervention before the age of 5 years is essential to optimize outcomes for children and families.[1–3]

Screening children at risk of intellectual disability and related impairment is an important challenge for health care providers and policy-makers. Global developmental delay has been defined in the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-V) 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-V emphasizes the difficulty in reliably assessing intellectual and related functioning among very young children during a period of intensive and variable developmental change.[4]

Within pediatric health care, policy decisions that address early detection and intervention for developmental delay involve a particularly complex balance of considerations. Parents of children with developmental disorders often report a sense of urgency to intervene as quickly as possible during essential developmental windows. There is also an important tension between intervening as early as possible and not over-identifying many children as developmentally delayed (i.e., too many false-positive findings). False-positive screen results for developmental delay can lead to unnecessary stigma, caregiver anxiety and excess cost, whereas false-negative results can lead to delays in treatment.

Existing guidelines and recommendations for screening children for developmental delay vary. In 1994, the Canadian Task Force on the Periodic Health Examination (now the Canadian Task Force on Preventive Health Care) found fair evidence to assess developmental milestones at each well-baby visit.[5] In another guideline in the same year,[6] the task force found good evidence to recommend against the use of the Denver Developmental Screening Test[7] in asymptomatic pre-school children, as well as insufficient evidence for other screening tools. In 2015, the US Preventive Services Task Force 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] Conversely, the American Academy of Pediatrics recommends screening for developmental delay using a standardized tool at 9, 18 and 24 or 30 months of age,[9] and screening for autism at 18 months and 24 months.[10] 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 developmental delay.[11,12]

More than 20 years have passed since the last Canadian Task Force on Preventive Health Care guideline on developmental screening, necessitating a new review of the empirical literature to determine the current state of the evidence. The systematic review on which this article is based provided evidence for the task force to inform new recommendations on screening for developmental delay in children aged 1–4 years. This systematic review synthesizes the effectiveness and harms of screening for developmental delay in children with respect to improving cognitive, academic and functional outcomes.

Methods

We conducted a systematic review to address the effectiveness of screening for developmental delay in children aged 1–4 years who are not at risk or suspected of having developmental delay. This search was part of a larger, staged systematic review that investigated treatment of developmental delay and test properties of screening tests (details about the full review protocol are available at www.crd.york.ac.uk/PROSPERO; registration no. CRD42014009809). Results from the full review will be available on the Canadian Task Force on Preventive Health Care website (http://canadiantaskforce.ca/). Similar methods have

been used by and are reported in other publications authored by our review team.[13–15]

Key questions

This systematic review aims to answer the following questions: What is the effectiveness of screening children aged 1–4 years without suspected developmental delay to improve their outcomes? What is the optimal interval for screening for developmental delay? What is the incidence of harms of screening children aged 1–4 years without suspected developmental delay?

Search strategy

We searched MEDLINE, Embase and PsychINFO with no beginning date limitations through to June 16, 2015. 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.[16]

Eligibility criteria

Details about the inclusion and exclusion criteria are shown in Table 1. Because developmental delay poses a high burden at individual, family and societal levels, rigorous methodologic standards are essential to assess whether developmental screening tools are effective in meaningful ways. For this reason, we included only controlled cohort studies and randomized controlled trials (RCTs) to answer the question of benefits.

Study selection, quality assessment and data abstraction

Six team members, working in pairs, reviewed titles and abstracts of papers in duplicate; articles marked for inclusion by either member went on to full-text screening, which was performed independently by the same 6 team members. Disagreements were resolved through discussion and consensus.

One team member completed full data abstraction using a Web-based systematic review software program (DistillerSR, Evidence Partners), and a second team member verified this extraction; disagreements were resolved through discussion or by a third team member. All studies included to answer the question about effectiveness of screening were assessed by one team member using the Cochrane risk-of-bias tool[18] and verified by another team member. The quality of the evidence was determined based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methods (www.gradeworkinggroup.org) using GRADEPro software (version 3.2 for Windows, available at http://ims.cochrane.org/revman/other-resources/gradepro/download).

Statistical analysis

We could not perform a meta-analysis because of the paucity of studies. The effect estimates from the studies were reported in the form of relative risk (RR) for binary outcomes of interest such as referral rate and academic performance, and rate ratio for time to event outcomes such as time to referral. For studies that showed a significant effect for the effectiveness of screening for developmental delay, we added the estimates of absolute risk reduction, absolute risk increase and number

needed to screen. The numbers needed to screen were estimated using the absolute numbers computed by GRADEpro, which are calculated using control group event rate and relative risk with the 95% confidence interval (CI).

Results

Of the 7782 unique citations identified, 393 were screened in full and 2 studies, both RCTs, were included in our review (Figure 1). The reference lists of 17 identified systematic reviews were searched; no citations were added to our database as a result. Characteristics of the included studies are provided in Table 2.

Effectiveness of screening
Data on effect sizes reported by the studies are shown in Tables 3 and 4, as are our GRADE ratings.

Referral outcomes
One RCT provided evidence on referral rates and time to referral for children less than 30 months of age who were screened for developmental delay using the Ages and Stages Questionnaire II (ASQ-II).[19] This 2013 American study included 2103 boys and girls who were randomly allocated to 1 of 2 intervention groups (screening with office support [group A] or screening without office support [group B]) or usual care. The mean ages were 10.5 (standard devision [SD] 8.2) months in intervention group A, 10.5 (SD 8.1) months in intervention group B and 10.4 (SD 8.6) months in the control

group. Families allocated to intervention group A met with trained office staff to complete the screening tool with the use of props; families in intervention group B completed the ASQ-II without support of office staff or the use of props. Children were considered screen positive if they scored 2 SDs below the mean for age on any of the 5 developmental domains, and they could be referred to early intervention services. Children in the control group were considered screen positive if they failed the usual-care developmental screen (milestones of 8–10 questions from 4 domains), and they could also be referred to early intervention services.

There were significantly more referrals to early intervention in intervention group A than in the control group (RR 1.95, 95% CI 1.49–2.54) (Tables 3 and 4). The absolute risk increase was 9.67%, and the number needed to screen for one child to be referred was 10 (95% CI 6–20). The referral rate was also significantly higher in intervention group B than in the control group (RR 1.71, 95% CI 1.30–2.25). The absolute risk increase was 7.24%, and the number needed to screen for one child to be referred was 14 (95% CI 8–33). Compared with the time to referral in the control group, the time was 70% shorter in intervention group A (rate ratio 0.30, 95% CI 0.19–0.48) and 64% shorter in intervention group B (rate ratio 0.36, 95% CI 0.23–0.59).

The GRADE ranking for time to referral and referral rates (for screening with or without office support) was moderate (Tables 3 and 4). Evidence from this study was downgraded for indirectness because participant age at enrolment was less than 12 months.

Table 1: Inclusion and exclusion criteria		
Variable	Inclusion	Exclusion
Population	Children aged 1–4 yr who were not at high risk of developmental delay or suspected of having developmental delay	Children born preterm (gestational age < 37 completed wk at birth) or with low birth weight (< 2500 g); children with other known disorders that may be associated with or affect development
Interventions	Screening with any test, tool or questionnaire used to screen for developmental delay, including tools for specific domains, tools for general developmental delay and tools for autistic disorder and autism spectrum disorders	Denver Developmental Screening Test[7]
Comparators	*Effectiveness of screening:* No screening comparison *Harms of screening:* No comparison group necessary	
Outcomes	*Effectiveness of screening:* 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, including anxiety, depression, oppositional defiant disorder and obsessive–compulsive disorder; overall quality of life; survival; and functionality as an adult (including employment, criminality and independence) *Harms of screening:* Parental anxiety and stigma (labelling)	
Study designs	*Effectiveness of screening:* RCTs and controlled cohort studies *Harms of screening:* Any study design	
Settings	Primary care settings and public health clinics	School settings
Language	English or French	
Note: DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, 4th edition,[17] RCT = randomized controlled trial.		

Academic performance

The other RCT provided data on academic performance in children screened for language delay.[20] This 2007 study from the Netherlands included 11 440 boys and girls aged 15 months at enrolment (mean age not reported). Children randomly assigned to the intervention group were screened twice, at 18 and 24 months, with the VroegTijdige Onderkenning Ontwikkelingsstoornissen (VTO) Language Screening instrument; children assigned to the control group received usual monitoring. A final score ranging from 0 to 7 was assigned; children with a total score of 2 or less were referred for additional assessment to confirm language delay. After screening, the study did not offer an intervention and did not indicate whether children received interventions elsewhere.

Assessment of academic performance at age 8 years showed no differences between the study groups in any of the outcomes measured (attendance of a special school: RR 0.71, 95% CI 0.48–1.04; repeating a grade: RR 0.99, 95% CI 0.81–1.21; repeating a grade because of language problems in regular primary school: RR 1.26, 95% CI 0.89–1.80; being below the 10th percentile of oral tests: RR 0.88, 95% CI 0.63–1.23; being below the 10th percentile of reading tests in grade 2: RR 1.00, 95% CI 0.72–1.40; or being below the 10th percentile of spelling tests in grade 2: RR 0.68, 95% CI 0.41–1.13) (Tables 3 and 4).

The GRADE ranking for all outcomes of academic performance was low (Tables 3 and 4). Evidence from this study was downgraded for potential risk of bias because of insufficient
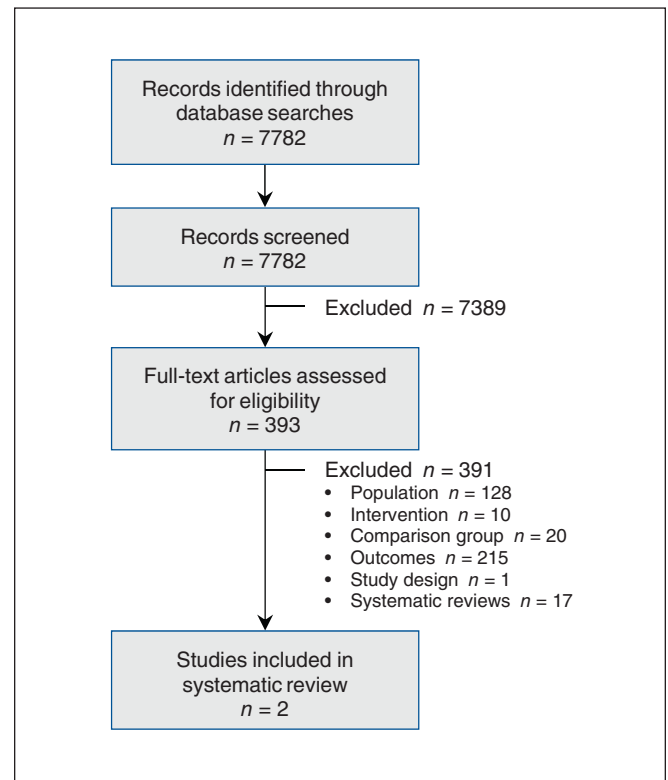

Figure 1: Selection of studies for the systematic review.

Table 2: Characteristics of included studies						
Study	Risk of bias	Participants	Intervention	Control	Length of follow-up	Eligibility criteria
Guevera et al.[18]	Low	**Sample: 2103** Intervention A: 707 Intervention B: 698 Control: 698 **Age, mo, mean ± SD** Intervention A: 10.5 ± 8.2 Intervention B: 10.5 ± 8.1 Control: 10.4 ± 8.6 **Female sex, no. (%)** Intervention A: 342 (48.4) Intervention B: 354 (50.9) Control: 351 (50.4) **Race/ethnicity data available, no. (%)** Intervention A: 553 (78.2) Intervention B: 521 (74.9) Control: 549 (78.9) **Loss to follow-up: NR**	Caregivers completed the ASQ-II at the 9-, 18- and 30-month well-child visits, and the Modified Checklist for Autism in Toddlers at the 18- and 24-month visits	Caregivers completed the screening tools without the aid of standardized props, either by mail before the well-child visits or at appointment check-in	18 mo	Age < 30 mo; estimated gestational age > 36 wk; no major congenital anomalies or genetic syndromes; not living in foster care and not currently receiving early intervention services
van Agt et al.[19]	Unclear	**Sample: 55 clusters** Intervention: 28 clusters (6485 children) Control: 27 clusters (4955 children) **Mean age: NR** **Female sex, %** Overall: 50% Intervention: 50.1% Control: 49.9% **Race/ethnicity: NR** **Loss to follow-up** Intervention: 1161 Control: 860	Structured screening tool (VTO) was completed twice, at 15–18 mo and at 24 mo; the VTO tool consists of a uniform set of questions for parents and test elements for the child	Usual care	81 mo	Age 15–24 mo during given inclusion period; living in area of intervention physician's or control physician's health care location

Note: ASQ-II = Ages and Stages Questionnaire II, NR = not reported, SD = standard devision, VTO = VroegTijdige Onderkenning Ontwikkelingsstoornissen Language Screening instrument.

Table 3: GRADE evidence profile for the effect of screening for developmental delay in children aged 1–4 years[18,19]

Measure	Quality assessment							No. (%) of participants		Relative risk (95% CI)	Effect			Quality	Importance
	No. of studies	Design	Serious risk of bias	Serious inconsistency	Serious indirectness	Serious imprecision	Other bias	Treatment	Control		Absolute effect per million	ARR or ARI	NNS (95% CI)		
Referral rates to intervention (screening tool ASQ-II), 18-mo follow-up															
Screening with office support	1	RCT[18]	No[a]	No[b]	Yes[c]	No[d]	No[e]	140/704 (19.9)	71/695 (10.2)	1.95 (1.49–2.54)	96 703 more (range 50 313 more to 157 211 more)	9.67%	10 (6–20)	Moderate	Critical
Screening without office support	1	RCT[18]	No[a]	No[b]	Yes[c]	No[f]	No[e]	121/693 (17.5)	71/695 (10.2)	1.71 (1.30–2.25)	72 440 more (range 30 668 more to 127 361 more)	7.24%	14 (8–33)	Moderate	Critical
Time to referral (screening tool ASQ-II), 18-mo follow-up															
Screening with office support	1	RCT[18]	No[a]	No[b]	Yes[c]	No[g]	No[c]	–/704	–/695	0.30 (0.19–0.48)	–	–	–	Moderate	Critical
Screening without office support	1	RCT[18]	No[a]	No[b]	Yes[c]	No[h]	No[e]	–/693	–/695	0.36 (0.23–0.59)	–	–	–	Moderate	Critical
Academic performance, by outcome measures (VTO screening), 81-mo follow-up															
Attendance at special school	1	RCT[19]	Yes[i]	No[b]	No[j]	Yes[k]	No[e]	83/3118 (2.7)	85/2288 (3.7)	0.71 (0.48–1.04)	10 762 fewer (range 19 144 fewer to 1 523 more)	–	–	Low	Critical
Grade repeated	1	RCT[19]	Yes[i]	No[b]	No[j]	Yes[l]	No[e]	443/3084 (14.4)	318/2250 (14.1)	0.99 (0.81–1.21)	1 413 fewer (range 26 754 fewer to 29 553 more)	–	–	Low	Critical
Grade repeated (language problems)	1	RCT[19]	Yes[i]	No[b]	No[j]	Yes[m]	No[e]	146/2401 (6.1)	84/1721 (4.9)	1.26 (0.89–1.80)	12 807 more (range 5 511 fewer to 38 871 more)	–	–	Low	Critical
< 10th percentile of oral tests	1	RCT[19]	Yes[i]	No[b]	No[j]	Yes[n]	No[e]	112/1270 (8.8)	90/925 (9.7)	0.88 (0.63–1.23)	11 685 fewer (range 36 068 fewer to 22 398 more)	–	–	Low	Critical
< 10th percentile of reading tests	1	RCT[19]	Yes[i]	No[b]	No[j]	Yes[o]	No[e]	86/1844 (4.7)	62/1328 (4.7)	1.00 (0.72–1.40)	0 fewer (range 13 231 fewer to 18 460 more)	–	–	Low	Critical
< 10th percentile of spelling tests	1	RCT[19]	Yes[i]	No[b]	No[j]	Yes[p]	No[e]	48/1728 (2.8)	52/1225 (4.2)	0.68 (0.41–1.13)	13 592 fewer (range 25 079 fewer to 5 489 more)	–	–	Low	Critical

Note: ARI = absolute risk increase, ARR = absolute risk reduction, ASQ-II = Ages and Stages Questionnaire II, CI = confidence interval, GRADE = Grading of Recommendations Assessment, Development and Evaluation, NNS = number needed to screen, RCT = randomized controlled trial, VTO = VroegTijdige Onderkenning Ontwikkelingsstoornissen Language Screening instrument.
See Table 4 for footnotes.

Table 4: GRADE summary of findings for effect of screening for developmental delay in children 1–4 years old

Outcome	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No. of participants (studies)	GRADE quality of evidence
	Assumed risk in control group, no. per million	Corresponding risk with treatment, no. per million			
Referral rates to intervention (ASQ-II screening tool), 18-mo follow-up					
Screening with office support	102 158	198 861 (152 471–259 370)	RR 1.95 (1.49–2.54)	1399 (1 study[18])	Moderate[a,b,c,d,e]
Screening without office support	102 158	174 599 (132 826–229 519)	RR 1.71 (1.30–2.25)	1388 (1 study[18])	Moderate[a,b,c,e,f]
Time to intervention referral (ASQ-II screening tool), 18-mo follow-up					
Screening with office support	0	0 (0 to 0)	RR 0.30 (0.19–0.48)	1399 (1 study[18])	Moderate[a,b,c,e,g]
Screening without office support	0	0 (0 to 0)	RR 0.36 (0.23–0.59)	1388 (1 study[18])	Moderate[a,b,c,e,h]
Academic performance, by outcome measures (VTO screening tool), 81-mo follow-up					
Attendance of special school	37 150	26 388 (18 007–38 674)	RR 0.71 (0.48–1.04)	5406 (1 study[19])	Low[b,e,i,j,k]
Repeating a grade	141 333	139 920 (114 579–170 886)	RR 0.99 (0.81–1.21)	5334 (1 study[19])	Low[b,e,i,j,l]
Repeating a grade (language problems)	48 809	61 616 (43 298–87 680)	RR 1.26 (0.89–1.80)	4122 (1 study[19])	Low[b,e,i,j,m]
< 10th percentile of oral test	97 297	85 612 (61 229–119 695)	RR 0.88 (0.63–1.23)	2195 (1 study[19])	Low[b,e,i,j,n]
< 10th percentile of reading test	46 687	46 687 (33 456–65 147)	RR 1.00 (0.72–1.40)	3172 (1 study[19])	Low[b,e,i,j,o]
< 10th percentile of spelling test	42 449	28 857 (17 370–47 938)	RR 0.68 (0.41–1.13)	2953 (1 study[19])	Low[b,e,i,j,p]

Note: ASQ-II = Ages and Stages Questionnaire II, CI = confidence interval, GRADE = Grading of Recommendations Assessment, Development and Evaluation, RR = relative risk, VTO = VroegTijdige Onderkenning Ontwikkelingsstoornissen Language Screening instrument.

*The assumed risk in the control group is based on event rates in control groups across studies. The corresponding risk (and its 95% CI) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

aUsing 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. Because 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, the evidence was not downgraded for serious study limitations.

bBecause this was a single study, inconsistency could not be assessed.

cThis study included boys and girls < 12 months of age with an average risk of developmental delay (mean ages: intervention group A, 10.5 ± standard deviation 8.2 months; intervention group B, 10.5 ± 8.1 months; control group, 10.4 ± 8.6 months). The intervention groups were screened using ASQ-II (one group with office support [group A], one group without [group B]); the control group received usual care. The study took place in a primary care setting in the United States, and the report was published 2013. The evidence was downgraded because the population was not restricted to children aged 1–4 years.

dAdequate sample size (704 in intervention group A, 695 in control group) and sufficient number of events (140 in intervention group A, 71 in control group). The pooled effect estimate is precise, with a narrow 95% CI (RR 1.95, 95% CI 1.49–2.54). The evidence was not downgraded for imprecision.

eThere were too few studies (< 10) to assess publication bias.

fAdequate sample size (693 in intervention group B, 695 in control group) and sufficient number of events (121 in intervention group B, 71 in control group). The pooled effect estimate is precise, with a narrow 95% CI (RR 1.71, 95% CI 1.30–2.25). The evidence was not downgraded for imprecision.

gAdequate sample size (704 in intervention group A, 695 in control group). The pooled effect estimate is precise, with a narrow 95% CI (RR 0.30, 95% CI 0.19–0.48). The evidence was not downgraded for imprecision.

hAdequate sample size (693 in intervention group B, 695 in control group). The pooled effect estimate is precise, with a narrow 95% CI (RR 0.36, 95% CI 0.23–0.59). The evidence was not downgraded for imprecision.

iUsing 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 of bias 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, the evidence was downgraded for serious study limitations.

jThis study included boys and girls aged 15 months at study entry with an average risk of developmental delay (mean ages not reported). The intervention group was screened using the VTO instrument; the control group received usual care. The study took place in a primary care setting in the Netherlands, and the report was published in 2007. There were no serious concerns regarding directness of this evidence.

kAdequate sample size (3118 in intervention group, 2288 in control group), but fairly low number of events (83 in intervention group, 85 in control group). The pooled effect estimate is not precise, with a 95% CI that includes the no-effect value (RR 0.71, 95% CI 0.48–1.04). The evidence was downgraded for imprecision.

lAdequate sample size (3084 in intervention group, 2250 in control group) and sufficient number of events (443 in intervention group, 318 in control group). The pooled effect estimate is not precise, however, with a 95% CI that includes the no-effect value (RR 0.99, 95% CI 0.81–1.21). The evidence was downgraded for imprecision.

mAdequate sample size (2401 in intervention group, 1721 in control group) and sufficient number of events (146 in intervention group, 84 in control group). The pooled effect estimate is not precise, however, with a 95% CI that includes the no-effect value (RR 1.26, 95% CI 0.89–1.80). The evidence was downgraded for imprecision.

nAdequate sample size (1270 in intervention group, 925 in control group) and sufficient number of events (112 in intervention group, 90 in control group). The pooled effect estimate is not precise, however, with a 95% CI that includes the no-effect value (RR 0.88, 95% CI 0.63–1.23). The evidence was downgraded for imprecision.

oAdequate sample size (1844 in intervention group, 1328 in control group), but fairly low number of events (86 in intervention group, 62 in control group). The pooled effect estimate is not precise, with a 95% CI that includes the no-effect value (RR 1.00, 95% CI 0.72–1.40). The evidence was downgraded for imprecision.

pAdequate sample size (1728 in intervention group, 1225 in control group), but low number of events (48 in intervention group, 52 in control group). The pooled effect estimate is not precise, with a 95% CI that includes the no-effect value (RR 0.68, 95% CI 0.41–1.13). The evidence was downgraded for imprecision.

information on allocation concealment and blinding of participants, and for imprecision because effect estimates included the null value.

Optimal interval and harms of screening

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

Interpretation

The evidence on the effectiveness of screening for developmental delay 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 developmental delay than among those in the control group.[19] 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 study reported longer-term follow-up data (81 mo) on academic performance outcomes among children screened at 18 and 24 months for speech and language delay. In this study, screening did not show a significant improvement in academic outcomes at age 8 years. 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 which result in the least harm.

We did not investigate treatment of developmental delay. As such, we have no evidence on the effectiveness of early intervention programs: the first study investigated only to the point of referral, and the second one did not indicate whether the children received an early intervention program between the initial screen and the 8-year assessment. The evidence on the effectiveness of screening is limited and inconclusive; as such, 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 developmental delay. The findings of this systematic review are in keeping with the most recent guideline (2015) from the US Preventive Services Task Force, which found insufficient evidence on screening for speech and language delay in children up to 5 years.[8]

Despite the lack of evidence found in this review and in the US Preventive Services Task Force's systematic review on screening for speech and language in children less than 5 years old,[21] screening of children is regularly implemented and endorsed. In the US, the American Academy of Pediatrics recommends screening for developmental delay at regular intervals up to 30 months.[9] In Canada, Ontario uses the NDDS as a surveillance tool to monitor for developmental delay in children at the 18-month well-baby visit. Despite common use, we found no peer-reviewed RCT evidence on the NDDS tool. A recent Canadian observational study evaluated the NDDS and found evidence that the tool should not be used on its own (unpublished data). Further investigation into these commonly used tools is required 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 or French were considered for inclusion. Second, for the question of effectiveness of screening, only RCTs and controlled cohort studies were included, thus excluding observational studies that may have reported on our outcomes of interest. Although this limited the breadth of evidence available, it ensured a higher quality of evidence. Third, because we selected high-level, long-term outcomes (e.g., cognitive, academic and adaptive functioning), studies reporting on shorter-term, specific outcomes (e.g., changes in expressive or receptive language or changes in social or motor functioning) and on outcomes related to symptoms of autism spectrum disorders were excluded. Based on our exclusion criteria, it is clear that there is research focused on these immediate outcomes, rather than on the long-term outcomes our review aimed to report on. Publication bias and methodologic inconsistency could not be assessed due to the paucity of studies.

Conclusion

The evidence on screening for developmental delay in asymptomatic children aged 1–4 years 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, 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;52:300-8.e1.
2. Campbell F, Conti G, Heckman JJ, et al. Early childhood investments substantially boost adult health. *Science* 2014;343:1478-85.
3. Wallace KS, Rogers SJ. Intervening in infancy: implications for autism spectrum disorders. *J Child Psychol Psychiatry* 2010;51:1300-20.
4. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders. Fifth edition*. Arlington (VA): American Psychiatric Publishing; 2013.
5. Feldman W. Well-baby care in the first 2 years of life. In: *The Canadian Guide to Clinical Preventive Health Care*. Ottawa: Canadian Task Force on the Periodic Health Examination; 1994:258-66.
6. Feightner JW. Preschool screening for developmental problems. In: *The Canadian Guide to Clinical Preventive Health Care*. Ottawa: Canadian Task Force on the Periodic Health Examination; 1994:289-96.
7. *Denver Developmental Screening Test*. Denver: Denver Developmental Materials. Available: http://denverii.com/ (accessed 2016 Jan. 12).
8. Siu AL; 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:e474-81.
9. 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;118:405-20.
10. Johnson CP, Myers SM. Identification and evaluation of children with autism spectrum disorders. *Pediatrics* 2007;120:1183-215.
11. *Getting it right at 18 months ... Making it right for a lifetime. Report of the Expert Panel on the 18 month well baby visit*. Toronto: Ontario Children's Health Network; Ontario College of Family Physicians; 2005. Available: www.children.gov.on.ca/htdocs/English/topics/earlychildhood/gettingitright.aspx (accessed 2016 Jan. 12).
12. NDDS Canada. Nipissing district development screen. North Bay (ON): Nipissing District Developmental Screen Intellectual Property Association; 2011. Available: www.ndds.ca/canada.html (accessed 2016 Jan. 12).

13. Peirson L, Fitzpatrick-Lewis D, Morrison K, et al. Prevention of overweight/ obesity in children and youth: a systematic review. *CMAJ Open* 2015;3:E23-33.
14. Peirson L, Douketis J, Ciliska D, et al. Prevention of overweight and obeisty in adult populations: a systematic review. *CMAJ Open* 2014;2:E268-72.
15. Peirson L, Douketis J, Ciliska D, et al. Treatment for overweight and obesity in adult populations: a systematic reivew and meta-analysis. *CMAJ Open* 2014; 2:E306-17.
16. McGowan J, Sampson M, Lefebvre C. An evidence based checklist for the peer review of electronic search strategies (PRESS EBC). *Evid Based Libr Inf Pract* 2010;5:149-54.
17. *Diagnostic and statistical manual of mental disorders*. 4th ed. Washington (DC): American Psychiatric Association; 1994.
18. Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Version 5.1.0. New York: John Wiley & Sons; 2011. Available: www. cochrane-handbook.org (accessed 2016 Jan. 12).
19. Guevara JP, Gerdes M, Localio R, et al. Effectiveness of developmental screening in an urban setting. *Pediatrics* 2013;131:30-7.
20. van Agt HME, van der Stege HA, de Ridder-Sluiter H, et al. A cluster-randomized trial of screening for language delay in toddlers: effects on school performance and language development at age 8. *Pediatrics* 2007;120:1317-25.
21. Berkman ND, Wallace I, Watson L, et al. Screening for speech and language delays and disorders in children age 5 years or younger: a systematic review for the U.S. Preventive Services Task Force. [Report no. 13-05197-EF-1]. Rockville (MD): Agency for Healthcare Research and Quality; 2015. Available: www.ncbi.nlm.nih.gov/pubmed/26225412 (accessed 2016 Jan. 12).

Affiliations: Evidence Review and Synthesis Centre, and Offord Centre for Child Studies, McMaster University, Hamilton, Ont.