Appendix 5 (as supplied by the authors): GRADE evidence profile: effect of screening for developmental delay (ages 1 to 4 years old) — RCTs*

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 (19.89%)	71/695 (10.22%)	RR 1.95 (1.50 to 2.54)	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 (17.46%)	71/695 (10.22%)	RR 1.7091 (1.30 to 2.25)	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 0.30 (0.19 to 0.48)	-	-	-	⊕⊕⊕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 0.35 (0.23 to 0.59)	-	-	-	⊕⊕⊕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 (2.66%)	85/2,288 (3.72%)	RR 0.71 (0.48 to 1.04)	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 (14.36%)	318/2,250 (14.13%)	RR 0.99 (0.81 to 1.21)	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 (6.08%)	84/1,721 (4.88%)	RR 1.26 (0.89 to 1.80)	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 (8.82%)	90/925 (9.73%)	RR 0.88 (0.63 to 1.23)	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 (4.66%)	62/1,328 (4.66%)	RR 1.00 (0.75to 1.40)	0 fewer (from 13,231 fewer to 18,460	-	-	⊕⊕OO LOW	CRITICAL

Appendix to: Canadian Task Force on Preventive Health Care. Recommendations on screening for developmental delay. *CMAJ* 2016. DOI:10.1503/cmaj.151437.

										more)				
Academic performance - By outcome measures (VTO screening) - Below 10 percentile of spelling test (follow-up 81 months)														
1	randomized trial[10]	seriousrisk[11]	no serious inconsistency[3]	no serious indirectness[12]	serious[18]	none[6]	48/1,728 (2.78%)	52/1,225 (4.25%)	RR 0.68 (0.41 to 1.13)	13,592 fewer (from 25,079 fewer to 5,489 more)	-	-	⊕⊕OO LOW	CRITICAL

*Reference: Warren R, Kenny, M, Fitzpatrick-Lewis D, et al. Screening and treatment for developmental delay in early childhood (ages 1-4 years): a systematic review. Calgary: Canadian Task Force on Preventive Health Care; 2016. Available: canadiantaskforce.ca/ctfphcguidelines/2015-developmental-delay/systematic-review/ (accessed 2016 Mar. 29).

[1] The single study is Guevara JP, Gerdes M, Localio R, et al. Effectiveness of developmental screening in an urban setting. *Pediatrics* 2013;131:30-7.
[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 HM, 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.
[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 0.49, 1.04)]. 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 0.99 (95% CI 0.81, 1.21)]. 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 1.26 (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 0.63, 1.23)]. 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 0.72, 1.40)]. 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 0.68 (95% CI 0.41, 1.13)]. This body of evidence was downgraded for imprecision.