Appendix 3 (as supplied by the authors): GRADE basis of recommendation decision table for screening for developmental delay

Questions: 1. What is the effectiveness of screening children without suspected developmental delay to improve outcomes?		
Populations: 1. Children aged 1 to 4 years not at high risk (screening); 2. Children starting intervention for developmental delay between the ages of 1 to 6		
Interventions: 1. Any tests, tool, or questionnaire used to screen for developmental delay; including tools for specific domains and tools for general developmental delay 2. Any intervention for developmental delay (including domain specific delays, and DD associated with Autism Spectrum Disorders) using behavioural, pharmacological, or psychological interventions		
Setting (if relevant): Primary care		
Decision domain	Summary of reason for decision	Subdomains influencing decision
Quality of evidence (QoE) for screening studies *Is there high or moderate quality of evidence* Yes☒ No ☐	**QoE for benefits of screening:** Moderate and low quality evidence but only on intermediate outcomes. No RCT evidence was found demonstrating that screening for developmental delay in children aged 1 to 4 improves long term outcomes (cognitive function, quality of life, mental health, survival, or functional status as an adult). A moderate quality[1] RCT (Guevara 2013[2]) found that screening increased the likelihood of referral to early intervention, reduced time to referral, and increased the number of children completing a multidisciplinary evaluation. A low quality RCT (van Agt 2007[3]) found that there were no differences in educational attainment between the groups (based on intention-to-screen analysis). Screening tests were found to have	Key reasons for downgrading or upgrading: **QoE for benefits of screening:** No evidence on long term clinical outcomes. Guevara 2013 was downgraded because the population was not restricted to children aged 1-4 years. van Agt 2007 was downgraded for the following reasons: 1. The study was rated as having unclear risk of bias and because all of the information for the outcome on educational attainment comes from this study, the body of evidence was downgraded for serious study limitations. 2. Although the sample size is adequate (3,118 intervention arm, 2,288 control arm) 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

	inconsistent accuracy and their low specificity would lead to a high rate of false positive tests: One study reported sensitivity and specificity ASQ of 82% and 78% (22% false positive rate) and for the PEDS of 74% and 64% (36% false positive rate). [4] A second study evaluated the ASQ aged 18 to 42 months reported sensitivity of 62% and specificity of 84% (16% false positive rate).[5] The false positive rate was 16 to 22% for the ASQ and 36% for the PEDS. [4,5] A third study that compared NDDS to the BSID-III for children between 1 month and 3 years reported sensitivity and specificity ranging from 29% to 65% and from 63% to 88%, respectively, depending on the age of the child and the cutpoint used to define an abnormal test.[6]	effect value [RR 0.71 (95% CI 0.48, 1.04)]. The evidence on the effectiveness of screening tests (accuracy) was not evaluated using the GRADE methods, therefore, the evidence was not downgraded or upgraded using the GRADE criteria. However, the following limitations were identified: 1. The review for test properties was limited to papers which had reported data; data on the components of test properties was not calculated. 2. 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. **QoE for harms of screening:** No direct evidence on the harms of screening.
(QoE) for treatment studies *Is there high or moderate quality of evidence* Yes☒ No ☐	**QoE for benefits of treatment in children aged 1 to 6:** Moderate and low quality evidence, but only for three critical outcomes. Moderate quality evidence from 3 RCTs[1, 7-9] for language impairment; low quality evidence from 1 RCT[7] for adaptive functioning. The 5 systematic reviews examining the effectiveness of treatment ASD were not quality assessed. No studies were found that assessed academic performance, mental health, survival or functionality as an adult. 3 moderate quality RCTs[7-9] showed a benefit of treatment on language impairment SMD of 0.8 [95%CI 0.02,	**QoE for benefits of treatment:** <u>RCT Evidence:</u> <u>Language impairment</u>: Two studies were rated as unclear risk of bias, one study was rated as high risk of bias, and therefore, overall the body of evidence was rated as moderate risk of bias. Given that all the data on this outcome comes from evidence at moderate risk of bias, the body of evidence was downgraded for serious study limitations. <u>Adaptive functioning:</u> 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

	1.6].[1] 1 RCT[9] on treatment for language impairment provided data for the outcome of social and personal activities of daily living (adaptive functioning) and found no effect of 0.60 (95% CI 3.05 to 4.25).[1] 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. [1] No other RCTs examining treatment of developmental delay were located. [1] 5 systematic reviews[1] evaluated the benefits of a different type of behavioural intervention on the treatment of autism spectrum disorders.[10-14]Results from one systematic review are excluded due to significant duplication with two other reviews.[14] Two reviews found a significant improvement in cognitive function with behavioural intervention [EIBI SMD 0.76 (95% CI 0.04 to 1.11); I^2=21%;[10]ABA1.34 (0.60 to 2.08); [11] one found no evidence that parent mediated behavioural intervention improved outcomes compared to standard care, [12] and one found inconsistent results (no significant improvements with acupuncture, while acupressure improved non-verbal comprehension and matching but not developmental aspects). [13] One review found a small to moderate improvement in quality of life [SMD 0.55 (95 CI 0.24, 0.87; n=171)], [10]while one other did not identify any primary studies that met their inclusion criteria. [13]	Imprecision due to effect estimate including null value. <u>Systematic review evidence</u>: Although the quality of the primary studies was not assessed by the ERSC, the authors of the systematic reviews all expressed concerns about the quality of these primary studies, including serious concerns about the risk of bias, lack of blinding, and imprecision due to small sample sizes and potential publication bias.[25,26] Also, the results of these reviews are difficult to interpret, as many of the control groups in each of the reviews received some form of intervention, and therefore most studies were comparing the effects of more intensive and less intensive interventions rather than intervention and standard care . **QoE for harms treatment:** The systematic reviews were not quality assessed.

	QoE for harms of treatment: 1 review of behavioural interventions found none[10] while the review on acupuncture/acupressure found inconsistent results (some studies identified no harms while for others mild harms such as crying or irritability were reported). [13] No RCT evidence on the harms of treatment was found. [1]	
Balance of benefits and harms *Is there certainty that the benefits outweigh the harms?* Yes☐ No ☒	In the judgment of the CTFPHC, the lack of convincing RCT evidence demonstrating the long-term benefits associated with screening for developmental delay, the limited evidence on the benefits of treating children with clinically identified developmental delay, and the relatively poor diagnostic properties of available screening tests warrant a *strong* recommendation *against* population-based screening.	Is the baseline risk for benefit similar across subgroups? Yes☒ No☐ Should there be separate recommendations for subgroups based on risk levels? Yes☐ No ☒ Is the baseline risk for harm similar across subgroups? Yes ☒ No ☐ Should there be separate recommendations for subgroups based on harms? Yes☐ No☒
Values and preferences *Is there confidence in the estimate of relative importance of outcomes and patient preferences?* Yes☒ No ☐	No studies examining preferences and values in relation screening for developmental delay were identified in the literature. The task force felt that parents would only want their children to be screened for developmental delay if benefit from screening and treatment had been confirmed.	Perspective taken: Patient Source of values and preferences: Relative value of importance of outcomes determined by the guideline panel. Patient preferences were determined by literature review. Relative value of importance of outcomes determined by the guideline panel. Patient preferences were determined by literature review. Source of variability, if any: No evidence identified No evidence identified Method for determining values satisfactory for this

		recommendation? Yes☒ No☐ All critical outcomes measured? Yes ☐ No☒
Resource implications *Are the resources worth the expected net benefit?* Yes☐ No ☒	Costs were not considered in developing the recommendations as we did not find any evidence on the cost-effectiveness of the intervention. However, the CTFPHC considers that given that screening tests had poor to moderate accuracy, screening would lead to a high rate of false positive screens, which may consume resources that would otherwise be available for the care of children that have clinically evident developmental delay.	What are the costs per resource unit? Not available. Feasibility: Is this intervention generally available? Yes☒ No ☐ Opportunity cost: Is this intervention and its effects worth withdrawing or not allocating resources from other interventions? Yes☒ No ☐ Is there lots of variability in resource requirements across settings? Not available.
Overall strength of recommendation: STRONG	**The guideline panel recommends against screening for developmental delay using standardized tools in children aged 1 to 4 years with no apparent signs of developmental delay and whose parents and clinicians have no concerns about development.**	
Remarks and values and preference statement	The CTFPHC recommendation places a relatively higher value on the lack of RCTs evaluating the benefits and harms of screening, the lack of evidence demonstrating the benefit and harm of treatment on long-term clinical important outcomes, the poor reliability The CTFPHC places a relatively higher value on the absence of direct evidence showing that screening is beneficial, the poor diagnostic accuracy of screening tests, the risk of false positives that could result from screening, and the potential for screening to divert resources from the treatment of children with clinically evident DD. The CTFPHC places a relatively lower value on indirect evidence from the few relatively small studies that suggest a benefit of treating certain forms of clinically evident DD, and on the lack of evidence on harms and parents/caregivers preferences and values in relation to screening. The evidence supporting this recommendation is rated overall as low quality because although the systematic review found low quality evidence examining the effect of screening on academic performance and moderate quality evidence examining the effect of treatment on language impairment and cognition, the review did not identify any evidence for the remaining 6 outcomes that were identified at the outset as critical for developing the recommendations: 1) academic performance, 2) improvement to gross and fine motor skills, 3) adaptive function, 4) mental health, 5) survival, and 6) functionality as an adult.	

References

1. 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/

2. Guevara JP, Gerdes M, Localio R, et al. Effectiveness of developmental screening in an urban setting. *Pediatrics*. 2013;131:30-7.

3. 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.

4. 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;32:499-511.

5. Steenis L, Verhoeven M, Hessen D, et al. Parental and professional assessment of early child development: the ASQ-3 and the Bayley-III-NL. *Early Hum Dev* 2015;91:217-25.

6. Cairney J, Clinton J, Veldhuizen S, et al. Evaluation of the revised Nipissing District Developmental Screening (NDDS) tool for use in general population samples of infants and children. *BMC Pediatrics.*

7. Hund-Reid CSP. Effectiveness of phonological awareness intervention for kindergarten children with language impairment. *Can J Speech Lang Pathol Audiol* 2013;37:6-25.

8. Buschmann A, Jooss B, Rupp A, et al. Parent based language intervention for 2-year-old children with specific expressive language delay: a randomised controlled trial. *Arch Dis Child* 2009;94:110-6.

9. Glogowska M, Roulstone S, Enderby P, et al. Randomised controlled trial of community based speech and language therapy in preschool children. *BMJ* 2000;321:923-6.

10. 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;30:387-99.

11. Reichow B, Barton EE, Boyd BA, et al. Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD). *Cochrane Database Syst Rev* 2012;10:CD009260.

13. 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.

14. Cheuk DK, Wong V, Chen WX. Acupuncture for autism spectrum disorders (ASD). *Cochrane Database Syst Rev* 2011;(9):CD007849.

15. 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;154:338-44