Appendix 2 (as supplied by the authors): Detailed methods

The CTFPHC uses a standard process for the development of all clinical practice guidelines.[1,2] (please also refer to the CTFPHC website http://canadiantaskforce.ca/methods/methods-manual/). CTFPHC guidelines are developed using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE)[3] system. Work on each set of recommendations is led by a workgroup of 2 to 6 members of the task force. The workgroup establishes an a priori research design with predetermined key questions, analytic framework, outcomes of interest and search strategy to guide a systematic review of evidence. An independent organization is commissioned to conduct the systematic review of evidence using the a priori framework. The systematic review also includes an assessment of the methodological quality of the individual studies included in the review. The work group further evaluates the strength and quality of the overall body of evidence for each of the outcomes of interest for each the research questions and considers of the balance of benefits and harms for specific interventions, patient values and preferences, and resource considerations. Recommendations are formulated based upon this comprehensive assessment of evidence. Rationale for the recommendations, and judgement and values applied by the guideline panel are reported as part of guideline. Each phase of process includes peer review by methodologists and content experts. Additionally, stakeholders are invited to provide comments on draft guidelines. All members of the CTFPHC reviews and approve each phase of guideline development.

The developmental delay (DD) guideline was led by a work group comprised of 5 CTFPHC members and one clinical expert, with support from scientific staff at the Public Health Agency of Canada (PHAC). The work group established the key and contextual questions, outcomes, analytical framework, and search strategy that were used to develop the research protocol.[4] The Evidence Review and Synthesis Centre (ERSC) at McMaster University (Hamilton, Ontario) independently conducted a systematic review in accordance with the research protocol.[4,5]

The ERSC systematic review[5] examined the benefits and harms of screening for developmental delay, optimal screening intervals, benefits and harms of treatment and accuracy of diagnostic tests. Studies examining externalizing and conduct disorders were excluded as these conditions are usually identified in school aged-children. Studies examining screening for reduced vision and hearing were excluded as these conditions are usually detected through other screening programs. Genetic, molecular, and metabolic tests and neuroimaging tests were also excluded. The systematic review took place in four stages. An outline of the analytic framework with key and contextual questions covering stages I, II and III of the review is available in Appendix 1.

Stage I examined the benefits and harms of screening in children aged 1 to 4 and focused on the following critical outcomes: improved cognitive function, academic performance, quality of life, mental health, survival, and functional status as an adult.[5,6] The outcome on cognitive function was only reported in the ERSC review if the included studies reported the outcome explicitly as "cognitive function". The review also examined two intermediate outcomes, which were referral rates and change in time to referral. The search for harms examined false positives and potential consequences. The search included RCTs or controlled cohort studies examining the effectiveness of screening for

developmental delay. There were no study design limitations on the harms literature. Inclusion criteria for screening studies required a focus on screening children aged 1 to 4 without suspected developmental delay. Studies that included children suspected of having a DD or already diagnosed with a DD were excluded, as such a study could not provide evidence on the effectiveness of screening for identifying otherwise unrecognized cases of DD. Screening studies required a comparator arm of no screening or standard care. Studies were limited to those carried out clinical practice and public health settings. The databases searched were Medline, Embase and PsychINFO with no beginning date limitations through to February 24th, 2014. Since no evidence was found examining the effect of screening on clinical outcomes or harms, stage II of the review was initiated.

Stage II examined the effectiveness of treatment in children aged 1 to 6 and focused on the same clinical outcomes as stage I. [5] The literature search included systematic reviews and RCTs. Age range was increased by two years to allow sufficient time to observe treatment effects in children diagnosed at age 4. Critical outcomes included: cognitive function; academic performance; incidence of mental health conditions; overall quality of life; survival; and functionality as an adult (for treatment of developmental delay and ASD); and improvement to gross and fine motor skills, language, adaptive functioning, and cognition and performance (for domain specific delays). Any harm associated with treatment for developmental delay was included. Systematic reviews examining the outcomes of treatment of Autism were also included as a potential source of indirect evidence on treatment for DD. The databases that were searched to identify systematic reviews included Medline, Embase, PsychINFO and the Cochrane Database of Systematic Reviews from 2000 to June 16th, 2015, and for RCTs it included Embase, Cochrane Central and PsycINFO from January 1st, 2000 to June 16th, 2015to June 16th, 2015. These systematic reviews were assessed for *methodological* quality using the AMSTAR rating tool[7] but the quality of the evidence included in the systematic reviews was not evaluated by the ERSC. The treatment searches were limited to RCTs and systematic reviews (Medline, Embase, and PsycINFO and the Cochrane Database of Systematic Reviews with search date from 2000 to Sept. 15, 2015).

Since stage II initially did not locate any RCTs examining treatment for DD that met the inclusion criteria, an additional search for RCTs examining treatment for domain specific DDs was triggered. The addendum included two additional research questions examining the effectiveness of treatment on domain specific developmental delays in children 1 to 6 years of age. The outcomes examined were clinically relevant changes in gross and fine motor skills, language impairment, adaptive functioning, intellectual disability (IQ), learning disability (academic testing) and academic underachievement. The search was limited to RCTs included in the Medline, Embase, Cochrane Central and PsycINFO databases from January 1st, 2000 to June 16th, 2015. Autism spectrum disorder (ASD) is a distinct condition and therefore the current guideline does not apply to the use of screening tests that aim to detect ASD specifically. However, since screening with tools for overall developmental delay might also detect some cases of ASD, and since children detected with ASD might potentially benefit from treatment, we also searched for systematic reviews describing the benefit of treating ASD. The complete systematic review on screening for developmental delay can be obtained from the CTFPHC at www.canadiantaskforce.ca

Stage III examined the sensitivity, specificity, positive and negative predictive values, and likelihood ratios of screening tests that may be used to assess developmental delay. The literature search did not have limitations based on study design although studies were required to compare the screening test (index test) to a valid reference standard and Index and reference tests had to be administered concurrently or within a brief time interval. Medline, Embase, and PsycINFO databases were searched with no beginning date limitation through March 13th, 2014. A grey literature search was also conducted to search for screening tests used in a Canadian setting. Diagnostic accuracy studies were not evaluated using GRADE methods, however were assessed for bias using the using Quality Assessment of Diagnostic Accuracy Studies 2nd Edition (QUADAS-II).[8]

The CTFPHC used the Grading of Recommendations Assessment, Development and Evaluation (GRADE)[3] system to determine the quality of evidence and strength of recommendations (Box 2 in the main article). The recommendations were reviewed and approved by the entire CTFPHC and underwent external review by content experts in the area and by stakeholders.

The CTFPHC uses a rigorous and collaborative usability testing process to develop knowledge translation tools targeting various end-user groups (e.g., clinicians and patients) to accompany its guidelines. All tools are informed by feedback from clinicians (for clinician and patient tools) and patients (for patient tools) obtained through interviews and/or focus groups and/or surveys."

More information about CTFPHC methods, including the process to update this guideline and the systematic reviews that support the new CTFPHC recommendations, can be found elsewhere[1,2] and on the CTFPHC's website (http://canadiantaskforce.ca/methods/methods-manual/).

References

1. Connor Gorber S, Singh H, Pottie K, et al. Process for guideline development by the reconstituted Canadian Task Force on Preventive Health Care. *CMAJ* 2012;184:1575-81.
2. Procedure manual. Calgary: Canadian Task Force on Preventive Health Care; 2011. Available: http://canadiantaskforce.ca/methods/procedural-manual/ (accessed 2014 Jan. 17)
3. Schunemann H, Brozek J, Oxman A, editors. *GRADE handbook for grading the quality of evidence and strength of recommendations*. GRADE Working Group; 2009.
4. Dunfield L, Mitra D, Tonelli M, et al. Protocol: screening and treatment for developmental delay in early childhood. Calgary: Canadian Task Force on Preventive Health Care; 2014. Available: canadiantaskforce.ca/ctfphc-guidelines/2015-developmental-delay/protocol/ (accessed 2015 Aug. 17).
5. 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/
6. Warren R, Kenny M, Bennett T, et al. Screening for developmental delay among children aged 1-4 years: a systematic review. *CMAJ Open* 2016;4:E20-7.

7. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*2007;7:10.
8. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36.