Diagnostic accuracy of developmental screening in primary care at the 18-month health supervision visit: a cross-sectional study

Meta van den Heuvel MD MSc, Cornelia M. Borkhoff PhD, Christine Koroshegyi MA, Weeda Zabih MD MPH, Sijmen A. Reijneveld MD PhD, Jonathon L. Maguire MD MSc, Catherine S. Birken MD MSc, Patricia C. Parkin MD; TARGet Kids! Collaboration

Abstract

Background: Communication delays are often the first presenting problem in infants with a range of developmental disabilities. Our objective was to assess the validity of the 18-month Nipissing District Developmental Screen compared with the Infant Toddler Checklist, a validated tool for detecting expressive language and other communication delays.

Methods: A cross-sectional design was used. Children aged 18–20 months were recruited during scheduled health supervision visits. Parents completed both the 18-month Nipissing District Developmental Screen and the Infant Toddler Checklist. We assessed criterion validity (diagnostic test properties, overall agreement) for 1 or more "no" responses (1+NDDS flag) and 2 or more "no" responses (2+NDDS flag) using the Infant Toddler Checklist as a criterion measure.

Results: The study included 348 children (mean age 18.6 ± 0.7 mo). The 1+NDDS flag had good sensitivity (94%, 95% confidence interval [CI] 70%–100%, and 86%, 95% CI 64%–96%), poor specificity (63%, 95% CI 58%–68%, and 63%, 95% CI 58%–69%), and fair agreement (0.26) to identify expressive speech and other communication delays, respectively. The 2+NDDS flag had low to fair sensitivity (50%, 95% CI 26%–74%, and 73%, 95% CI 50%–88%), good specificity (86%, 95% CI 82%–90%, and 88%, 95% CI 84%–92%) and moderate agreement (0.45) to identify expressive speech and other communication delays, respectively.

Interpretation: The low specificity of the 1+NDDS flag may lead to overdiagnosis, and the low sensitivity of the 2+NDDS flag may lead to underdiagnosis, suggesting that infants who could benefit from early intervention may not be identified. The Nipissing District Developmental Screen does not have adequate characteristics to accurately identify children with a range of communication delays.

D evelopmental screening in early childhood is recommended by several leading professional organizations, including the Canadian Paediatric Society and the American Academy of Pediatrics.[1–3] In contrast, the recent 2016 evidence-based guidelines from the Canadian Task Force on Preventive Health Care recommends against developmental screening using standardized tools.[4] In the province of Ontario, an expert panel recommended an enhanced 18-month well-baby visit, which was introduced in 2009.[1] The panel recommended the use of standardized tools including the Nipissing District Developmental Screen, which is available free of charge and now commonly used.[5] The conflicting recommendations from professional organizations, evidence-based guidelines and policy-makers present a dilemma for primary care practitioners. Several Canadian experts have expressed concerns regarding the task force recommendation against developmental screening, calling for more research on this topic.[6,7]

Communication delays are often the first presenting problem in young children with a range of developmental disabilities, including isolated expressive speech delay, autism spectrum disorder and intellectual disability.[8–10] About 20% of toddlers have expressive speech delay, which resolves in about 50% of children by 5 years of age.[11] Children with persistent communication delays have poorer outcomes in reading, spelling and math skills, as well as increased emotional and behavioural difficulties.[11,12]

The Nipissing District Developmental Screen is a parent-completed questionnaire that addresses different areas of development, including communication.[13] The accuracy of the screen in identifying early communication delays is unknown. The Infant Toddler Checklist was developed for early identification of children who have or are at risk for developing a communication impairment.[8,14,15] More recently, the checklist was assessed for the accuracy of detecting a range of developmental delays, including language delay, global developmental delay and autism spectrum disorder.[16–18] The Infant Toddler Checklist has been shown to be valid and reliable, and has good diagnostic test properties.[8,15,17,18] The checklist is freely available and is one of 2 measures recommended for early detection of autism spectrum disorder in Canada.[19,20]

We speculate that many Canadian primary care practitioners will continue to use standardized developmental screening tools while waiting for further evidence to guide practice. We recently introduced the Infant Toddler Checklist screening tool into a network of primary care practices whose usual practice is use of the Nipissing District Developmental Screen. Therefore, the primary objective of this study was to assess the validity of the 18-month Nipissing District Developmental Screen compared with the Infant Toddler Checklist. Because parents and clinicians commonly monitor the number of words spoken in early childhood as a measure of language development, our secondary objective was to describe parent responses regarding their 18-month-old child from both screening tools.

Methods

Participants and setting
A cross-sectional prospective design was used. Children aged 18–20 months were recruited from January 2012 to February 2015 on a convenience basis, during scheduled health supervision visits, from primary care practices participating in TARGetKids!, a practice-based research network in Toronto, Ontario (www.targetkids.ca).[21] There are currently 9 large group practices in this network, each of which has between 3 and 10 practising physicians. Trained research assistants in the practices obtained survey and questionnaire data. The study protocol and cohort profile were recently published.[21]

Children were included if their parents completed both the 18-month Nipissing District Developmental Screen and the Infant Toddler Checklist. Participant characteristics were collected using a standardized instrument based on the Canadian Health Measures Survey.[22] Exclusion criteria were health conditions affecting growth, acute or chronic conditions (other than asthma), known severe developmental delay and families

unable to communicate in English. All data were entered into a Web-based data management system (Medidata Rave).

Measures

18-month Nipissing District Developmental Screen
The 18-month Nipissing District Developmental Screen is a 1-page, 17-item, parent-completed tool. Each question addresses 1 or more areas of development: fine motor, gross motor, emotional, social, self-help, learning, thinking and communication.[13] Ten questions address communication. The Nipissing District Developmental Screen is not a standardized tool. Recently, Cairney and colleagues[23] evaluated the psychometric properties of the screen with the Bayley Scales of Infant Development (3rd edition). In that study, the Nipissing District Developmental Screen had modest test–retest reliability (Spearman ρ = 0.62), low sensitivities ranging from 29% to 68% and specificities ranging from 58% to 88% in the identification of developmental delays.[23]

In 2011, the wording of question 6 ("Does your child say 5 or more words?") was changed to "Does your child say 20 or more words?" The screen's authors provide a rationale for this change, highlighting the importance of early referral to a speech pathologist for children with a suspected delay and refer to the MacArthur–Bates Communicative Developmental Inventories.[24,25] We used the 2011 version of the Nipissing District Developmental Screen.

The screen takes 5 minutes to complete. Response options are "yes" or "no." One or more "no" responses (i.e., the child does not demonstrate the behaviour) suggests the need for further assessment or referral. This is known as the "one flag" rule (1+NDDS flag). The "two flag" rule (2+NDDS flag) requires a minimum of 2 "no" responses for referral and follow-up. Currently, the instructions of the 18-month Nipissing District Developmental Screen recommend a "one flag" rule to follow up with the health care or child care professional regarding the child's development.[13] We assessed both the "one flag" and "two flag" rule for the Nipissing District Developmental Screen.

Infant Toddler Checklist
The Infant Toddler Checklist is a 1-page, 24-item, parent-completed tool that can be used in primary care.[14,19] The Infant Toddler Checklist was developed to measure language predictors to determine whether a communication evaluation is needed.[8,14,15] It is a standardized tool with normative scores for monthly intervals for children 6–24 months of age. The Infant Toddler Checklist was originally developed, normed and validated in 2 samples involving American children.[8,26] In 2 further studies that included children 12–24 months of age (n = 232 and n = 915, respectively), the checklist was found to have sensitivity of 87% and 86%, and specificity of 75% and 83%.[15,17] In a large study in pediatric primary care in the United States, the positive predictive value of the checklist was 75% for a range of developmental disorders (autism spectrum disorder, language delay and global development delay).[18]

The checklist takes 5 minutes to complete and 2 minutes to score.[8,18] Most questions are answered with a 3-point scale:

"not yet," "sometimes" and "often." Scoring produces 3 composite scores (social, expressive speech, symbolic) and a total score. The total score and each of the 3 composite scores are dichotomized as "concern/no concern" using the 10th percentile cutoff point. It is recommended that a child should be carefully monitored if the expressive speech composite is below the 10th percentile (concern for expressive speech delay), and the checklist should be re-administered in 3 months.[14,19] In addition, it is recommended that a child should be referred for an evaluation if the social composite, symbolic composite or the total score is below the 10th percentile (concern for other communication delay).[14,19]

Statistical analysis

Means, standard deviations (SDs) and percentages were calculated to characterize the study participants. If a participant's response for an individual question on the Infant Toddler Checklist was missing, we replaced it with a response corresponding to the level below the mode for that domain (e.g., if most answers were "often," then we replaced the missing response with "sometimes"). We assessed criterion validity of the 18-month Nipissing District Developmental Screen by calculating diagnostic test properties and overall agreement. For the purpose of our analysis, the Infant Toddler Checklist was considered a criterion measure. Diagnostic test properties (sensitivity, specificity, positive predictive value, negative predictive value) of the 18-month Nipissing District Developmental Screen compared with the dichotomized scores of the Infant Toddler Checklist were calculated (with 95% confidence intervals [CIs]). For most screening tools, sensitivity of 80% and specificity of 90% are generally recommended.[27] For developmental screening tools, because of behavioural noncompliance, sensitivity between 70% and 80%, and specificity of 80% have been suggested.[28]

Cohen κ coefficients were calculated as a measure of overall agreement between the 2 tools. This measure gives a quantitative assessment of how well the 2 tools agree. Levels of agreement below 0.20, between 0.21 and 0.40, between 0.41 and 0.60, and above 0.61 are considered poor, fair, moderate and good, respectively.[29]

Finally, we described the number of words spoken by 18-month-old children according to parent responses on a comparable single question on both the Nipissing District Developmental Screen and Infant Toddler Checklist, calculating proportions. All analyses were performed with IBM SPSS 20.

Ethics approval

Ethics approval was obtained from the Hospital for Sick Children (Toronto) and St. Michael's Hospital (Toronto) research ethics boards.

Results

Participant characteristics

We included 348 children with a mean age of 18.6 (SD 0.7) months (Table 1). The Nipissing District Developmental Screen and Infant Toddler Checklist were completed at the same visit, with no intervention between measures. On the Nipissing District Developmental Screen, 138 (39.7%) children had 1+NDDS flag and 54 (15.5%) had 2+NDDS flags (Figure 1). On the Infant Toddler Checklist, 16 (4.6%) children were identified as having an expressive speech delay (concern on the expressive speech composite); 22 (6.3%) children were identified as having a communication delay needing referral (concern on the social composite, symbolic composite or total score).

Diagnostic test properties of the 18-month Nipissing District Developmental Screen compared with the Infant Toddler Checklist

To identify children with an expressive speech delay, the 1+NDDS flag was sensitive (94%, 95% CI 70%–100%) but not specific (63%, 95% CI 58%–68%). The 2+NDDS flag

Table 1: Participant characteristics, n = 348

Characteristic	No. (%)*
Age, mean ± SD, mo	18.6 ± 0.7
Sex, male	192 (55.2)
Maternal education	
College or university	328 (94.3)
Nipissing District Developmental Screen	
≥ 1 flag	138 (39.7)
≥ 2 flags	54 (15.5)
Infant Toddler Checklist	
Expressive speech delay†	16 (4.6)
Other communication delays‡	22 (6.3)

Note: SD = standard deviation.
*Unless stated otherwise.
†Need to monitor: expressive speech composite below the 10th percentile.
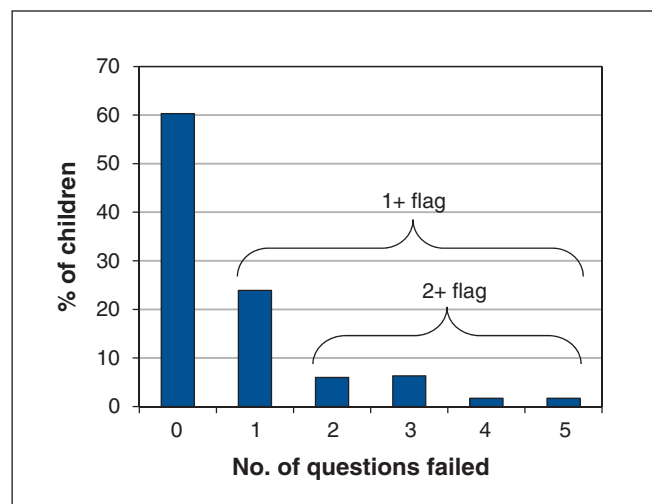‡Need for referral: social composite, symbolic composite or the total score below the 10th percentile.

Figure 1: Percentage of children and the total number of questions failed on the 18-month Nipissing District Developmental Screen.

reduced the sensitivity (50%, 95% CI 26%–74%) but increased the specificity (86%, 95% CI 82%–90%). To identify children with other communication delays (concerns on the communication composite, symbolic composite or total score), the 1+NDDS flag was sensitive (86%, 95% CI 64%–96%) but not specific (63%, 95% CI 58%–69%). The 2+NDDS flag reduced the sensitivity (73%, 95% CI 50%–88%) but increased the specificity (88%, 95% CI 84%–92%) (Tables 2 and 3).

Overall agreement of the 18-month Nipissing District Developmental Screen with the Infant Toddler Checklist

The 1+NDDS flag had fair agreement (Cohen κ 0.26) with the need for monitoring or referral for communication delays on the Infant Toddler Checklist (concerns on any composite or total score). The 2+NDDS flag had moderate agreement (Cohen κ 0.45) with the need for monitoring or referral for communication delays on the Infant Toddler Checklist (concerns on any composite or total score).

Description of parent responses regarding the number of words spoken

Question 6 of the 18-month Nipissing District Developmental Screen ("Does your child speak more than 20 words?") was the question with the greatest number (*n* = 85, 24.4%) of parents responding "no" (Figure 2). Question 17 on the Infant Toddler Checklist ("About how many different words does your child use meaningfully that you recognize?") identi-

Table 2: Diagnostic test properties of the +1NDDS flag compared with the Infant Toddler Checklist

	Infant Toddler Checklist			
	Expressive speech delay*		Other communication delays†	
1+NDDS flag	Yes	No	Yes	No
Positive	15	123	19	119
Negative	1	209	3	207
Sensitivity (95% CI), %	94 (70–100)		86 (64–96)	
Specificity (95% CI), %	63 (58–68)		63 (58–69)	
Positive predictive value (95% CI), %	11 (6–17)		14 (9–21)	
Negative predictive value (95% CI), %	99 (97–100)		99 (96–100)	

Note: 1+NDDS flag = 1 or more "no" responses on the Nipissing District Developmental Screen, CI = confidence interval.
*Need to monitor: expressive speech composite below the 10th percentile.
†Need for referral: social composite, symbolic composite or the total score below the 10th percentile.

Table 3: Diagnostic test properties of the +2NDDS flag compared with the Infant Toddler Checklist

	Infant Toddler Checklist			
	Expressive speech delay*		Other communication delays†	
2+NDDS flag	Yes	No	Yes	No
Positive	8	46	16	38
Negative	8	286	6	288
Sensitivity (95% CI), %	50 (26–74)		73 (50–88)	
Specificity (95% CI), %	86 (82–90)		88 (84–92)	
Positive predictive value (95% CI), %	15 (7–28)		30 (18–44)	
Negative predictive value (95% CI), %	97 (95–99)		98 (95–99)	

Note: 2+NDDS flag = 2 or more "no" responses on the Nipissing District Developmental Screen, CI = confidence interval.
*Need to monitor: expressive speech composite below the 10th percentile.
†Need for referral: social composite, symbolic composite or the total score below the 10th percentile.

fied 5 (1.4%) children who spoke no words at all, 21 (6.0%) who spoke 1–3 words, 81 (23.3%) who spoke 4–10 words, 134 (38.5%) who spoke 11–30 words and 107 (30.7%) who spoke more than 30 words.

Interpretation

We investigated whether the 18-month Nipissing District Developmental Screen, commonly used by Canadian primary care practitioners, can identify communication delays in early childhood in a primary care setting. Compared with the Infant Toddler Checklist, the 1+NDDS flag had good sensitivity, poor specificity and fair agreement to identify expressive speech delay and other communication delays. Thus, the 1+NDDS flag cutoff may result in a large number of false positives, leading to overdiagnosis. The 2+NDDS flag had low to fair sensitivity, good specificity and moderate agreement to identify expressive speech delay and other communication delays. Thus, the 2+NDDS flag cutoff may result in a large number of false negatives, leading to underdiagnosis. Taken together, these findings suggest that the Nipissing District Developmental Screen, at either cutoff, may not have adequate characteristics to accurately identify children with a range of communication delays.

Only a single study has been published that evaluated the psychometric properties of the Nipissing District Developmental Screen. The authors concluded that the screen had poor agreement with the reference standard (Bayley Scales of Infant Development, 3rd ed.), and recommended not to use the Nipissing District Developmental Screen on its own for the identification of developmental delays.[23] Our findings are consistent with this research. In contrast to the Nipissing

District Developmental Screen, the Infant Toddler Checklist has undergone substantial development and validation over the past 2 decades. The checklist was recently assessed in 137 US primary care pediatrician practices at the 12-month visit.[18] Of 10 479 checklists completed, 1318 were positive (12.5%).[18] The 10.9% prevalence of a positive Infant Toddler Checklist in our study is similar to that of the study in US primary care.[18] In their recent guideline, the Canadian Task Force on Preventive Health Care recommended against screening using standardized tools in children (aged 1–4 yr) with no apparent signs of developmental delay or developmental concerns.[4] However, this recommendation was based on low-quality evidence, and studies evaluating the diagnostic properties of the Infant Toddler Checklist were not included.[30] A systematic review of 11 studies involving primary care practitioners who worked without screening tools showed that practitioners achieved poor sensitivity (range 14%–54%) in correctly identifying children who had a developmental concern, suggesting that sensitive screening tools may be valuable.[31]

In our study, almost a quarter of the 18-month-old children did not speak 20 words or more as measured by the Nipissing District Developmental Screen. Almost a third of the 18-month-old children spoke 10 words or fewer according to the Infant Toddler Checklist. Data from "Wordbank," an open database about children's vocabulary growth that archives data from the MacArthur–Bates Communicative Developmental Inventories including data from more than 5000 children, shows a wide vocabulary range at 18 months from 13.9 words (10th percentile) to 269 words (90th percentile).[32] The Bayley Scales of Infant Development (3rd ed.) suggests that fewer than 8 words spoken for a 24-month-old is 1.33 SDs below the normative mean.[33] Our data show that
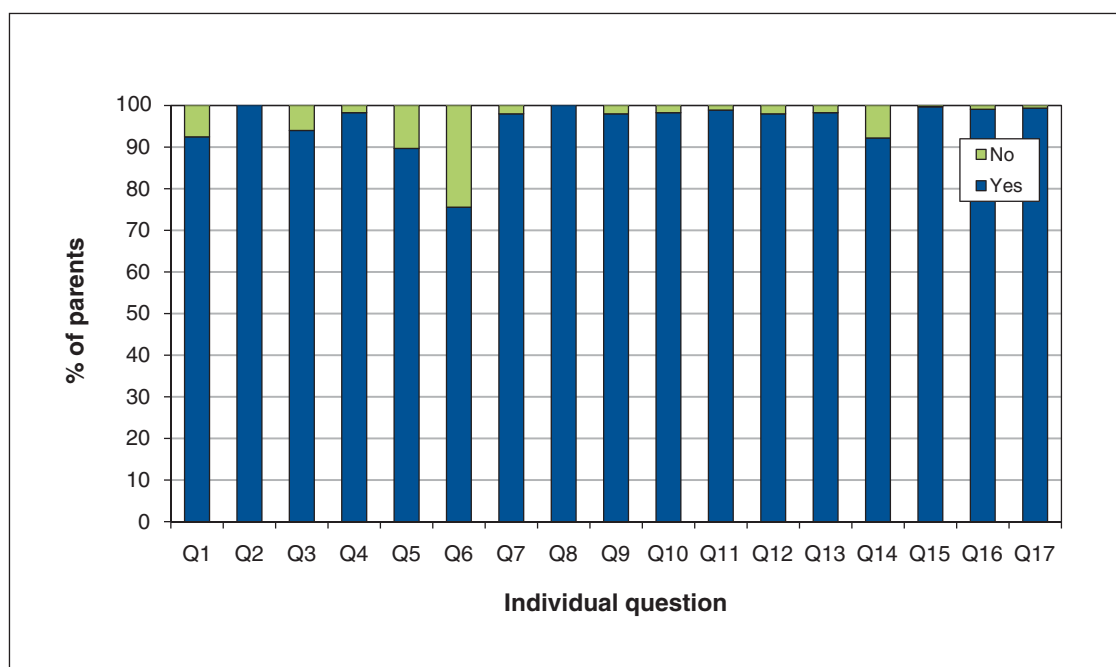

Figure 2: Percentage of "no" and "yes" responses to the individual questions of the 18-month Nipissing District Developmental Screen.

many more parents responded "no" to question 6 than any other question on the 18-month Nipissing District Developmental Screen. This single question accounted for more than 60% of infants with a 1+NDDS flag. Because parents commonly monitor their children's vocabulary, they should be informed about the broad vocabulary range at 18 months to reduce unnecessary concerns identified with the Nipissing District Developmental Screen. In addition, it is important for health care practitioners who work with children to be aware of the fair agreement and low specificity of the 1+NDDS flag for identifying children with expressive speech delay as compared with the Infant Toddler Checklist.

Limitations

Children's communication was not assessed with an independent, standardized measure for comparison, such as the MacArthur–Bates Communicative Developmental Inventories, the Bayley Scales of Infant Development or clinical evaluation. In addition, our study population included only parents who were able to communicate in English, and therefore results may not be generalizable to other populations. Finally, the Nipissing District Developmental Screen is a general development screening tool, whereas the Infant Toddler Checklist was initially developed to target communication delays, so it would be expected that some items (e.g., gross motor) on the Nipissing District Developmental Screen would not correlate with the Infant Toddler Checklist. However, most of the questions on the 18-month Nipissing District Developmental Screen address communication, and the Infant Toddler Checklist has a high positive predictive value for detecting global developmental delay, making the comparison of both tools justifiable. In addition, the authors of the Nipissing District Developmental Screen did not provide independent domains (e.g., "communication domain"); therefore, a specific validation was not feasible.

Conclusion

Our results show that infants with communication delays will not be adequately identified with the 18-month Nipissing District Developmental Screen. The low specificity of the 1+NDDS flag may lead to overdiagnosis and cause unnecessary concern for parents. The lower sensitivity of the 2+NDDS flag may lead to underdiagnosis, suggesting that infants who could benefit from early intervention may not be identified. The Nipissing District Developmental Screen does not have adequate characteristics to accurately identify children with a range of communication delays. With communication delays often being the first presenting problem in young children with a range of developmental disabilities, the Infant Toddler Checklist may be a more promising screening tool for assessing developmental delay in early childhood. A rigorous research agenda is needed to investigate the use of developmental screening tools in primary care.[6]

References

1. Williams R, Clinton J; Canadian Paediatric Society, Early Years Task Force. Getting it right at 18 months: in support of an enhanced well-baby visit. *Paediatr Child Health* 2011;16:647-54.
2. Council on Children With Disabilities; Section on Developmental Behavioral Pediatrics; Bright Futures Steering Committee, et al. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. *Pediatrics* 2006;118:405-20.
3. Committee on Practice and Ambulatory Medicine; Bright Futures Periodicity Schedule Workgroup. 2016 recommendations for preventive pediatric health care. *Pediatrics* 2016;137:25-7.
4. Canadian Task Force on Preventive Health Care. Recommendations on screening for developmental delay. *CMAJ* 2016;188:579-87.
5. Guttmann A, Cairney J, MacCon K, et al. *Uptake of Ontario's enhanced 18-month well baby visit: an AHRQ report.* Toronto: Institute for Clinical Evaluative Sciences; 2016.
6. LeBlanc JC, Belanger SA, Andrews D, et al. Take home message of Task Force report: NOT the strong recommendation against developmental screening, BUT the need for rigorous research and practice [letter]. *CMAJ* 2016 June 10. Available: www.cmaj.ca/content/188/8/579.figures-only/reply#cmaj_el_731967 (accessed 2016 Sept. 8).
7. Dosman CF, Andrews D, Goulden KJ. Screening for surveillance: Let's use validated tools for evidence-based surveillance [letter]. *CMAJ* 2016 May 2. Available: www.cmaj.ca/content/188/8/579.figures-only/reply#cmaj_el_731967 (accessed 2016 Sept. 8).
8. Wetherby AM, Allen L, Cleary J, et al. Validity and reliability of the communication and symbolic behavior scales developmental profile with very young children. *J Speech Lang Hear Res* 2002;45:1202-18.
9. Reilly S, McKean C, Morgan A, et al. Identifying and managing common childhood language and speech impairments. *BMJ* 2015;350:h2318.
10. Carey W, Crocker A, Coleman W, et al. *Developmental-behavioral pediatrics.* 4th ed. Philadelphia: Elsevier; 2009.
11. Reilly S, Wake M, Ukoumunne OC, et al. Predicting language outcomes at 4 years of age: findings from Early Language in Victoria Study. *Pediatrics* 2010;126:e1530-7.
12. Clegg J, Hollis C, Mawhood L, et al. Developmental language disorders — a follow-up in later adult life. Cognitive, language and psychosocial outcomes. *J Child Psychol Psychiatry* 2005;46:128-49.
13. NDDS Ontario [homepage]. North Bay (ON): Nipissing District Developmental Screen. Available: www.ndds.ca/ontario (accessed 2016 May 15).
14. Wetherby A, Prizant B. *Communication and symbolic behavior scales developmental profile.* Preliminary normed ed. Baltimore: Paul H. Brookes Publishing; 2001.
15. Wetherby A, Goldstein H, Cleary J, et al. Early identification of children with communication disorders: concurrent and predictive validity of the CSBS developmental profile. *Infants Young Child* 2003;16:161-74.
16. Wetherby AM, Woods J, Allen L, et al. Early indicators of autism spectrum disorders in the second year of life. *J Autism Dev Disord* 2004;34:473-93.
17. Wetherby AM, Brosnan-Maddox S, Peace V, et al. Validation of the Infant-Toddler Checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. *Autism* 2008;12:487-511.
18. Pierce K, Carter C, Weinfeld M, et al. Detecting, studying, and treating autism early: the one-year well-baby check-up approach. *J Pediatr* 2011;159:458-65. e1-6.
19. Wetherby AM, Barry MP. Communication and Symbolic Behavior Scales Developmental Profile: Infant/Toddler Checklist. Available: http://firstwords.fsu.edu/pdf/Checklist_Scoring_Cutoffs.pdf (accessed 2015 Dec. 29).
20. Anagnostou E, Zwaigenbaum L, Szatmari P, et al. Autism spectrum disorder: advances in evidence-based practice. *CMAJ* 2014;186:509-19.
21. Carsley S, Borkhoff CM, Maguire JL, et al. Cohort profile: the Applied Research Group for Kids (TARGet Kids!). *Int J Epidemiol* 2015;44:776-88.
22. Canadian Health Measures Survey (CHMS). Detailed information for January 2012 to December 2013 (Cycle 3). Ottawa: Statistics Canada; [modified 2014 Oct. 9]. Available: www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5071 (accessed 2015 June 24).
23. 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 Pediatr* 2016;16:42.
24. NDDS [FAQ]. North Bay (ON): Nipissing District Developmental Screen. Available: www.ndds.ca/faq-on.html#FAQ12 (accessed 2015 Nov. 4).
25. Fenson L, Marchman VA, Thal DJ, et al. *The MacArthur-Bates Communicative Developmental Inventories user's guide and technical manual.* Baltimore: Brooks Publishing; 2007.
26. Wetherby A, Prizant B. *Communication and Symbolic Behavior Scales (CSBS).* Normed edition. Chicago: Applied Symbolix; 1993.
27. Streiner D, Norman G, Cairney J. *Health measurement scales: a practical guide to their development and use.* Oxford (UK): Oxford University Press; 2014.
28. Glascoe FP. Early detection of developmental and behavioral problems. *Pediatr Rev* 2000;21:272-80.
29. Kwiecien R, Kopp-Schneider A, Blettner M. Concordance analysis: part 16 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2011;108:515-21.
30. 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.
31. Sheldrick RC, Merchant S, Perrin EC. Identification of developmental-behavioral problems in primary care: a systematic review. *Pediatrics* 2011;128:356-63.
32. Wordbank: an open database of children's vocabulary development. Stanford (CA): Stanford University. Available: http://wordbank.stanford.edu (accessed 2015 Dec. 29).

33. Bayley N. *Bayley scales of infant and toddler development*. 3rd ed. San Antonio (TX): Harcourt Assessment; 2006.

Affiliations: Paediatric Outcomes Research Team (PORT) (van den Heuvel, Borkhoff, Koroshegyi, Zabih, Birken, Maguire, Parkin), Division of Paediatric Medicine, Department of Paediatrics, The Hospital for Sick Children; Institute for Health Policy, Management and Evaluation (Borkhoff, Birken, Maguire, Parkin); The Applied Health Research Centre of the Li Ka Shing Knowledge Institute, and Department of Pediatrics, St. Michael's Hospital (Maguire), University of Toronto; Department of Paediatrics (van den Heuvel, Birken, Maguire, Parkin), Faculty of Medicine, University of Toronto, Toronto, Ont.; Department of Health Sciences (Reijneveld), University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

Contributors: All of the authors critically reviewed the study design. Meta van den Heuvel performed the statistical analysis. Meta van den Heuvel and Patricia Parkin drafted the manuscript, which all of the authors revised. All of the authors gave final approval of the version to be published and agreed to act as guarantors of the work.

Supplemental information: For reviewer comments and the original submission of this manuscript, please see www.cmajopen.ca/content/4/4/E634/suppl/DC1.