Screening for Speech and Language Delay in Children 5 Years Old and Younger: A Systematic Review

Ina F. Wallace, PhD, Nancy D. Berkman, PhD, Linda R. Watson, EdD, Tamera Coyne-Beasley, MD, MPH, Charles T. Wood, MD, Katherine Cullen, BA, Kathleen N. Lohr, PhD

abstract

BACKGROUND AND OBJECTIVES: No recommendation exists for or against routine use of brief, formal screening instruments in primary care to detect speech and language delay in children through 5 years of age. This review aimed to update the evidence on screening and treating children for speech and language since the 2006 US Preventive Services Task Force systematic review.

METHODS: Medline, the Cochrane Library, PsycInfo, Cumulative Index to Nursing and Allied Health Literature, ClinicalTrials.gov, and reference lists. We included studies reporting diagnostic accuracy of screening tools and randomized controlled trials reporting benefits and harms of treatment of speech and language. Two independent reviewers extracted data, checked accuracy, and assigned quality ratings using predefined criteria.

RESULTS: We found no evidence for the impact of screening on speech and language outcomes. In 23 studies evaluating the accuracy of screening tools, sensitivity ranged between 50% and 94%, and specificity ranged between 45% and 96%. Twelve treatment studies improved various outcomes in language, articulation, and stuttering; little evidence emerged for interventions improving other outcomes or for adverse effects of treatment. Risk factors associated with speech and language delay were male gender, family history, and low parental education. A limitation of this review is the lack of well-designed, well-conducted studies addressing whether screening for speech and language delay or disorders improves outcomes.

CONCLUSIONS: Several screening tools can accurately identify children for diagnostic evaluations and interventions, but evidence is inadequate regarding applicability in primary care settings. Some treatments for young children identified with speech and language delays and disorders may be effective.
Speech and language delays and disorders are common, with an estimated prevalence between 5% and 12% (median, 6%) in children 2 to 5 years of age. A speech or language delay implies that the child is developing speech or language in the correct sequence but at a slower rate than expected, whereas a speech or language disorder suggests that the child’s speech or language ability is qualitatively different from what is typical. In this review, we use speech and language “delay,” “disorder,” “impairment,” and “disability” interchangeably.

The American Speech-Language-Hearing Association guidelines describe a speech disorder as an impairment of the articulation of speech sounds, fluency, or voice and a language disorder as impaired comprehension or use of spoken, written, or other symbol systems. A disorder may involve the form of language (phonology, morphology, syntax), the content of language (semantics), and the function of language in communication (pragmatics) in any combination. Because prelinguistic communication behaviors (eg, gestures, babbling, joint attention) are associated with language delays, this review considers screening of both verbal and preverbal communication skills.

Young children with speech and language delay in the preschool years may be at increased risk for learning disabilities once they reach school age. Children with both speech sound disorders and language impairment are at greatest risk for language-based learning disabilities (eg, difficulties in reading and written language). Estimates of the increased risk for poor reading outcomes in grade school are 4 to 5 times greater for children with speech and language impairment than for children with appropriate development; risk persists into adulthood. Adults who had speech and language disorders as children may hold lower-skilled jobs and are more likely to experience unemployment than other adults. Behavior problems and impaired psychosocial adjustment associated with speech and language may also persist into adulthood.

Identifying speech and language problems before children enter school can foster initiation of early interventions before these problems interfere with formal education and behavioral adjustment. AAP clinical guidelines recommend that pediatric health care providers perform surveillance at every well-child visit for children <36 months of age; should concerns arise, screening should be administered using standardized developmental tools. Irrespective of concerns, the guidelines identify 9, 18, and 24 or 30 months as appropriate ages for developmental screening.

In 2006, the US Preventive Services Task Force (USPSTF) concluded that evidence was insufficient to recommend for or against (“1 statement”) routine use of brief, formal screening instruments in primary care to detect speech and language delay in children up to 5 years of age. In 2013, the USPSTF commissioned a new systematic review of the current evidence on brief, formal screening for speech and language delays and disorders in children 5 years old and younger. The USPSTF used it to update its 2006 recommendations about screening in primary care settings.

**METHODS**

Following the USPSTF Procedure Manual, we developed an analytic framework (Supplemental Fig 2), list of key questions (KQs), and supporting contextual questions. We searched Medline (via PubMed), the Cochrane Library, PsycINFO, and Cumulative Index to Nursing and Allied Health Literature for English-language articles published from January 1, 2004, through July 20, 2014. We conducted targeted searches for unpublished literature in ClinicalTrials.gov. Appendix A of the full report documents the search strategy. To supplement electronic searches, we reviewed reference lists of pertinent review articles and included studies. We used a PICOTS (populations, interventions, comparators, outcomes, timing, settings, and study designs) approach to identify studies that met inclusion and exclusion criteria that we developed for each key question (see Appendices B and C of the full report). Two reviewers independently applied inclusion and exclusion criteria to all studies in the 2006 review and to all new studies from our update searches. An investigator abstracted evidence from included full-text articles for each key question; a second investigator checked and confirmed each abstraction. We also checked for errors in the abstractions of studies in the 2006 review. Two reviewers independently rated the quality of each study based on USPSTF guidelines as good, fair, or poor (see Appendix D of the full report); they resolved discrepancies by discussion. We reassessed the quality rating of studies in the 2006 review to ensure that they met current criteria. If 1 reviewer disagreed with this earlier assessment, we rerated the quality of that study through dual review.

We abstracted accuracy statistics when available from screening studies. When investigators did not provide accuracy statistics, we calculated sensitivity, specificity, prevalence, positive and negative predictive values, positive and negative likelihood ratios (LRs), and 95% confidence intervals (CIs) for sensitivity and specificity (see Appendix E of the full report). We evaluated applicability to US primary care populations based on demographics, coexisting conditions,
representativeness of the population, study refusal rate, severity of the delay, and recruitment source and applicability of the intervention/screening (ie, how well the clinical experience is liable to be reproduced in other settings).

This review was funded by the Agency for Healthcare Research and Quality (AHRQ). The USPSTF members and AHRQ Medical Officers helped develop the scope, KQs, and analytic framework that guided our literature search and review.

RESULTS

We document the impact of screening using evidence derived from included studies identified through the 2006 report,22,23 our database and manual searches,19,24 and recommendations from peer reviewers. We had evidence for 5 of 7 KQs (Supplemental Fig 2): we had no evidence for KQ3 (adverse effects of screening) or KQ4 (surveillance by primary care clinicians). Figure 1 shows the flow of studies from initial identification of titles and abstracts to final inclusion or exclusion.

KQ1: Improvements in Outcomes

No study met the 2006 inclusion criteria to determine whether screening improved either speech and language or other outcomes. One randomized controlled trial (RCT) met our inclusion criteria by randomizing a large national sample of children who received regularly scheduled care at child health centers to early screening and measuring outcomes at 8 years of age.25,26 We did not include evidence from this trial owing to a rating of poor quality caused by very high attrition.

KQ2: Accurate Identification of Children for Diagnostic Evaluations and Interventions: Screening Accuracy

We examined the accuracy of screening techniques and whether accuracy varies by demographic and screening source. We included 24 good and fair studies (26 articles):

---

FIGURE 1

Flow diagram of study retrieval and selection. Abstracts of potentially relevant articles reviewed, identified through database searching (1) and other sources (2): (1) Databases include PubMed, Cochrane, PsycInfo, and Cumulative Index to Nursing and Allied Health Literature. (2) Other sources include searching for specific screening instruments, review of reference lists, and suggested by peer reviewers. Some studies are included for more than one key question or contextual question. One systematic review was the review being updated for this report.
8 newly identified studies (9 articles\textsuperscript{27–35}), and 16 studies (17 articles) from the 2006 review\textsuperscript{36–52} (Supplemental Table 3). Supplemental Table 4 describes relevant screening instruments.

**Detailed Synthesis of Evidence on Screening Accuracy**

Tables 1 and 2 present accuracy statistics separately for parent and trained-examiner instruments, respectively. We report sensitivity and specificity (and 95% CIs), prevalence, positive and negative predictive values, and positive and negative LRs. We present median (not mean) values because accuracy statistics were skewed. We report the accuracy statistics by age group when possible.

**Accuracy of Screening Instruments Used by Parents**

Altogether, 14 studies (16 articles\textsuperscript{27–30,32–35,40,42,43,46–49,52}) examined the accuracy of screeners in which parents rated the speech and language skills of their young children (mostly 2 or 3 years of age) (Table 1). Cutoff scores for positive screening (ie, a speech or language problem), when provided, varied by instrument.

Sensitivity for detecting a true speech and language delay or disorder using parent-report screeners ranged between 50% and 94% (median, 81%); specificity for detecting a child without speech and language delays ranged between 45% and 96% (median, 87%). Children with positive screening results (ie, those who failed the screening test) had a moderately\textsuperscript{53} higher likelihood of language delay than children with negative screening results (ie, those who passed) in at least 1 study investigating the Ages and Stages Questionnaire (ASQ), the Communicative Development Inventory (CDI), the Language Development Survey (LDS), the Parent Questionnaire, and Ward’s screening tool. With respect to negative LRs, results from ≥1 studies using the CDI, the Infant-Toddler Checklist (ITC), and LDS suggested a moderately lower likelihood of language delay for those children who passed the screening test relative to those who did not.

**Accuracy of Screening Instruments Used by Trained Examiners**

Twelve studies examined the accuracy of instruments administered by trained examiners, including nurses, primary care providers, teachers, and paraprofessionals (Table 2).\textsuperscript{27,31,36–39,41,44,45,48,50,51} These studies tended to focus on older preschool-age children: 3 studies included children 2 to 3 years of age\textsuperscript{44,45,48}; 1 of children 3 to 4 years of age\textsuperscript{37}; 5 of children 4 to 5 years of age\textsuperscript{27,31,36,50,51}; and 3 of children across different ages (18 to 72 months).\textsuperscript{38,39,41} Several studies included >1 screening instrument. All but 2 instruments require some direct testing of the child; the Developmental Nurse Screen\textsuperscript{48} and the Davis Observation Checklist for Texas (DOCT)\textsuperscript{36} involve ratings made after observing the child.

Sensitivity for detecting a true delay or disorder ranged between 17% and 100% (median, 74%); specificity for detecting typical speech and language ranged between 46% and 100% (median, 91%). In studies of the Battelle Developmental Inventory Screening Test,\textsuperscript{27} DOCT,\textsuperscript{36} Screening Kit of Language Development (SKOLD),\textsuperscript{38} Sentence Repetition Screening Test,\textsuperscript{51} Structured Screening Test,\textsuperscript{44} and the Trial Speech Screening Test,\textsuperscript{31} positive LRs indicated at least a moderately higher likelihood of language delay for those who screened positive; the studies of the Brigance Preschool Screening Test,\textsuperscript{27} DOCT,\textsuperscript{36} Early Screening Test,\textsuperscript{27} Hackney Early Language Screening Test,\textsuperscript{45} Northwestern Syntax Screening Test,\textsuperscript{37} and SKOLD,\textsuperscript{38} indicated at least a moderately lower likelihood of language delay for those who screened negative.

**Accuracy by Age of Child**

ASQ sensitivity was marginally higher for older children (4.5 years) in 1 study\textsuperscript{27} than for younger children (2 to 3 years) in 2 other studies.\textsuperscript{28,29} However, in the latter 2 studies, the positive LRs indicated at least a moderately higher likelihood of a language delay in children who screened positive relative to children who screened negative; we saw no such increase in the likelihood of delay in the study of older children. The negative LRs were small and equivalent for both younger and older samples.

Four of the 5 CDI studies examined the accuracy of the toddler version (18 to 36 months).\textsuperscript{29,30,32–34} The fifth study used the preschool version with children 36 to 62 months of age.\textsuperscript{26} Accuracy of the 2 versions was similar. The 1 ITC study separately considered 2 age groups of toddlers (12 to 17 months; 18 to 24 months); accuracy was similar for younger and older toddlers.\textsuperscript{35}

**Accuracy of Longer-Term Prediction**

Two studies examined the accuracy of parent-reported screeners for predicting long-term language delay.\textsuperscript{32,33,42,43} Both studies examined the accuracy of the screener at 2 years in relation to the reference standard (a diagnostic tool) at both 2 years and 3 years. In the LDS study,\textsuperscript{43} sensitivity for detecting a language delay at 3 years was 67% (91% at 2 years). Specificity for detecting typical language development at 3 years was 93% (96% at 2 years). In the ELFRA-2 (ie, German CDI) study,\textsuperscript{32,33} sensitivity and specificity at 3 years were 94% (93% at 2 years) and 61% (88% at 2 years), respectively.
<table>
<thead>
<tr>
<th>Instrument and Version</th>
<th>Decision Cutoff Point</th>
<th>Reference</th>
<th>USPSTF Quality Rating</th>
<th>Child Age</th>
<th>n</th>
<th>Reference Instrument</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>Prevalence, %a</th>
<th>PPV, %ab</th>
<th>NPV, %ab</th>
<th>PLR, %a</th>
<th>NLR, %a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ages and Stages Questionnaire, 2nd ed.</td>
<td>“Recommended cutoff”</td>
<td>Frisk et al 200927</td>
<td>Fair</td>
<td>4.5 y</td>
<td>110</td>
<td>PLS-4 Receptive</td>
<td>67 (45–88)</td>
<td>73 (64–82)</td>
<td>16</td>
<td>32</td>
<td>92</td>
<td>2.4</td>
<td>0.46</td>
</tr>
<tr>
<td>Ages and Stages Questionnaire, Spanish version</td>
<td></td>
<td>Guiberson et al 201129</td>
<td>Fair</td>
<td>24–35 mo</td>
<td>45</td>
<td>PLS-4, Spanish edition</td>
<td>75 (54–91)</td>
<td>76 (67–85)</td>
<td>20</td>
<td>43</td>
<td>92</td>
<td>3.0</td>
<td>0.36</td>
</tr>
<tr>
<td>Ages and Stages Questionnaire, Spanish version</td>
<td></td>
<td>Guiberson and Rodríguez 201028</td>
<td>Fair</td>
<td>32–36 mo</td>
<td>48</td>
<td>PLS-4, Spanish edition</td>
<td>58 (38–80)</td>
<td>62 (82–100)</td>
<td>46</td>
<td>87</td>
<td>73</td>
<td>7.7</td>
<td>0.44c</td>
</tr>
<tr>
<td>SC518: Swedish CDI WS</td>
<td></td>
<td>Westerlund et al 200454</td>
<td>Fair</td>
<td>18 mo</td>
<td>881</td>
<td>Language Observation, 3 y</td>
<td>50 (34–66)</td>
<td>90 (89–92)</td>
<td>4</td>
<td>18c</td>
<td>89c</td>
<td>4.8c</td>
<td>0.08c</td>
</tr>
<tr>
<td>CDI WS</td>
<td>&lt;19th percentile</td>
<td>Heilmann et al 200550</td>
<td>Fair</td>
<td>24 mo</td>
<td>100</td>
<td>PLS-3</td>
<td>81 (69–94)</td>
<td>79 (69–89)</td>
<td>38</td>
<td>70c</td>
<td>89c</td>
<td>3.9</td>
<td>0.23c</td>
</tr>
<tr>
<td>ELFRA-2: German CDI</td>
<td>Words and Sentences</td>
<td>Sachse and Von Suchodeletz 2008, 200932,33</td>
<td>Good</td>
<td>24–26 mo</td>
<td>117</td>
<td>SETK2</td>
<td>93 (87–99)</td>
<td>87 (78–87)</td>
<td>59</td>
<td>91c</td>
<td>89c</td>
<td>7.3</td>
<td>0.08c</td>
</tr>
<tr>
<td>Short Form Inventarios del Desarrollo de Habilidades Comunicativas: Spanish CDI WS</td>
<td></td>
<td>Guiberson et al 201129</td>
<td>Fair</td>
<td>24–35 mo</td>
<td>45</td>
<td>PLS-4, Spanish edition</td>
<td>87 (73–100)</td>
<td>86 (72–100)</td>
<td>51</td>
<td>87c</td>
<td>86c</td>
<td>6.4</td>
<td>0.15c</td>
</tr>
<tr>
<td>Piloto Inventario III: Spanish CDI</td>
<td></td>
<td>Guiberson and Rodríguez 201028</td>
<td>Fair</td>
<td>32–36 mo</td>
<td>48</td>
<td>PLS-4, Spanish edition</td>
<td>82 (66–96)</td>
<td>81 (66–96)</td>
<td>46</td>
<td>78</td>
<td>84</td>
<td>4.2c</td>
<td>0.22c</td>
</tr>
<tr>
<td>General Language Screen</td>
<td>≥2 of 11 items endorsed</td>
<td>Stott et al 200240</td>
<td>Fair</td>
<td>36 mo</td>
<td>506</td>
<td>DPII (37 mo)</td>
<td>75 (67–83)</td>
<td>81 (77–84)</td>
<td>8%c</td>
<td>47</td>
<td>94</td>
<td>3.9c</td>
<td>0.31c</td>
</tr>
<tr>
<td>Parent Language Checklist: previous version of the General Language Screen</td>
<td>1 failed item</td>
<td>Burden et al 199640</td>
<td>Good</td>
<td>36 mo</td>
<td>425</td>
<td>Renfrew Action Picture Test, Bus Story, study-derived tests of phonology and comprehension</td>
<td>87 (82–93)</td>
<td>85 (59–51)</td>
<td>32</td>
<td>42</td>
<td>89</td>
<td>1.6</td>
<td>0.28</td>
</tr>
<tr>
<td>Infant-Toddler Checklist</td>
<td></td>
<td>Wetherby et al 200355</td>
<td>Fair</td>
<td>12–17 mo</td>
<td>151</td>
<td>CSBS Behavior Sample</td>
<td>88 (80–87)</td>
<td>74 (68–83)</td>
<td>35</td>
<td>65c</td>
<td>92c</td>
<td>3.5c</td>
<td>0.15c</td>
</tr>
<tr>
<td>Language Development Survey</td>
<td>&lt;50 words or no word combinations</td>
<td>Klee et al 199842</td>
<td>Fair</td>
<td>24–26 mo</td>
<td>64</td>
<td>Clinical judgment on infant MSEL language scales, MLU</td>
<td>91 (74–100)</td>
<td>87 (78–96)</td>
<td>17</td>
<td>59</td>
<td>98</td>
<td>6.9</td>
<td>0.10</td>
</tr>
<tr>
<td>Language Development Survey Study 2</td>
<td>≥28 screening score</td>
<td>Klee et al 2000445</td>
<td>Fair</td>
<td>24–26 mo</td>
<td>64</td>
<td>RDLS Expressive</td>
<td>91 (74–100)</td>
<td>96 (91–100)</td>
<td>17</td>
<td>83</td>
<td>98</td>
<td>24.1</td>
<td>0.09</td>
</tr>
<tr>
<td>Language Development Survey Study 2</td>
<td>&lt;50 words or no word combinations</td>
<td>Rescorla and Alley 200147</td>
<td>Fair</td>
<td>25.4 mo</td>
<td>66</td>
<td>RDLS Expressive</td>
<td>94 (84–100)</td>
<td>67 (53–80)</td>
<td>27</td>
<td>52c</td>
<td>97c</td>
<td>2.8</td>
<td>0.08c</td>
</tr>
</tbody>
</table>
to 36 months, 37 to 42 months, and 43 to 48 months, median sensitivity rates were 94%, 94%, and 97%, respectively; median specificity rates were 92%, 88%, and 85%. Across the 3 age levels, median sensitivity and specificity were 88% and 86% for the African American dialect versions and 100% and 93% for the Standard English versions.

KQ5: Treatment: Speech and Language Outcomes

Thirteen RCTs (6 newly identified)54–59 in 14 articles evaluating speech and language interventions and 1 systematic review met criteria for inclusion (Supplemental Table 5). Of these, 11 examined language outcomes and 8 measured speech outcomes. The systematic review of treatment of childhood apraxia of speech failed to find any studies that met our inclusion criteria, so we did not consider it further.60

Language

Of 11 studies measuring language outcomes (Supplemental Table 6), 4 tested treatments primarily or exclusively delivered in English. Group training on language activities for parents of toddlers with limited expressive language found significant effects on expressive and receptive language.61 Finally, 1 group of parents learned activities to target speech sounds and a second group of parents shared storybooks with their children63; neither treatment was associated with gains in child expressive syntax or semantic knowledge compared with the control group. Two trials tested treatments primarily or exclusively delivered in Spanish.91,92 In 1 trial testing variations of the Hanen Parent Program93,94 for toddlers with language delays, 1 found significant effects on expressive language measures favoring the treatment group.94 In 2 trials testing interventions that included language and social skills training, 1 found significant expressive language effects associated with gains in social skills.96,97

TABLE 1 Continued

<table>
<thead>
<tr>
<th>Instrument and Version</th>
<th>Decision Cutoff Point</th>
<th>Reference</th>
<th>Quality Rating</th>
<th>Child Age</th>
<th>n</th>
<th>Reference Instrument</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>Prevalence, %</th>
<th>PPV, %a,b</th>
<th>NPV, %a,b</th>
<th>PLR, %a</th>
<th>NLR, %a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Development Survey Study 3</td>
<td>&lt;50 words or no word combinations</td>
<td>Rescorla 198946</td>
<td>Fair</td>
<td>24–34 mo</td>
<td>81</td>
<td>RDLS Expressive</td>
<td>89 (80–98)</td>
<td>88 (75–97)</td>
<td>56</td>
<td>89</td>
<td>86</td>
<td>6.4</td>
<td>0.13</td>
</tr>
<tr>
<td>Parent Questionnaire</td>
<td>NR</td>
<td>Stokes 199748</td>
<td>Fair</td>
<td>34–40 mo</td>
<td>381</td>
<td>SLP rating using language sample, RDLS Comprehension</td>
<td>78 (66–89)</td>
<td>91 (88–94)</td>
<td>13</td>
<td>58c</td>
<td>95c</td>
<td>8.3</td>
<td>0.24</td>
</tr>
<tr>
<td>Ward's Created Screening Tool</td>
<td>≥1 item</td>
<td>Ward 19942</td>
<td>Fair</td>
<td>7–23 mo</td>
<td>1070</td>
<td>REEL</td>
<td>80 (75–85)</td>
<td>92 (60–94)</td>
<td>24c</td>
<td>75</td>
<td>94</td>
<td>96</td>
<td>0.22</td>
</tr>
</tbody>
</table>

BPVS, British Picture Vocabulary Scales; CDI, Communicative Development Inventory; WS, Words and Sentences; CSBS, Communication and Symbolic Behavior Scales; DPII, Developmental Profile II; ELFRA, Elternfragebogen fur die Fruberkenntung von Riskokindern; MLU, mean length of utterances; MSEL, Mullen Scales of Early Learning; NR, not reported; NLR, negative likelihood ratio; NPV, negative predictive value; PLS, Preschool Language Scale; PLR, positive likelihood ratio; PPV, positive predictive value; REEL, Receptive Expressive Emergence of Language; RDLS, Reynell Developmental Language Scales; SCS18, Swedish Communication Screening at 18 mo of age; SETK/2, Sprachentwicklungstest fur zweijahrige kinder; SLP, speech language pathologist.

a Calculated by EPC authors unless otherwise noted that study investigators provided data. Prevalence values were not estimated or weighted to reflect sampling.

b Predictive values may be questionable for studies in which prevalence exceeded 10%; the problem arises when investigators choose a random sample of children with negative screens to complete the reference measures.

c Study investigators provided data.

d Could not calculate because of lack of data in article.

e Same data using a different decision rule for failing screener.
<table>
<thead>
<tr>
<th>Instrument and Component</th>
<th>Decision Cutoff Point</th>
<th>Reference</th>
<th>USPSTF Quality Rating</th>
<th>Child Age</th>
<th>n</th>
<th>Reference Instrument</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>Prevalence</th>
<th>PPV&lt;sup&gt;a&lt;/sup&gt;&lt;sup&gt;b&lt;/sup&gt;</th>
<th>NPV&lt;sup&gt;a&lt;/sup&gt;&lt;sup&gt;b&lt;/sup&gt;</th>
<th>PLR&lt;sup&gt;a&lt;/sup&gt;</th>
<th>NLR&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Battelle Developmental Inventory Screening Test, Receptive</td>
<td>&lt;1 SD</td>
<td>Frisk et al 2009&lt;sup&gt;27&lt;/sup&gt;</td>
<td>Fair</td>
<td>4.5 y</td>
<td>110</td>
<td>PLS-4 Receptive</td>
<td>56 (33–78)</td>
<td>70 (60–79)</td>
<td>16</td>
<td>26</td>
<td>89</td>
<td>1.8</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PLS-4 Expressive</td>
<td>68 (49–88)</td>
<td>86 (79–94)</td>
<td>20</td>
<td>56</td>
<td>92</td>
<td>5.0</td>
<td>0.37</td>
</tr>
<tr>
<td>Brigance Preschool Screen</td>
<td>&lt;1 SD</td>
<td>Frisk et al 2009&lt;sup&gt;27&lt;/sup&gt;</td>
<td>Fair</td>
<td>4.5 y</td>
<td>110</td>
<td>PLS-4 Receptive</td>
<td>61 (39–84)</td>
<td>70 (58–79)</td>
<td>20</td>
<td>51</td>
<td>97</td>
<td>0.12</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PLS-4 Expressive</td>
<td>68 (49–88)</td>
<td>86 (79–94)</td>
<td>20</td>
<td>51</td>
<td>97</td>
<td>0.12</td>
<td>0.65</td>
</tr>
<tr>
<td>Davis Observation Checklist for Texas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PLS-4 Expressive</td>
<td>80 (55–100)</td>
<td>98 (94–100)</td>
<td>17</td>
<td>89</td>
<td>96</td>
<td>39.2</td>
<td>0.20</td>
</tr>
<tr>
<td>Denver Articulation Screening Test</td>
<td>&lt;15th percentile</td>
<td>Drumwright et al 1973&lt;sup&gt;41&lt;/sup&gt;</td>
<td>Fair</td>
<td>30–72 mo</td>
<td>150</td>
<td>Henja Articulation Test</td>
<td>92 (&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>97 (&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Denver Developmental Screening Test Language Sector</td>
<td>NR</td>
<td>Borowitz and Glascoe 1986&lt;sup&gt;30&lt;/sup&gt;</td>
<td>Fair</td>
<td>18–46 mo</td>
<td>71</td>
<td>PLS</td>
<td>46 (34–58)</td>
<td>100 (100–100)</td>
<td>92</td>
<td>100</td>
<td>15</td>
<td>&lt;sup&gt;e&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Developmental Nurse Screen</td>
<td>NR</td>
<td>Stokes 1997&lt;sup&gt;48&lt;/sup&gt;</td>
<td>Fair</td>
<td>34–40 mo</td>
<td>378</td>
<td>SLP rating using language sample, RDLS Comprehension</td>
<td>78 (&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>96 (&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Early Screening Profile Verbal Concepts</td>
<td>&lt;1 SD</td>
<td>Frisk et al 2009&lt;sup&gt;27&lt;/sup&gt;</td>
<td>Fair</td>
<td>4.5 y</td>
<td>110</td>
<td>PLS-4 Auditory</td>
<td>94 (84–100)</td>
<td>68 (59–78)</td>
<td>16</td>
<td>40</td>
<td>98</td>
<td>3.0</td>
<td>0.08</td>
</tr>
<tr>
<td>Fluharty Preschool Screening Test</td>
<td>Failure ≥1 subtests</td>
<td>Allen and Bliss 1987&lt;sup&gt;5,7&lt;/sup&gt;</td>
<td>Fair</td>
<td>36–47 mo</td>
<td>182</td>
<td>SICD</td>
<td>86 (72–100)</td>
<td>81 (72–89)</td>
<td>20</td>
<td>53</td>
<td>96</td>
<td>4.5</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PLS-4 Expressive</td>
<td>60 (41–78)</td>
<td>81 (75–87)</td>
<td>14</td>
<td>33</td>
<td>93</td>
<td>3.1</td>
<td>0.49</td>
</tr>
<tr>
<td>FPSLST Articulation</td>
<td>NR</td>
<td>Sturner et al 1993&lt;sup&gt;50&lt;/sup&gt;</td>
<td>Fair</td>
<td>53–68 mo</td>
<td>51</td>
<td>MASC, GFTA, informal language sample</td>
<td>74 (&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>96 (&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>4&lt;sup&gt;c&lt;/sup&gt;</td>
<td>50</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>FPSLST Language</td>
<td>NR</td>
<td>Sturner et al 1993&lt;sup&gt;50&lt;/sup&gt;</td>
<td>Fair</td>
<td>53–68 mo</td>
<td>51</td>
<td>TACL-R</td>
<td>38 (&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>85 (&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>17&lt;sup&gt;c&lt;/sup&gt;</td>
<td>42</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>FPSLST Articulation</td>
<td>NR</td>
<td>Sturner et al 1993&lt;sup&gt;50&lt;/sup&gt;</td>
<td>Fair</td>
<td>55–69 mo</td>
<td>147</td>
<td>TD</td>
<td>43 (&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>93 (&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>5&lt;sup&gt;c&lt;/sup&gt;</td>
<td>26</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>FPSLST Language</td>
<td>NR</td>
<td>Sturner et al 1993&lt;sup&gt;50&lt;/sup&gt;</td>
<td>Fair</td>
<td>55–69 mo</td>
<td>147</td>
<td>TOLD-P</td>
<td>17 (&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>97 (&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>22&lt;sup&gt;c&lt;/sup&gt;</td>
<td>50</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
<td>&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Northwestern Syntax Screening Test</td>
<td>Failure ≥1 subtests</td>
<td>Allen and Bliss 1987&lt;sup&gt;5,7&lt;/sup&gt;</td>
<td>Fair</td>
<td>36–47 mo</td>
<td>182</td>
<td>SICD</td>
<td>92 (81–100)</td>
<td>48 (41–56)</td>
<td>14</td>
<td>22</td>
<td>97</td>
<td>1.8</td>
<td>0.16</td>
</tr>
<tr>
<td>SKOLD</td>
<td></td>
<td>Bliss and Allen 1984&lt;sup&gt;5,8&lt;/sup&gt;</td>
<td>Fair</td>
<td></td>
<td></td>
<td></td>
<td>81–100</td>
<td>41–56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKOLD&lt;sub&gt;DS30&lt;/sub&gt;</td>
<td>&lt;11</td>
<td></td>
<td>30–36 mo</td>
<td>47</td>
<td>SICD</td>
<td>100 (100–100)</td>
<td>98 (93–100)</td>
<td>6</td>
<td>75</td>
<td>100</td>
<td>44.0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SKOLD&lt;sub&gt;DS37&lt;/sub&gt;</td>
<td>&lt;10</td>
<td></td>
<td>37–42 mo</td>
<td>93</td>
<td>SICD</td>
<td>100 (100–100)</td>
<td>91 (85–97)</td>
<td>11</td>
<td>33</td>
<td>100</td>
<td>11.1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SKOLD&lt;sub&gt;DS43&lt;/sub&gt;</td>
<td>&lt;19</td>
<td></td>
<td>43–48 mo</td>
<td>100</td>
<td>SICD</td>
<td>100 (100–100)</td>
<td>93 (88–98)</td>
<td>9</td>
<td>60</td>
<td>100</td>
<td>15.2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>African American dialect</td>
<td>SKOLD&lt;sub&gt;DS30&lt;/sub&gt;</td>
<td>&lt;9</td>
<td></td>
<td>30–36 mo</td>
<td>75</td>
<td>SICD</td>
<td>89 (68–100)</td>
<td>86 (78–95)</td>
<td>12</td>
<td>47</td>
<td>98</td>
<td>6.5</td>
<td>0.13</td>
</tr>
</tbody>
</table>

<sup>a</sup>PPV = Positive Predictive Value, NPV = Negative Predictive Value, PLR = Positive Likelihood Ratio, NLR = Negative Likelihood Ratio.
A small group format by researcher-trained staff to toddlers or preschoolers was included in the latter study. Both trials reported significant improvement on measures of language skills.

Four trials tested individual treatment to children by research staff or speech-language pathologists. One examined the effects of providing young children (18 to 42 months) with language or phonological delays with access to usual speech-language therapy services in the community. With an average of only 6.2 hours of therapy over 12 months, children showed small but significant gains in receptive, but not expressive, language relative to controls.

Another trial involving 4-year-olds with specific language impairments tested a manualized intervention that addressed individualized language goals, phonological awareness, and letter knowledge. The intervention had no significant effect on expressive, receptive, or pragmatic language.

A third trial tested the effects of a strategy called recasting (repeating what is said by a child, with correct articulation or with a grammatical expansion of the child's utterance). The intervention had no overall effect on children's mean length of utterances but did produce improvements among children with the lowest baseline articulation skills.

The fourth trial tested whether an individualized treatment of children with speech sound disorders affected mean length of utterance but found no significant language effect. Finally, preschoolers with language impairments who played with age-appropriate language peers in the house play area of their classroom over a 3-week period showed significant improvement on activity-specific expressive language skills.

<table>
<thead>
<tr>
<th>Instrument and Component</th>
<th>Decision Cutoff Point</th>
<th>Reference</th>
<th>USPSTF Quality Rating</th>
<th>Child Age</th>
<th>n</th>
<th>Reference Instrument</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>Prevalence a</th>
<th>PPVb PPVc</th>
<th>NPVb NPVc</th>
<th>PLRa NLRb</th>
</tr>
</thead>
<tbody>
<tr>
<td>SKOLDB87</td>
<td>&lt;14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>37–42 mo 91 SiCD</td>
<td>88 (65–100) 96 (79–99)</td>
<td>10</td>
<td>6.0</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>SKOLDB83</td>
<td>&lt;14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>43–48 mo 54 SiCD</td>
<td>94 (84–100) 97 (94–99)</td>
<td>15</td>
<td>6.9</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Sentence Repetition</td>
<td>&lt;20th percentile</td>
<td>Sturmer et al 1989</td>
<td>Fair</td>
<td>54–96 mo 323 AAPS-R</td>
<td>57 (45–68) 95 (93–98)</td>
<td>19i</td>
<td>1.25 NR 0.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structured Screening Test</td>
<td>&lt;10</td>
<td>Laing et al 2002</td>
<td>Good</td>
<td>282 RDLS</td>
<td>62 (45–78) 91 (87–94)</td>
<td>11i</td>
<td>6.6 NR 0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hackney Early Language</td>
<td>≤10</td>
<td>Law 1994</td>
<td>Good</td>
<td>30 mo 189 RDLS</td>
<td>98 (94–100) 93 (91–99)</td>
<td>26</td>
<td>53 99 3.17 0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial Speech Screening Test</td>
<td>&lt;12 elements</td>
<td>Rigby and Chesham 1981</td>
<td>Fair</td>
<td>54 mo 438 SLP evaluation of Remfrew, RDLS, Edinburgh Articulation</td>
<td>80 (68–92) 93 (91–99)</td>
<td>10</td>
<td>58 98 12.1 0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AAP, Arizona Articulation Proficiency Scale; Revised; FLRST, Fluharty Preschool Speech and Language Screening Test; GFTA, Goldman-Fristoe Test of Articulation; ITPA, Illinois Test of Psycholinguistic Abilities; MSCA, McCarthy Scales of Children's Abilities; NR, not reported; PLS, Preschool Language Scale; RDLS, Reynell Developmental Language Scales; SiCD, Sequenced Inventory of Communication Development; SKOLDB, Screening Kit of Language Development; SLP, speech language pathologist; TACL-R, Test for Auditory Comprehension of Language – Revised; TD, Templin-Darley Tests of Articulation Consonant Single Subtest; TOLD-P, Test of Language Development Primary.

a Calculated by EPC authors unless otherwise noted that study investigators provided data. Prevalence values were not estimated or weighted to reflect sampling.
b Predictive values may be questionable for studies in which prevalence exceeded 10%; the problem arises when investigators choose a random sample of children with negative screens to complete the reference measures.
c Study investigators provided data.
d Could not calculate because of lack of data in article.
e Calculated as infinity.
Eight trials reported on various speech sounds\textsuperscript{54,58,59,63,65–68} (Supplemental Table 6). Of 2 trials of parent-mediated interventions, 1 found that a modified Hanen Parent Program had significant effects on consonant inventory and syllable structure\textsuperscript{67}. In the other trial, parents engaged the child in activities directed at discrimination of sounds.\textsuperscript{63} Children in the control condition improved more in auditory discrimination in the presence of background noise than experimental subjects.

A small group intervention for toddlers significantly improved the percentage of intelligible utterances for treated children.\textsuperscript{68} Two studies examined individual treatment by speech-language pathologists. One examined the effects of the “cycles” approach to phonological therapy (wherein rule-based errors in the child’s speech sound production are treated through recursive cycles of therapy) for preschoolers with severe phonological disorders; the intervention produced significant effects on standardized tests and percentage of correct consonants from a speech sample.\textsuperscript{66} The other study found no improvement in phonology error rate for children randomized to usual community speech-language pathology services for a year; however, treated children were 2.7 times less likely to exhibit the severity of speech sound problems used as a criterion for initial study eligibility.\textsuperscript{65}

The recasting trial found no main effects on children’s intelligibility, but did find improvements among children with the lowest baseline articulation skills.\textsuperscript{58}

Two studies reported that their interventions significantly improved phonological awareness skills in preschoolers. In one, teaching assistants delivered small group and individual lessons\textsuperscript{54}; in the other, language assistants provided individual home-based interventions.\textsuperscript{59}

**Fluency**

Two trials examined the Lidcombe Program of Early Stuttering intervention.\textsuperscript{69} Both significantly reduced stuttering in preschoolers, when delivered in a clinic setting\textsuperscript{55} and when using a telephone-based health delivery model.\textsuperscript{56}

**KQ6: Treatment: Outcomes Other Than Speech and Language**

Two trials examined effects on socialization. One, among children receiving community-based speech-language services, produced no significant effect.\textsuperscript{65} The other, among language-delayed toddlers receiving small-group therapy, produced large and significant differences favoring the treated children.\textsuperscript{68}

For reducing behavior problems, one trial tested the effectiveness of a low-intensity parent group program\textsuperscript{57} and another an in-home individualized program provided by a language assistant\textsuperscript{59}; neither found significant effects. Similarly, measures of well-being of toddlers\textsuperscript{65} and health-related quality of life of preschoolers\textsuperscript{59} yielded nonsignificant effects of treatment.

In 2 trials, toddlers randomized to speech-language services were no different from controls on attention level or play.\textsuperscript{65,68} Parents of language-delayed toddlers participating in small-group language therapy reported significantly greater improvements in parental stress.\textsuperscript{68}

Two trials measuring emergent literacy skills among preschoolers\textsuperscript{54,59} found that letter knowledge improved significantly, but one failed to find a significant effect for a broader construct of literacy.\textsuperscript{54} However, treatment did significantly improve a measure of reading comprehension administered at 6-months of follow-up.

**KQ7. Adverse Effects of Treatments**

Three studies examined potential adverse effects of interventions but reported no negative impacts on children or parents.\textsuperscript{59,65,68}

**DISCUSSION**

**Screening Accuracy**

Some screening instruments accurately identify children for language delays or disorders. As in the 2006 review, however, we observed wide ranges of reported sensitivity and specificity; no one instrument clearly demonstrated the best characteristics or 1 age as optimal for screening. We compared findings from the same instrument in different populations; specifically, accuracy of 3 parent-rated screeners (ASQ, CDI, and ITC) and 2 trained-examiner screeners (Fluharty Preschool Speech and Language Screening Test and SKOLD) across ages. CDI, ITC, and SKOLD displayed consistency and acceptable levels of sensitivity and specificity (≥70%)\textsuperscript{70} at each age level; this suggests that they are more robust across different ages than ASQ and Fluharty Preschool Speech and Language Screening Test, which had generally low sensitivity across age levels.

Accuracy apparently drops over time. In the 2 studies\textsuperscript{32,33,42,43} that examined whether a parent-report screener administered at 2 years would be as accurate at 3 years, sensitivity was lower in 1 study and specificity was lower in the other. Decreasing specificity with time may mean some that some children with language delays will “catch up” and display more typical language skills as they age.\textsuperscript{71}

The comparison between parent-rated and trained-examiner screeners indicated many similarities in performance characteristics. Aside from the Denver Developmental Screening Test (now known as the Denver II), most trained-examiner tools are not used in primary care.
offices and would require a dedicated, trained examiner to test the child directly. Three parent-rated screeners (CDI, ITC, and LDS) display acceptable sensitivity and specificity. Moreover, because parents complete these screeners, adopting them in a screening program would not burden a primary care practice with training someone in test administration. The more extensive information that parents provide related specifically to their children's language skills may help explain their greater accuracy in identifying children with speech and language delays than broad-based screeners that include other domains but fewer speech and language items. Moreover, staff in primary care settings could likely interpret results from parent screeners with little difficulty.

Treatment Outcomes
The majority of the 13 trials support the effectiveness of treating young children with language delays and disorders (6 of 11 trials reporting significant positive results) and those with problems with speech sounds (6 of 8 trials reporting significant positive results), and toddlers and preschoolers for fluency problems (2 of 2 trials reporting significant positive results). Individual and small-group service delivery models and various intervention agents, including parents supported or trained by professionals, speech-language pathologists, and trained teaching or therapy assistants, generally favored intervention groups.

Multiple factors limit the confident interpretation of this body of evidence on speech and language treatment. These factors involve (1) the small size of many trials, which constrains investigating moderators and mediators of treatment effectiveness; (2) the lack of replicated positive findings for any treatment approach except the Lidcombe program for stuttering; (3) the wide variability across trials in the age of children treated, intervention agents (eg, speech-language pathologists, teaching assistants, parents, research staff), intensity, content, and strategies; (4) the relatively small number of trials using manualized treatments or providing enough details of the treatment to permit replication; (5) a corresponding lack of data detailing treatment fidelity in many trials; (6) a lack of common outcome measures; and (7) inconsistency in how results are reported. Because of this degree of heterogeneity, we could not do any meta-analysis. Overall, the evidence offers little guidance about specific factors associated with effective treatments for young children with speech and language delays.

Contextual Issues: Risk Factors
One contextual issue involved whether consistent, reliable, and valid risk factors exist that clinicians could use to identify children at highest risk for speech and language delay and disorders.19 We examined 31 cohort studies, 24 with multivariate analysis to control for other factors, and 1 review of studies on characteristics of late-talking toddlers; 20 cohort studies involved English-speaking children (Supplemental Tables 7 and 8). Potential risk factors for speech and language problems include male gender, family history of speech or language impairment, lower levels of parental education, and various perinatal risk factors (eg, prematurity, birth difficulties, and low birth weight). Studies about risk factors varied in the type of delay or disorder being considered, used inconsistent measurement of risk factors, included heterogeneous patient populations, and inconsistently adjusted for confounders in multivariate models. Future research should account for the heterogeneity across populations of children, consider a multifactorial perspective of child development, examine social determinants of health as possible risk factors, and adopt more standardized outcome measures over a longer-term period of follow-up than has been customary to date.

Limitations of the Review
Numerous limitations of the literature base continue to plague the field. Some date to the 2006 review, but additional limitations we encountered further reduce the applicability of the findings.

Most serious is the lack of well-designed, well-conducted studies addressing the overarching question: does screening for speech and language delay or disorders improve outcomes? Moreover, neither the 2006 review nor our update found any studies that addressed the questions of adverse effects of screening or the role of enhanced surveillance by primary care clinicians in accurately identifying children for diagnostic evaluations and interventions, 2 important issues in screening. We identified some instruments that can accurately screen children with speech and language delays. However, many studies included potentially inappropriate populations, such as “samples” of children identified (randomly or otherwise) on the basis of their language status. Using such “predetermined” samples hampers investigators from determining certain accuracy statistics (other than sensitivity and specificity) and may bias conclusions about screening accuracy and, thus, can limit applicability to pediatric populations in general. Moreover, few studies examined how well screeners detected speech and language disorders over the long term. Such studies are critical in calculating the real benefit of early detection. In addition, few of the screening accuracy studies occurred in primary care settings, and none in the United States. The extent to which conclusions reached from screening in primary care settings in Sweden, Australia, and the United Kingdom are generalizable to the United States is not known. Most treatment studies were also conducted outside the United States.
Whether conclusions reached from trials in countries with different medical, health insurance, and educational systems apply in this country remains an open question. Additional limitations relate to interpreting treatment outcomes and replicating interventions. Much of the literature lacks information about important features of the intervention, such as whether children received community services for speech and language outside the study, and does not adequately document intervention models. Finally, control groups in numerous trials were children offered intervention on a delayed schedule. This condition likely would make parents more willing to consent to enrolling their children in a RCT, but it constrains our ability to look at long-range outcomes for treated versus untreated children.

**Future Research Needs**

To determine whether screening for speech and language delay or disorders improves speech, language, or other outcomes, studies need to be specifically designed and executed to examine these issues. Furthermore, they need to be implemented with little risk of bias. This research gap presents an opportunity for a large study in primary care settings to test the efficacy of systematic routine screening for speech and language delays and disorders in comparison with not implementing routine screening. In tandem with this, the field would benefit from a study to examine the feasibility of speech and language-specific screening as part of the more general developmental screening that is already recommended.18

Given federal mandates under the Individuals with Disabilities Education Act that all children with a documented speech or language delay receive early intervention, conducting RCTs to examine the efficacy of interventions may be difficult in future. Protocols may adopt rigorous quasi-experimental designs, such as regression discontinuity designs, to answer intervention questions. Well-designed and implemented regression discontinuity designs meet standards for rigor for evaluations of evidence sponsored by the Institute of Education Sciences.

We recommend that stakeholders with an interest in screening develop research agendas and funding targeted to answer the important questions that we could not address. Future systematic reviews will benefit from an enhanced literature base.

**CONCLUSIONS**

We found no evidence to answer the overarching question of whether screening for speech and language delay or disorders improves speech and language outcomes. Studies from the 2006 review and our newly identified studies suggest that some screening instruments can accurately pinpoint these disorders. Although the parent-rated instruments require only that the primary care provider interpret the findings, studies have not examined this in practice. As in the 2006 review, we found no studies that addressed the harms of screening for speech and language delays. Neither did we find any evidence about the role of enhanced surveillance by a primary care clinician once a child elicits clinical concern for speech and language delay. Building on the 2006 review, we found evidence supporting the effectiveness of treating speech and language delays and disorders in children. Nevertheless, the whole body of evidence does not provide guidance regarding specific factors associated with effective treatments for young children with speech and language delays or disorders. Finally, we found no evidence relating to the harms of treating speech and language delays or disorders.

**ACKNOWLEDGMENTS**

This work was supported by the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services contract no. HHS290201200015I02. The views expressed in this manuscript do not represent and should not be construed to represent a determination or policy of the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services.

The authors gratefully acknowledge the following individuals for their contributions to this project and deeply appreciate their considerable support, commitment, and contributions: our AHRQ Medical Officer, Karen C. Lee, MD, MPH; the USPSTF leads; Evidence-based Practice Center (EPC) Project Manager, Carol Woodell, BSPH; RTI-UNC EPC Director, Meera Viswanathan, PhD; our EPC Librarian, B Lynn Whitener, MLSL; and our EPC publications specialist, Loraine Monroe.

**FINANCIAL DISCLOSURE:** The authors have indicated they have no financial relationships relevant to this article to disclose.

**FUNDING:** Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services Contract No. HHS290201200015I02. The views expressed in this manuscript do not represent and should not be construed to represent a determination or policy of the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services.

**POTENTIAL CONFLICT OF INTEREST:** The authors have indicated they have no potential conflicts of interest to disclose.

**COMPANION PAPERS:** Companions to this article can be found on pages e474 and e494, online at www.pediatrics.org/cgi/doi/10.1542/peds.2015-1711 and www.pediatrics.org/cgi/doi/10.1542/peds.2015-0211.
REFERENCES


27. Frisk V, Montgomery L, Boychyn E, et al. Why screening Canadian preschoolers...
for language delays is more difficult than it should be. Infants Young Child. 2009;22(4):290–308


57. Wake M, Tobin S, Girolametto L, et al. Outcomes of population based language promotion for slow to talk toddlers at ages 2 and 3 years: Let's Learn Language cluster randomised controlled trial. BMJ. 2011;343:d4741


Screening for Speech and Language Delay in Children 5 Years Old and Younger: A Systematic Review
Ina F. Wallace, Nancy D. Berkman, Linda R. Watson, Tamera Coyne-Beasley, Charles T. Wood, Katherine Cullen and Kathleen N. Lohr
Pediatrics 2015;136;e448; originally published online July 7, 2015;
DOI: 10.1542/peds.2014-3889

Updated Information & Services
including high resolution figures, can be found at:
/content/136/2/e448.full.html

Supplementary Material
Supplementary material can be found at:
/content/suppl/2015/07/02/peds.2014-3889.DCSupplemental.html

References
This article cites 100 articles, 19 of which can be accessed free at:
/content/136/2/e448.full.html#ref-list-1

Subspecialty Collections
This article, along with others on similar topics, appears in the following collection(s):
Developmental/Behavioral Pediatrics
/cgi/collection/development:behavioral_issues_sub
Cognition/Language/Learning Disorders
/cgi/collection/cognition:language:learning_disorders_sub

Permissions & Licensing
Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
/site/misc/Permissions.xhtml

Reprints
Information about ordering reprints can be found online:
/site/misc/reprints.xhtml
Screening for Speech and Language Delay in Children 5 Years Old and Younger: A Systematic Review
Ina F. Wallace, Nancy D. Berkman, Linda R. Watson, Tamera Coyne-Beasley, Charles T. Wood, Katherine Cullen and Kathleen N. Lohr
*Pediatrics* 2015;136;e448; originally published online July 7, 2015; DOI: 10.1542/peds.2014-3889

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/136/2/e448.full.html