Identification of Developmental-Behavioral Problems in Primary Care: A Systematic Review

abstract

CONTEXT: Recent mandates and recommendations for formal screening programs are based on the claim that pediatric care providers underidentify children with developmental-behavioral disorders, yet the research to support this claim has not been systematically reviewed.

OBJECTIVE: To review research literature for studies regarding pediatric primary care providers’ identification of developmental-behavioral problems in children.

METHODS: On the basis of a Medline search conducted on September 22, 2010, using relevant key words, we identified 539 articles for review. We included studies that (1) were conducted in the United States, (2) were published in peer-reviewed journals, (3) included data that addressed pediatric care providers’ identification of developmental-behavioral problems in individual patients, (4) included an independent assessment of patients’ developmental-behavioral problems, such as diagnostic interviews or validated screening instruments, and (5) reported data sufficient to calculate sensitivity and specificity. Studies were not limited by sample size. Eleven articles met these criteria. We used Quality Assessment of Diagnostic Accuracy Studies (QUADAS) criteria to evaluate study quality. Although the studies were similar in many ways, heterogeneous methodology precluded a meta-analysis.

RESULTS: Sensitivities for pediatric care providers ranged from 14% to 54%, and specificities ranged from 69% to 100%. The authors of 1 outlier study reported a sensitivity of 85% and a specificity of 61%.

CONCLUSIONS: Pediatricians are often the first point of entry into developmental and mental health systems. Knowing their accuracy in identifying children with developmental-behavioral disabilities is essential for implementing optimal evaluation programs and achieving timely identification. Moreover, these statistics are important to consider when planning large-scale screening programs. Pediatrics 2011; 128:356–363

AUTHORS: R. Christopher Sheldrick, PhD, Shela Merchant, MA, and Ellen C. Perrin, MD

Department of Pediatrics, Floating Hospital, Tufts Medical Center, Boston, Massachusetts

KEY WORDS

pediatric, screening, behavior disorders/problems, developmental disabilities, systematic reviews

ABBREVIATIONS

PPV—positive predictive value
NPV—negative predictive value
QUADAS—Quality Assessment of Diagnostic Accuracy Studies

All the authors made substantive contribution to the study and/or manuscript and reviewed the final paper before its submission.

www.pediatrics.org/cgi/doi/10.1542/peds.2010-3261
doi:10.1542/peds.2010-3261

Accepted for publication Apr 20, 2011

Address correspondence to R. Christopher Sheldrick, PhD, Department of Pediatrics, Floating Hospital, Tufts Medical Center, 800 Washington St, Box 854, Boston, MA 02111. E-mail: rsheldrick@tuftsmedicalcenter.org

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2011 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.
Estimates indicate that at least 1 in 5 children has a developmental and/or behavioral disability.\textsuperscript{1,2} Several recent recommendations demonstrate the growing consensus that early identification is essential for providing adequate treatment to children with such disabilities.\textsuperscript{3,4} To facilitate early identification, the American Academy of Pediatrics Council on Children With Disabilities has recommended that pediatricians and other child health care providers perform ongoing developmental surveillance during all routine health supervision visits, supplemented with standardized screening instruments at specified ages.\textsuperscript{5} More recently, several states have instituted programs that encourage or even require all child health providers to administer screening instruments at well-child visits.\textsuperscript{6,7}

Such mandates and recommendations are based on the claim that pediatric care providers underidentify children with developmental-behavioral disorders. A systematic review of the evidence for this claim is critical for several reasons. First, to evaluate the utility of any new screening program or recommendation, its effectiveness must be compared with the clinical accuracy of standard pediatric practice that does not include validated screening instruments. Second, most claims about pediatric providers’ accuracy focus on their sensitivity and largely ignore their specificity. Hence there is a focus on correctly identifying children who have disorders while ignoring whether the providers correctly identify children without disorders. For a full understanding of clinicians’ accuracy in identifying developmental and behavioral problems, both sensitivity and specificity are essential. Third, for many families, especially those with young children, pediatric care providers serve as gatekeepers to mental health and developmental services.\textsuperscript{7} If providers are, in fact, underidentifying children with developmental-behavioral disorders, finding feasible methods for improving identification is essential for effective treatment.

We have systematically reviewed the research literature for evidence regarding pediatric primary care providers’ identification of developmental and behavioral problems in children. We make a distinction between developmental and behavioral disorders because most professional recommendations,\textsuperscript{3} screening instruments, and diagnostic tests follow this dichotomy and focus on 1 or the other but not both. We recognize that developmental and behavioral disorders (1) cause significant impairment in children and deserve meticulous attention in the pediatric care setting, (2) are not distinct categories (some disorders, such as attention-deficit/hyperactivity disorder, are alternately classified in either category)\textsuperscript{6-10} and (3) are often present in the same child (research has suggested that 75%–80% of children with an autism spectrum disorder also have a comorbid behavioral disorder).\textsuperscript{11,12} Nevertheless, most research studies focus on either behavioral or developmental problems, so we review each as an independent category.

**METHOD**

To identify articles for this literature review, we conducted a search of the Medline database on September 22, 2010. There was no preexisting protocol for this type of search, so we devised our own method and included search terms particular to the topic at hand.

**Criteria for Inclusion**

We included studies that (1) were conducted in the United States, (2) were published in a peer-reviewed journal, (3) included data regarding whether pediatric care providers identified a developmental or behavioral problem in individual patients, (4) included an independent assessment of patients’ developmental-behavioral problems, such as diagnostic interviews or validated screening instruments, and (5) reported data for a sufficient number of cases to determine sensitivity and specificity. Studies were not limited by sample size.

**Search Strategy**

We began by accessing all the articles in which an assessment of childhood developmental or behavioral disorders in pediatric settings was described. Specifically, we searched for the intersection (ie, using the “and” term) of articles in the following groups:

- articles with subject headings that included “developmental disabilities,” “delay,” “mental health,” “mental disorders,” “child behavior disorders,” “language development disorders,” “depression,” “anxiety,” or “autistic disorders” (note that these terms were combined with an “or” term);
- articles with the subject heading “pediatrics” or “pediatric providers”;
- articles with subject headings “diagnosis,” “identification,” “screening,” or “surveillance” (note that these terms were combined with an “or” term).

Because research on this topic is sparse and began only in the 1980s, we did not set a date limit on the studies we searched. The final yield was 539 articles.

We included 2 additional search strategies. First, we reviewed the reference section for additional citations in any article that focused on screening or surveillance. We included this additional step because some studies that met inclusion criteria were conducted...
for a purpose other than to assess the accuracy of pediatric care providers (eg, to validate a screening instrument). Second, for all articles that met inclusion criteria, we used the “cited by” function in Medline to identify articles that refer to studies that met our inclusion criteria. These 2 additional search strategies yielded 174 new articles for review.

Two of us (Dr Sheldrick and Ms Merchant) conducted the literature review. Studies with titles or abstracts that referred to pediatric care providers’ recognition of developmental or behavioral problems were reviewed in greater detail. Examples of excluded studies are those that focused on pediatric issues not relating to child behavior and development, such as screening for parent psychopathology; neurologic bases of psychiatric disorders; treatment of psychiatric disorders; studies not conducted in the United States; and general summaries of common developmental and behavioral problems seen in the pediatric population. For the remaining articles, we read the abstracts, introductions, and results sections to determine if articles (1) met the inclusion criteria described earlier or (2) included references to other articles that might meet inclusion criteria.

Dr Sheldrick and Ms Merchant reviewed the final set of studies for data from which sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) could be calculated. In the case of multiple studies that reported on the same data set, we included only the earliest study in our analyses. In the case of studies that seemed to meet inclusion criteria (ie, the methods section described a procedure for eliciting pediatricians’ assessment of developmental or behavioral problems as well as a diagnostic interview) but from which such statistics were not reported, the author was contacted to determine if any such analyses were conducted. Two authors were contacted. In 1 case, the author reported that analyses relevant to our study had not been performed; thus the study was not included in our analyses. In the second case, the author referred us to another study that used the same data set as a study that was already included in our sample.

Assessment of Clinical Accuracy

Studie s’ Methodology

To determine if the studies in our final sample were comparable, we assessed all final studies that met inclusion criteria by using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) criteria. QUADAS provides a framework for evaluating the quality of each study’s methodology and conclusions. It helps to determine if the quality of each study is high enough to have reliable and valid conclusions and whether information across different studies can be integrated.

Clinicians’ Accuracy

Sensitivity, specificity, PPV, NPV, and a diagnostic odds ratio were calculated for each study along with their 95% confidence intervals. For studies with weighted data, true-positive, false-negative, true-negative, and false-positive results were estimated on the basis of the size of the final sample that completed the gold-standard reference test, the sensitivity and specificity reported, and the weighted prevalence and identification rate. For unweighted data, these numbers were recorded verbatim from the original article or estimated directly from reported results. In addition to conducting assessments with parents, some authors also reported data from assessments directly with children. Because only some articles reported such information, we abstracted data only from parent interviews and screening instruments for all studies.

RESULTS

Of the 713 articles identified by the electronic search strategy, 445 were eliminated because of duplications and/or because they did not focus on pediatricians’ recognition of developmental-behavioral problems. Detailed review of the remaining 268 articles yielded 11 articles that met the criteria outlined earlier such that we were able to derive indices of clinicians’ accuracy (see Fig 1).

There were many consistent findings with regard to study quality. For all studies, the time period between the pediatric assessment and the criterion (ie, gold-standard) test was reasonable; all participants in each study received the same criterion test; criterion tests were conducted regardless of the results of the pediatric examination; and pediatric examinations and criterion tests were conducted independently. Withdrawals were generally well explained with 1 exception, in which the 56% enrollment rate was not explained.

In other ways, studies varied greatly in quality and design with respect to the purpose of our review. No accepted summary of study quality is available, so relevant study details and QUADAS criteria are instead listed individually in Table 1. For example, 5 studies included only screening instruments as their criterion tests. Because 4 of these 5 studies had large sample sizes, and all 5 of them administered the criterion test to all participants, we deemed their inclusion worthwhile. The remaining 6 studies relied on diagnostic interviews, but the specific interviews used varied. For 10 studies, data regarding whether pediatric providers identified a developmental-behavioral problem derived from brief, provider-completed question-
naires; 1 study relied on an extensive record review. Although most study samples were representative of pediatric populations, 2 were drawn from ongoing studies with high-risk samples. Studies also varied greatly in the proportion of the study population that were successfully enrolled in the study (range: 36%–91%), which indicates different levels of external validity. Some studies focused only on young children, others focused on adolescents, and yet others focused on more comprehensive age ranges.

Nine studies assessed behavioral problems. Only 2 studies assessed developmental problems, 1 of which relied on a screening instrument as a criterion test. No study we identified included diagnostic tests for autism or pervasive developmental disorder. Because of the high degree of heterogeneity in study methodology and quality, we determined that meta-analytic statistics were inappropriate. Therefore, we list descriptive statistics for each study in Table 1 and show each study’s sensitivity and specificity (with 95% confidence intervals) in Fig 2. One study seems to be an outlier. Brown and Wissow reported that physicians identified 48.6% of their patients as having a behavior problem, which yielded a sensitivity of 0.85 and a specificity of 0.61 when compared with a screening instrument that identified 21.5% of the sample as being at risk. For the remaining studies, sensitivity ranged from 14% to 54% and specificity ranged from 69% to 100%. PPVs ranged from 24% to 66%, and NPVs ranged from 61% to 94%. However, because PPV rises and NPV falls as prevalence increases (all else being equal), these statistics should be interpreted in light of the prevalence rates according to the gold-standard diagnostic or screening test.

**DISCUSSION**

The American Academy of Pediatrics recommends that developmental screening instruments maintain sensitivity and specificity rates higher than 70%. Among the studies examined, pediatric care providers who worked without screening instruments achieved specificity (ie, the proportion of children correctly identified as not having a problem) that was consistently near or higher than 70%. The sensitivity of pediatric care providers was also consistent among the studies but in the opposite direction: in all cases except 1 it was lower than 54%. Thus, the number of children correctly identified as having a developmental or behavioral problem was quite low.

The lack of data on identification of developmental disorders was notable. Nine studies in our sample examined the identification of behavioral problems only, 1 focused on language delays, and 1 focused on a range of developmental problems. This distinction is important, because pediatricians’ accuracy might vary according to the type of problem in question. We were unable to identify any study that directly compared pediatric care providers to a clinically determined diagnosis of a developmental disorder. Thus, although diagnoses of autism or pervasive developmental disorder are often reported to be delayed, we have no direct evidence regarding the role of pediatric care providers in causing this delay. Moreover, no study assessed both developmental and behavioral disorders in the same children, which makes it impossible to determine if identification of 1 type of problem influences detection of the other.

We note several limitations to this study. The major limitation to improved understanding of pediatricians’ accuracy in identifying developmental and behavioral disorders stems from the paucity of research that focuses on this topic (especially identification of developmental delays). In many cases, the primary purpose of studies we included was not to determine if pediatricians accurately
<table>
<thead>
<tr>
<th>Study</th>
<th>Construct Assessed</th>
<th>Criterion Test</th>
<th>Pediatric Care Provider</th>
<th>Sample</th>
<th>Additional QUADAS Concern Noted</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
<th>DOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costello et al (1988)</td>
<td>Behavioral problems</td>
<td>DISC</td>
<td>Physician questionnaire</td>
<td>11.7</td>
<td>None</td>
<td>0.17 (0.08–0.35)</td>
<td>0.96 (0.93–0.98)</td>
<td>0.36 (0.18–0.59)</td>
<td>0.90 (0.86–0.93)</td>
<td>4.9 (1.7–14.1)</td>
</tr>
<tr>
<td>Lavigne et al (1993)</td>
<td>Behavioral problems</td>
<td>Agreement between 2 diagnostic interviews</td>
<td>Physician questionnaire</td>
<td>12.9</td>
<td>Whole or random sample did not receive criterion test</td>
<td>0.21 (0.12–0.32)</td>
<td>0.95 (0.90–0.95)</td>
<td>0.31 (0.19–0.45)</td>
<td>0.89 (0.86–0.91)</td>
<td>3.5 (1.7–7.1)</td>
</tr>
<tr>
<td>Merritt et al (1993)</td>
<td>Behavioral problems</td>
<td>CAS</td>
<td>Physician questionnaire</td>
<td>4.15</td>
<td>None</td>
<td>0.29 (0.15–0.55)</td>
<td>0.79 (0.60–0.91)</td>
<td>0.50 (0.24–0.78)</td>
<td>0.61 (0.42–0.78)</td>
<td>1.6 (0.4–6.6)</td>
</tr>
<tr>
<td>Chang et al (1988)</td>
<td>Behavioral problems</td>
<td>SADS</td>
<td>Physician questionnaire</td>
<td>4.5</td>
<td>Sample not representative of pediatric primary care patients</td>
<td>0.43 (0.29–0.59)</td>
<td>0.69 (0.55–0.80)</td>
<td>0.52 (0.35–0.68)</td>
<td>0.61 (0.48–0.73)</td>
<td>1.7 (0.7–4.1)</td>
</tr>
<tr>
<td>Costello et al (1985)</td>
<td>Behavioral problems</td>
<td>DISC</td>
<td>Physician questionnaire</td>
<td>7.0</td>
<td>Inclusion criteria unclear</td>
<td>0.14 (0.06–0.51)</td>
<td>1.00 (0.76–1.00)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>King et al (2005)</td>
<td>Language delay</td>
<td>Preschool Language Scale-3rd edition</td>
<td>Medical record review</td>
<td>10.1</td>
<td>Sample not representative of pediatric primary care patients</td>
<td>0.53 (0.22–0.46)</td>
<td>0.82 (0.78–0.85)</td>
<td>0.17 (0.11–0.26)</td>
<td>0.92 (0.88–0.94)</td>
<td>2.2 (1.2–4.1)</td>
</tr>
<tr>
<td>Keilhauer et al (1997)</td>
<td>Behavioral problems</td>
<td>PSC-35</td>
<td>Physician questionnaire</td>
<td>1.22</td>
<td></td>
<td>0.54 (0.51–0.56)</td>
<td>0.87 (0.86–0.88)</td>
<td>0.37 (0.35–0.39)</td>
<td>0.63 (0.92–0.94)</td>
<td>7.9 (6.9–9.0)</td>
</tr>
<tr>
<td>Leaf et al (2004)</td>
<td>Behavioral problems</td>
<td>CBCL</td>
<td>Physician questionnaire</td>
<td>1.56</td>
<td></td>
<td>0.50 (0.44–0.57)</td>
<td>0.76 (0.73–0.78)</td>
<td>0.24 (0.21–0.29)</td>
<td>0.91 (0.89–0.92)</td>
<td>3.1 (2.5–4.2)</td>
</tr>
<tr>
<td>Murphy et al (1992)</td>
<td>Behavioral problems</td>
<td>PSC-35</td>
<td>Physician questionnaire</td>
<td>2.11</td>
<td></td>
<td>0.50 (0.22–0.79)</td>
<td>0.90 (0.74–0.97)</td>
<td>0.57 (0.25–0.84)</td>
<td>0.87 (0.71–0.95)</td>
<td>9.0 (1.5–56.1)</td>
</tr>
<tr>
<td>Brown and Wissow (2010)</td>
<td>Behavioral problems</td>
<td>SDQ</td>
<td>Physician questionnaire</td>
<td>2.15</td>
<td></td>
<td>0.85 (0.79–0.90)</td>
<td>0.61 (0.57–0.65)</td>
<td>0.38 (0.35–0.43)</td>
<td>0.94 (0.91–0.96)</td>
<td>8.9 (5.6–14.0)</td>
</tr>
<tr>
<td>Hix-Small et al (2007)</td>
<td>Developmental problems</td>
<td>ASQ</td>
<td>Physician questionnaire</td>
<td>2.16</td>
<td></td>
<td>0.52 (0.26–0.40)</td>
<td>0.95 (0.83–0.97)</td>
<td>0.66 (0.55–0.76)</td>
<td>0.84 (0.81–0.86)</td>
<td>10.1 (5.9–17.0)</td>
</tr>
</tbody>
</table>

ID indicates identification; DOR, diagnostic odds ratio; CI, confidence interval; DISC, Diagnostic Interview Schedule for Children; CAS, Child Assessment Schedule; SKDS, Schedule for Affective Disorders and Schizophrenia for School-Aged Children; NA, not applicable/not available; PSC-35, Pediatric Symptom Checklist (raw score ≥ 28); 35-item version; CBCL, Child Behavior Checklist (T score ≥ 64); SDQ, Strengths & Difficulties Questionnaire (high symptoms + impairment); ASQ, Ages & Stages Questionnaire (18- and 24-month forms).

a Sample that received criterion test; total sample size was larger.
b Oversampled children with high CBCL scores; children with low CBCL scores were matched, not randomly sampled.
c Sampled from study of depressed mothers and matched controls (Yale–National Institute of Mental Health Collaborative Family Study).
d For any mention of concern regarding developmental delay.
e High-risk sample.
identified children’s developmental and behavioral disorders. In fact, 1 study that collected data that could have been used to address this question did not analyze it (S. M. Horwitz, PhD, Department of Pediatrics, Stanford University, written correspondence, June 5, 2009). Six studies used a structured diagnostic interview as the criterion, and 5 used validated parent-report screening instruments. Given that the reliability and validity of these criterion instruments vary, this information could affect the estimate of pediatricians’ accuracy as well. In addition, studies included in this review predominantly relied on parent report for criterion assessments, but there might be an advantage to including standardized examinations of the child (eg, the Bayley scales) to assess development. Moreover, cutoff scores for various screening instruments also differ; some are set with the aim of maximizing accuracy with respect to diagnostic interviews, whereas others are set to conform to state-level policies. Results of studies that compared pediatric care providers to screening instruments should, therefore, be interpreted with caution. For reasons such as these, we were not able to identify enough articles with similar design, quality, and sample characteristics to support a meta-analysis; thus, interpretation must occur at the level of each individual study. Finally, our search strategy might have missed some studies that would otherwise have met our inclusion criteria. Because relevant articles may be published under diverse subject headings, identification of studies for inclusion was difficult.
Pediatric care providers have increasingly been encouraged to use standardized screening instruments to identify children with behavioral and developmental problems. The utility of screening instruments depends on how much they are able to improve on pediatricians’ standard care, which was the focus of our study. Many screening instruments have been reported to display higher sensitivity than the pediatricians in the studies sampled for the review. However, few screening instruments claim to improve on pediatricians’ specificity, and many fall far short.

Given what we know about pediatric care providers’ accuracy in identifying developmental and behavioral problems, it is important to consider what the downstream effects of implementing a screening program might be. Assuming that the screening instrument is truly effective, the proportion of true cases identified could be expected to rise. However, this rise would not occur without cost. Along with the rise in the number of true cases would come an increase in the number of “false-positives,” or patients who are incorrectly identified as having a condition. False-positive results require additional assessment, and if they occur in increasing numbers, it would have an effect on pediatricians’ time and the capacity of referral sources. These changes might also affect patient satisfaction. A positive screening result suggests the presence of disorder even if that suggestion is later reversed. For parents who have real concerns about their children despite subclinical symptoms, such attention might be welcome. However, other parents might still experience anxiety or stigma associated with false-positive results, as has been noted in other areas of medicine.

CONCLUSIONS

Before instituting a screening program, it is essential to consider how that program will change both sensitivity and specificity as well as the downstream effects of these changes. In this article we have summarized what is known about the accuracy of pediatricians’ judgments of children’s developmental and behavioral status in the absence of formal screening programs.

REFERENCES

19. Brown JD, Wissow LS. Screening to identify
THE VINES HAVE IT: Each spring and fall, I clear the grape vines from the trees that edge our property. If I don’t, the vines proliferate wildly and overwhelm the tree. Eventually the weight of the vines pulls down branches leading to ill health and even the death of some trees. The problem of robust vine growth does not seem confined to northern Vermont. In the Central and South American rain forests, woody vines appear to be winning their clash with the trees. As reported in The New York Times (Science: May 23, 2011), woody vines, such as liana, are increasing in tropical forests. Lianas are long-stemmed woody vines rooted in the soil that use trees to support their growth toward the sunlight in the forest canopy. Once the vine reaches the top, it produces a cornucopia of leaves. In several areas that have been monitored for the past few decades, the tops of three-fourths of all trees with trunks greater than eight inches in diameter are infested with lianas. While the vines may form bridges between trees that make it easier for arboreal animals to move about the forest, they compete very successfully with the trees for water, soil nutrients, and sunlight. Why the vines are so successful is not known but it may be that they are better able than trees to utilize atmospheric carbon dioxide and also to extract water from the soil, particularly in the dry season. Trees infested with liana have stunted growth and higher mortality rates. The sheer mass of the vines can also lead to mechanical failure and death of the tree. The long term effects of liana growth are worrisome. Trees are essential for storing terrestrial carbon. Because lianas store less carbon than trees a reduction in tree mass will result in a net reduction in carbon storage. Carbon storage is essential for regulating atmospheric carbon dioxide. Some experts suggest that the liana proliferation has led to a 10 percent net decrease in carbon storage and could have implications for global warming. As for me, while my tree line is not like a tropical forest, removing the vines is hard work. Still, I like my trees and their role in the environment so I will continue to battle the vines twice a year.