HHS Public Access

Author manuscript

Acad Pediatr. Author manuscript; available in PMC 2018 July 01.

Published in final edited form as:

Acad Pediatr. 2017 July ; 17(5): 464–470. doi:10.1016/j.acap.2017.01.016.

Is a Positive Developmental-Behavioral Screening Score Sufficient to Justify Referral? A Review of Evidence and Theory

R. Christopher Sheldrick, PhD and **Daryl Garfinkel, BA**

Developmental-Behavioral Pediatrics, Floating Hospital for Children, Tufts Medical Center, Boston, Mass

Abstract

In their recommendations on screening for autism and developmental disabilities, the American Academy of Pediatrics recommends referral subsequent to a positive screening result. In this article, we argue that positive screening results are not always sufficient to justify a referral. We show that although positive predictive values are often low, they actually overstate the probability of having a disorder for many children who screen positive. Moreover, recommended screening thresholds are seldom set to ensure that the benefits of referral will equal or exceed the costs and risk of harm, which is a necessary condition for an optimal threshold in decision analysis. Drawing on recent recommendations for the Institute of Medicine/National Academy of Medicine, we discuss the implications of this argument for pediatric policy, education, and practice. In particular, we recommend that screening policies be revised to ensure that the costs and benefits of actions recommended in the event of a positive screen are appropriate to the screening threshold. We recommend greater focus on clinical decision-making in the education of physicians, including shared decision-making with patients and their families. Finally, we recommend broadening the scope of screening research to encompass not only the accuracy of specific screening instruments, but also their ability to improve decision-making in the context of systems of care.

Keywords

autistic disorder; decision-making; developmental disabilities; mass screening; sensitivity and specificity

The council on Children with Disabilities and the Bright Futures Steering Committee of the American Academy of Pediatrics (AAP) recommends that pediatricians conduct developmental screening at well-child visits and address positive screening results by making referrals for “developmental and medical evaluations and early developmental intervention/early childhood services.”[1] Many other groups, including the Centers for Disease Control and Prevention,[2] base their recommendations for managing positive screening results on the AAP’s policy statement. In this article, we focus on evidence and theory supporting this recommendation.

Address correspondence to R. Christopher Sheldrick, PhD, 800 Washington St, Box 854, Boston, MA 02111 (rsheldrick@tuftsmedicalcenter.org).

The authors have no conflicts of interest to disclose.

On the surface, a recommendation to refer subsequent to a positive result on a developmental-behavioral screening instrument reflects sound judgment. A range of evidence suggests that developmental-behavioral disabilities (a term we use broadly to include developmental delays, autism, internalizing disorders, and externalizing disorders) are underidentified in pediatrics, with approximately a third of children with disabilities being identified in primary care settings.[3] In response, professional guidelines typically recommend use of screening instruments that detect at least 70% of children with developmental-behavioral disabilities (ie, sensitivity = 70%). If these assumptions are true and if all pediatricians began to refer all children who scored positive on valid screening instruments, then detection of developmental-behavioral disabilities at a sufficiently young age could be expected to increase dramatically (ie, from approximately a third of children with developmental-behavioral disabilities to at least 70%).

However, published evidence clearly shows that pediatricians typically do not refer all children who screen positive. A recent analysis of a systematic review identified 8 published studies on the implementation of developmental-behavioral screening programs that reported rates of referral subsequent to positive screens.[4] Among these studies, referral rates ranged from 10% to 86%.[5] Whereas results such as these are widely interpreted as failures to implement evidence-based screening protocols with fidelity, we suggest an alternative explanation on the basis of at least 2 limitations in the logic underlying many screening recommendations. First, pediatricians make clinical decisions about patients as individuals, one at a time—they do not make decisions for groups of patients. We therefore begin by showing how reliance on group-level statistics like positive predictive value (PPV) overstates the value of recommended screening thresholds (also known as “cut scores”) for making decisions about individual patients. Second, we argue that recommended screening thresholds typically fail to account for the expected costs and benefits of available referral options. Although pediatricians are clearly aware of the benefits of identifying and referring children with developmental-behavioral disabilities so that they can receive treatment, they are also acutely aware of the costs and potential for harm, such as those from false positive results. It is therefore reasonable to hypothesize that pediatricians often do not refer all children who score positive on screening instruments because they perceive the expected costs and harms resulting from referrals to outweigh the benefits, especially when the probability of disability is low. In short, it is reasonable that some children with positive screening results might not be referred. We present each of these arguments in greater detail and conclude by discussing the implications of our argument for pediatric research, education, and policy—including our suggestion that standard recommendations be revised.

Overstated Values

Group-level statistics like PPV overstate the value of recommended screening thresholds. Screening instruments are often evaluated according to the proportion of children with disabilities who score positive (ie, sensitivity), and the proportion of children without disabilities who score negative (ie, specificity). Many guidelines—including the AAP’s—recommend minimum values of 70%for sensitivity as well as specificity.[1,6] Statistics like sensitivity and specificity are extremely useful for addressing problems in public health because they set expectations for how a screening instrument will perform when used to

identify children with disabilities over an entire population. They are also often used to determine screening thresholds (ie, using the Youden index) to identify the threshold with the greatest combined value of sensitivity and specificity. However, their clinical utility is more limited.

When seeing patients, clinicians must interpret and act on screening results. To inform decisions about referrals for further evaluation or treatment, screening instruments would ideally tell pediatricians the chance that a given child truly has a developmental-behavioral disability. A statistic known as PPV is helpful in this regard. PPV is the probability of having a diagnosis with a positive score. Concretely, it is measured by analyzing all children who score positive and by determining the proportion of children who have a diagnosis. Because it is essentially an indicator of prevalence among the subgroup of children who score positive, PPV is highly sensitive to prevalence overall. If clinicians were to refer all children who scored positive, PPV would reflect how often their referral decisions would be correct using diagnoses as a criterion.

However, as a group-level statistic, PPV suffers from a significant limitation with respect to clinical decision-making: not every child who scores positive on a screening instrument has the same chance of having a disability. As clinicians have often observed, children with extreme scores are typically far more likely to have a disability than are children who score at or just beyond the threshold. Particularly for instruments with a single screening threshold, PPV overestimates the probability of having a diagnosis for some children who score positive and under-estimates that probability for others. If pediatricians were to refer all children who scored positive, PPV would offer valuable information about their overall performance (ie, the proportion of children referred who truly have disabilities). However, PPV is often not a useful indicator for individual cases.[7]

To understand the implications for screening policies, it is useful to consider decisions regarding children who score precisely at a screening threshold. For pediatricians to refer all children who screen positive, by definition they must refer all children who score at the threshold. If any positive score is sufficient to justify a referral, then a positive score at the threshold must also be sufficient to justify a referral. Unfortunately, for children who score at the screening threshold, the probability of having a disability is typically less than the PPV. For these children, what we want to know is the threshold probability (ie, the probability of having a developmental-behavioral disability for a child who scores at the screening threshold). Only if threshold probability is high enough to justify a referral should all children with positive screens be referred.

Figure 1 offers a way to visualize sensitivity, specificity, PPV, and threshold probability. Figure 1A includes 2 sets of scores—1 for a healthy population, and the other for a population affected by developmental-behavioral disabilities. Using these distributions of scores, Figure 1B shows how sensitivity and PPV are calculated. In both cases, the “true positives” in the lower right are those who are correctly classified and are shaded in light gray. Sensitivity represents the proportion of the population affected by disabilities who are correctly classified—reflected in this case by the fact that 77% of the curve below the horizontal axis is shaded in light gray. PPV represents the proportion of children who score

positive who are correctly classified— reflected in this case by the fact that 46% of the curves to the right of the screening threshold are shaded in light gray. Figure 1C offers a way to visualize the predictive value of individual scores. For each score, predictive value is equal to the frequency of the score in the affected population (represented by the length of the vertical gray) divided by the frequency of the score in the total population (represented by the total length of the vertical line). In Figure 1C, the probability at the threshold is 20% —much lower than the PPV of 46%. Whereas PPV is calculated on the basis of all children who score beyond the screening threshold, regardless of their individual scores (as reflected by areas under the appropriate curves), threshold probability applies only to children who score at the threshold itself (ie, reflected by the length of the lines).

Because PPV can be directly calculated on the basis of sensitivity, specificity, and prevalence, we can make inferences about the upper limits of PPV for any screening instrument for which these statistics are known.[8] Figure 2 shows PPV as a function of prevalence and screener accuracy (assuming sensitivity and specificity to be equal). For a positive screening result to be more likely to truly reflect a diagnosis than not, PPV must equal at least 50%.[9] It is therefore worth considering what is necessary to reach this level of performance by following the line labeled “PPV = 50%” in Figure 2. For a screener with minimally acceptable accuracy (ie, sensitivity = 70% and specificity = 70%), PPV would only increase to 50% if a screener was used in a population in which 30% of children had disabilities. If a clinician were working with a population of children in which only 15% of children had disabilities, a screener with much greater accuracy (eg, sensitivity = 85% and specificity = 85%) would be required to reach a PPV of 50%.

Whatever the value of PPV, threshold probabilities are unfortunately typically lower still.[7] A recent study that evaluated threshold probability for 2 well validated behavioral screening instruments (the Strengths and Difficulties Questionnaire and the Childhood Behavior Checklist) reported threshold probabilities ranging from 9% to 54%, despite high prevalence in 2 of the 3 samples considered (35% and 43% vs 14%).[7] These estimates of threshold probability were as low as a third of corresponding estimates of PPV. Because threshold probability is dependent on prevalence (just like PPV), the high base rates reported in the studies cited might overestimate the probability that children from a general population who score at the threshold would truly have developmental-behavioral disabilities.

Unaccounted Costs

Recommended screening thresholds often fail to account for the expected costs and benefits of available referral options. How high must threshold probability be to justify a referral? Decision analysis offers a clear response: A child should be referred if the probability of disability is high enough that the benefits of referral are likely to outweigh costs and potential for harm.[10] Because of uncertainty about the precise extent and severity of a child’s symptoms, it is often unclear whether a child’s problems truly reflect a disability that demands treatment. Thus, as is true of so many decisions in medicine, referral decisions are typically made in the face of uncertainty.[11] Clinicians must consider whether the likelihood of benefit (eg, amelioration or prevention of symptoms and impairment attributable to treatment) outweighs costs (eg, family’s time and copays or use of scarce resources) and the

likelihood of harm (eg, the family's experiences of anxiety and stigma and use of scarce mental health and developmental resources).

Rational decision-making demands consideration of the benefits as well as the costs of referral. For example, primary care providers often have limited staff time and mental health resources at their disposal, and even accurate screening instruments might yield many positive results. As shown in Figure 2, if 15% of children have developmental-behavioral disabilities, then 25% will score positive on a screener that shows 85% sensitivity and specificity. For a screener that shows 70% sensitivity and specificity, 36% will score positive. Considering the implementation studies of pediatric screening identified from the review cited previously, if pediatricians had suddenly begun to refer all children who scored positive, referrals would increase to levels between 1.2 and 4.3 times the original. It is worth considering whether local service systems would have the capacity to handle this many referrals, or instead whether demand for services would outstrip supply, possibly leading to longer wait times and diminished access for evaluations and treatment. Moreover, demand on primary care physicians to see patients has increased dramatically in recent years,[12] as have professional recommendations regarding the care each patient should receive.[13] These demands leave less and less time to address complicated referrals.

Furthermore, the benefits of treatment cannot be assumed in every case. Primary care pediatricians' referrals to many types of specialists are often not completed.[14–16] Even in high-quality research studies on screening, referrals often do not result in evaluations or treatment,[17,18] thus detracting from the expected efficacy of referrals overall. Moreover, whereas the efficacy of evidence-based treatments is typically well documented with respect to symptoms or diagnoses, influence on patient-centered outcomes like quality of life is typically less understood. Finally, interventions by community providers vary widely, and evidence for their efficacy is even less well studied.

Likewise, clinicians must consider the effect of referrals on families who might not have previously considered the possibility that their child might benefit from specialty services.[19] A recent report from the Institute of Medicine (IOM)/National Academy of Medicine (NAM) strongly emphasizes the importance of family engagement in the diagnostic process, noting that, "Because patients are a heterogeneous population with varying needs, values, and preferences, their roles in diagnosis need to be individually tailored."[11pp166] A large literature shows the psychological effect of receiving a diagnosis,[20] especially with regard to mental health or development, yet the challenge to pediatricians is to find a way to engage with families to conduct ongoing surveillance and to maximize participation with treatment plans.

Overall, discussion of the costs, harms, and benefits of referral highlights 2 themes that are essential to understanding the use of screening instruments to support referral decisions. First, as the IOM/NAM report states, "absolute certainty in diagnosis is unattainable."[11pp166] Referral decisions are often made in the face of tremendous uncertainty, not only regarding the true status of the child but also with regard to the potential for benefits, costs, and harms.[21] Second, many essential elements to the decision-making process are specific to the individual patient or to the local community. Stigma and patient-centered outcomes are

strongly influenced by families' perceptions and the communities in which they live. Availability and quality of treatment resources also vary from community to community. Thus, a screening score that is high enough to justify a referral for a particular child in one context might not be high enough for a child from a different family from a different community. Blanket recommendations for referrals fail to address these complexities.

As described previously, screening thresholds represent tradeoffs between sensitivity on the one hand and predictive value (including PPV and threshold probability, both of which covary with specificity) on the other. Low thresholds favor sensitivity at the cost of predictive value—that is, they detect more children with disabilities, but at the cost of more false positive results. High thresholds favor predictive value, but at the cost of sensitivity—that is, positive scores are more likely to be correct, but more children with disabilities will be missed. In practice, thresholds for many instruments are set by maximizing the Youden index,[22] which is equivalent to the sum of sensitivity and specificity and typically comes close to balancing the two. However, as shown in Figure 2, balancing sensitivity and specificity often yields predictive values that are inadequate to justify referrals in many cases. Thus, use of the Youden index seldom "optimizes" screening thresholds from a cost/benefit perspective.[23,24] One hypothesis for why pediatricians do not refer all children with positive screening results is that thresholds set using the Youden index are too low to justify referrals for individual patients in many settings. If pediatricians respond by not referring most children who score just above the threshold, then they are effectively raising the screening threshold, thereby favoring predictive value at the cost of sensitivity.[7] For screening to be effective, understanding the tradeoffs in any threshold is essential.

If not all children who score positive should be referred, there are significant implications for screening policy, education, and research. The argument that positive screening results are often insufficient to justify referrals suggests that a more nuanced approach to identification and diagnosis is needed. Notably, other professional screening recommendations for A1C, prostate-specific antigen, and allergen-specific immunoglobulin E tests include referral as an option, but they also discuss preventive interventions, further assessment to clarify risk, and informed decision-making with patients.[25–27] Consistent with the recommendations of the recent IOM/NAM report on improving diagnosis in health care, we discuss potential implications of this approach for policy, education, and research on developmental-behavioral screening.

Screening Policy

We argue that policy statements should not recommend referral for all children with positive screening results unless there is evidence that the benefits are likely to outweigh costs and harms for children who score at recommended screening thresholds. Because of the absence of such evidence for developmental-behavioral screening, current recommendations should be significantly revised to offer more specific guidance regarding expectations for individual practitioners. Referrals for all positive scores might still be recommended if threshold probabilities are sufficiently high, but only in very specific conditions. For example, a recommendation to refer all children who screen positive might be justified if screening instruments are extremely accurate—a goal that computerized adaptive testing has shown

promise of achieving.[28,29] A recommendation to refer all children with positive screens might also be justified if thresholds are set extremely high,[30] or if the risks and costs of referral are extremely low—as might be the case for preventive services (eg, interventions to promote positive parenting)[31–33] and/or to a colocated provider who is trained to administer a brief follow-up assessment (eg, the Modified Checklist for Autism in Toddlers Follow-Up Interview).[34] In other cases, developers of screening instruments would ideally begin using alternative methods for setting thresholds that more carefully consider perspectives on potential costs and benefits.[35] For example, developers might consider recommending multiple thresholds that are appropriate for different clinical decisions such as a treatment, assessment, and further monitoring. As noted by a helpful reviewer, multiple thresholds are found in recommendations for other pediatric disorders, such as serum bilirubin monitoring to detect jaundice,[36] and similar logic could be used for developmental-behavioral disorders. However, in more typical cases when pediatricians conduct universal screening in community settings, it might be more reasonable to recommend that positive screening scores—particularly those close to a threshold on the basis of the Youden index—be addressed through additional assessment in which providers incorporate all available information, including evidence from screening as well as from surveillance (ie, observations of the child, additional parent reports, and reports from other professionals). For this process, the recommendation of the IOM/NAM for shared decision-making with families to assess perspectives on the benefits, costs, and potential for harm from the full range of available referral options is particularly relevant.[11]

Education

With increased emphasis on clinical judgement, improvements in education are also warranted. To improve the diagnostic process, the IOM/NAM specifically recommends enhancing professional education and training in “clinical judgement” and “appropriate use of diagnostic tests and the application of these results on subsequent decision-making.”[11pp166] If screening is not sufficient to drive referral decisions, then enhanced training in clinical reasoning might be needed. A large literature in evidence-based medicine recommends specific methods for interpreting test results to support decision-making[37–41] as well as closer attention to strengths and limitations in the evidence used to support the validity of different screening instruments.[42] In particular, we argue that an understanding of the tradeoffs inherent in any screening threshold is essential for appropriate interpretation.

Research

Most research on screening focuses on the standardization, construct validity, and diagnostic accuracy of screening instruments. In contrast, the IOM/NAM describes diagnosis as a longitudinal, iterative process that often involves multiple team members and close engagement with families. This perspective suggests that research on screening should not be confined to the accuracy of screening instruments when used alone, but should also address their ability to improve clinical decision-making in the context of systems of care. Consistent with this view, one of the most cited articles in the journal, *Medical Decision Making*, argues that the effectiveness of a clinical test rests not only on the accuracy of the

instrument, but also on how it is used, whether and how it influences clinical decision-making and referral status, and ultimately how it affects the health of patients.[43]

In developmental-behavioral pediatrics, more research is needed on how pediatricians can best integrate risk information from a range of sources (eg, observation, screeners, reports from child care providers, direct communications with parents, etc). For example, improved health information technology systems that rely on human factors research to efficiently show relevant information to decision-makers have the potential to incorporate predictive models that synthesize evidence across a range of risk factors, like those commonly used in cardiology.[44] More research is needed on families' and physicians' perspectives on the costs and benefits of referrals to support shared decision-making and the development of patient-informed screening thresholds. In particular, practicing pediatricians are often reluctant to refer if they are worried about causing undo stress or stigma to families, yet the research literature is often silent with regard to the risk of harm resulting from referrals. For example, a recent statement from the United States Preventive Services Task Force reported no studies that addressed harms of autism screening,[17] leading to a finding that harms were likely to be small.[45] However, the statement noted that the high dropout rate in screening studies suggests that the process might be difficult for some families.

Finally, further research is needed on practical approaches to increase identification of developmental-behavioral disabilities that complement the use of screening instruments and enhance their effectiveness. For example, models that conceptualize screening as but one element in a larger system of care[5] might motivate research on ways to reduce burden in connecting patients with local treatment resources (eg, Help Me Grow),[46] research to advance models of collaborative, colocated mental health care (eg, Healthy Steps),[47] and research to reduce the burden of follow-up assessments to improve diagnostic validity (eg, Modified Checklist for Autism in Toddlers Follow-Up Interview).[34] Together, interventions such as these have the potential to improve identification far beyond what the use of screening instruments can accomplish alone.

Acknowledgments

Financial disclosure: Research support was provided by NICHD R01 HD072778. The opinions expressed in this article are the author's own and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

References

1. Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee and Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics. 2006; 118:405–420. [PubMed: 16818591]
2. Centers for Disease Control and Prevention. [Accessed August 16, 2016] Developmental Monitoring and Screening for Health Professionals. Available at http://www.cdc.gov/ncbddd/childdevelopment/screening-hcp.html
3. Sheldrick RC, Merchant S, Perrin EC. Identification of developmental-behavioral problems in primary care: a systematic review. Pediatrics. 2011; 128:356–363. [PubMed: 21727101]

4. Wissow LS, Brown J, Fothergill KE, et al. Universal mental health screening in pediatric primary care: a systematic review. J Am Acad Child Adolesc Psychiatry. 2013; 52:1134–1147. [PubMed: 24157388]
5. Sheldrick RC, Breuer DJ, Hassan R, et al. A system dynamics model of clinical decision thresholds for the detection of developmental-behavioral disorders. Implement Sci. 2016; 11:156. [PubMed: 27884203]
6. Glascoe F. Screening for developmental and behavioral problems. Ment Retard Dev Disabil Res Rev. 2005; 11:173–179. [PubMed: 16161092]
7. Sheldrick RC, Benneyan JC, Kiss IG, et al. Thresholds and accuracy in screening tools for early detection of psychopathology. J Child Psychol Psychiatry. 2015; 56:936–948. [PubMed: 26096036]
8. Lavigne JV, Feldman M, Meyers KM. Screening for mental health problems: addressing the base rate fallacy for a sustainable screening program in integrated primary care. J Pediatr Psychol. 2016; 41:1081–1090. [PubMed: 27289070]
9. Meehl PE, Rosen A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. Psychol Bull. 1955; 52:194–216. [PubMed: 14371890]
10. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med. 1980; 302:1109–1117. [PubMed: 7366635]
11. Institute of Medicine. Improving Diagnosis in Health Care. Washington, DC: National Academies of Sciences, Engineering, and Medicine; 2015.
12. Baron RJ. What's keeping us so busy in primary care? A snapshot from one practice. N Engl J Med. 2010; 362:1632–1636. [PubMed: 20427812]
13. Belamarich PF, Gandica R, Stein RE, et al. Drowning in a sea of advice: pediatricians and American Academy of Pediatrics policy statements. Pediatrics. 2006; 118:e964–e978. [PubMed: 17015516]
14. Zuckerman KE, Cai X, Perrin JM, et al. Incomplete specialty referral among children in community health centers. J Pediatr. 2011; 158:24–30. [PubMed: 20801461]
15. Zuckerman KE, Perrin JM, Hobrecker K, et al. Barriers to specialty care and specialty referral completion in the community health center setting. J Pediatr. 2013; 162:409–414. [PubMed: 22929162]
16. Forrest CB, Shadmi E, Nutting PA, et al. Specialty referral completion among primary care patients: results from the ASPN Referral Study. Ann Fam Med. 2007; 5:361–367. [PubMed: 17664503]
17. McPheeters, ML., Weitlauf, AS., Vehorn, A., et al. Screening for Autism Spectrum Disorder in Young Children: A Systematic Evidence Review for the U.S. Preventive Services Task Force. Evidence Synthesis No. 129. AHRQ Publication No. 13-05185-EF-1. Rockville, Md: Agency for Healthcare Research and Quality; 2016.
18. Guevara JP, Gerdes M, Localio R, et al. Effectiveness of developmental screening in an urban setting. Pediatrics. 2013; 131:30–37. [PubMed: 23248223]
19. Stuart M, McGrew JH. Caregiver burden after receiving a diagnosis of an autism spectrum disorder. Res Autism Spectrum Disord. 2009; 3:86–97.
20. Jutel, AG., Conrad, P. Putting a Name to It: Diagnosis in Contemporary Society. Baltimore, Md: JHU Press; 2014.
21. Djulbegovic B, Ende J, Hamm RM, et al. When is it rational to order a diagnostic test, or prescribe treatment: the threshold model as an explanation of practice variation. Eur J Clin Invest. 2015; 45:485–493. [PubMed: 25675907]
22. Youden WJ. Index for rating diagnostic tests. Cancer. 1950; 3:32–35. [PubMed: 15405679]
23. Smits N. A note on Youden's J and its cost ratio. BMC Med Res Methodol. 2010; 10:89. [PubMed: 20920288]
24. Bewick V, Cheek L, Ball J. Statistics review 13: receiver operating characteristic curves. Crit Care. 2004; 8:508–512. [PubMed: 15566624]
25. Crawford ED, Rosenberg MT, Partin AW, et al. An approach using PSA levels of 1.5 ng/mL as the cutoff for prostate cancer screening in primary care. Urology. 2016; 96:116–120. [PubMed: 27450937]

26. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2012; 35(suppl 1):S64–S71. [PubMed: 22187472]
27. Sicherer SH, Wood RA. Allergy testing in childhood: using allergen-specific IgE tests, the section on allergy and immunology. Pediatrics. 2012; 129:193–197. [PubMed: 22201146]
28. Gibbons RD, Weiss DJ, Pilkonis PA, et al. Development of a computerized adaptive test for depression. Arch Gen Psychiatry. 2012; 69:1104–1112. [PubMed: 23117634]
29. Gardner W, Shear K, Kelleher KJ, et al. Computerized adaptive measurement of depression: a simulation study. BMC Psychiatry. 2004; 4:13. [PubMed: 15132755]
30. Tomlinson M, Rotheram-Borus MJ, Harwood J, et al. Community health workers can improve child growth of antenatally-depressed, South African mothers: a cluster randomized controlled trial. BMC Psychiatry. 2015; 15:225. [PubMed: 26400691]
31. Weisleder A, Cates CB, Dreyer BP, et al. Promotion of positive parenting and prevention of socioemotional disparities. Pediatrics. 2016; 137:e20153239. [PubMed: 26817934]
32. Perrin EC, Sheldrick RC, McMenamy JM, et al. Improving parenting skills for families of young children in pediatric settings: a randomized clinical trial. JAMA Pediatr. 2014; 168:16–24. [PubMed: 24190691]
33. Sanders MR. Triple P-Positive Parenting Program: towards an empirically validated multilevel parenting and family support strategy for the prevention of behavior and emotional problems in children. Clin Child Fam Psychol Rev. 1999; 2:71–90. [PubMed: 11225933]
34. Robins, DL., Fein, D., Barton, M. Follow-up interview for the modified checklist for autism in toddlers (M-CHAT FUI). Self-published; 1999. p. 131-144.
35. Swets JA. The science of choosing the right decision threshold in high-stakes diagnostics. Am Psychol. 1992; 47:522–532. [PubMed: 1595983]
36. American Academy of Pediatrics Subcommittee on Hyperbilirubinemia. Management of hyperbilirubinemia in the newborn infant 35 or more weeks of gestation [erratum in: 2004;114:1138]. Pediatrics. 2004; 114:297–316. [PubMed: 15231951]
37. Guyatt, G., Drummond, R., Meade, MO., et al. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. 2. Chicago, Ill: McGraw Hill Medical; 2008.
38. Jaeschke R, Guyatt GH, Sackett DL, et al. Users' guides to the medical literature: III. How to use an article about a diagnostic test B. What are the results and will they help me in caring for my patients? JAMA. 1994; 271:703–707. [PubMed: 8309035]
39. Straus, SE., Glasziou, P., Richardson, WS., et al. Evidence-Based Medicine: How to Practice and Teach EBM. 4. New York: Churchill Livingstone; 2011.
40. Youngstrom EA. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. J Pediatr Psychol. 2014; 39:204–221. [PubMed: 23965298]
41. Youngstrom EA, Choukas-Bradley S, Calhoun CD, et al. Clinical guide to the evidence-based assessment approach to diagnosis and treatment. Cogn Behav Pract. 2014; 22:20–35.
42. Camp BW. What the clinician really needs to know: questioning the clinical usefulness of sensitivity and specificity in studies of screening tests. J Dev Behav Pediatr. 2006; 27:226–230. [PubMed: 16775521]
43. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making. 1991; 11:88–94. [PubMed: 1907710]
44. Lloyd-Jones DM, Wilson PW, Larson MG, et al. Framingham risk score and prediction of lifetime risk for coronary heart disease. Am J Cardol. 2004; 94:20–24.
45. Siu AL, Bibbins-Domingo K, Grossman DC, et al. Screening for autism spectrum disorder in young children: US preventive services task force recommendation statement. JAMA. 2016; 315:691–696. [PubMed: 26881372]
46. Bogin J. Enhancing developmental services in primary care: the Help Me Grow experience. J Dev Behav Pediatr. 2006; 27:S8–S12. [PubMed: 16715785]
47. Zuckerman B, Parker S, Kaplan-Sanoff M, et al. Healthy Steps: a case study of innovation in pediatric practice. Pediatrics. 2004; 114:820–826. [PubMed: 15342859]

What's New

We argue that professional guidelines that recommend referral subsequent to positive results on developmental-behavioral screening instruments should be revised because: 1) positive predictive value often overstates the probability of disorder; and 2) positive scores do not ensure that the benefits of referral outweigh the risks.

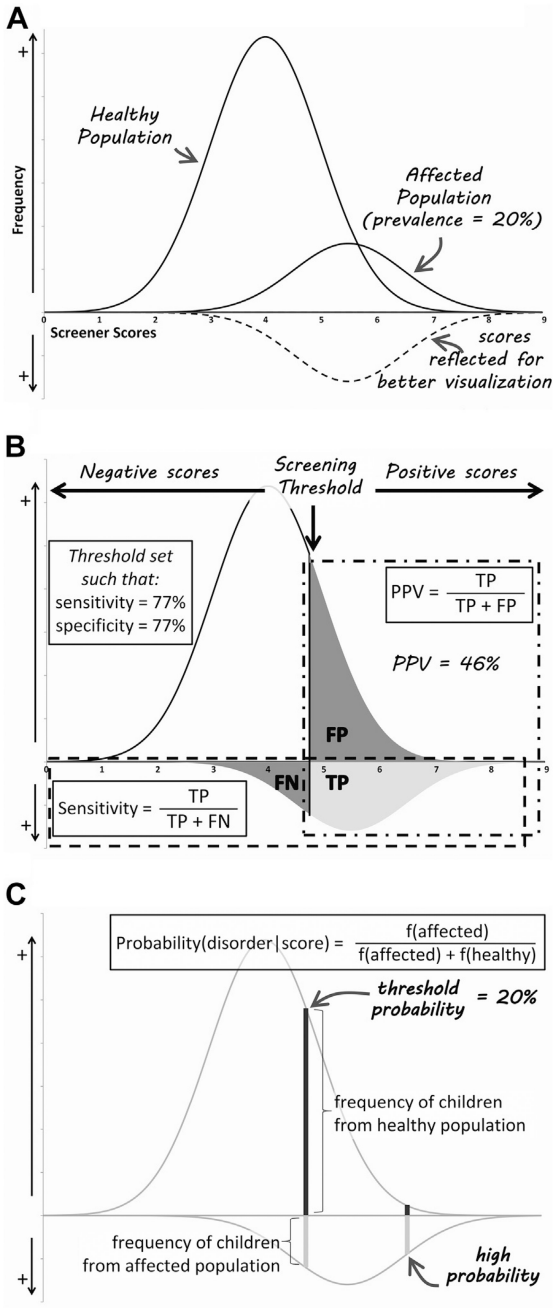

Figure 1.

Visualization of sensitivity, positive predictive value (PPV), and threshold probability. (A) Normal distributions of screening scores. (B) Sensitivity and PPV: calculated from all true positive (TP), false positive (FP), or false negative (FN) results. (C) Threshold probability: calculated on the basis of frequency (f) of children from healthy and affected populations who score at the threshold.

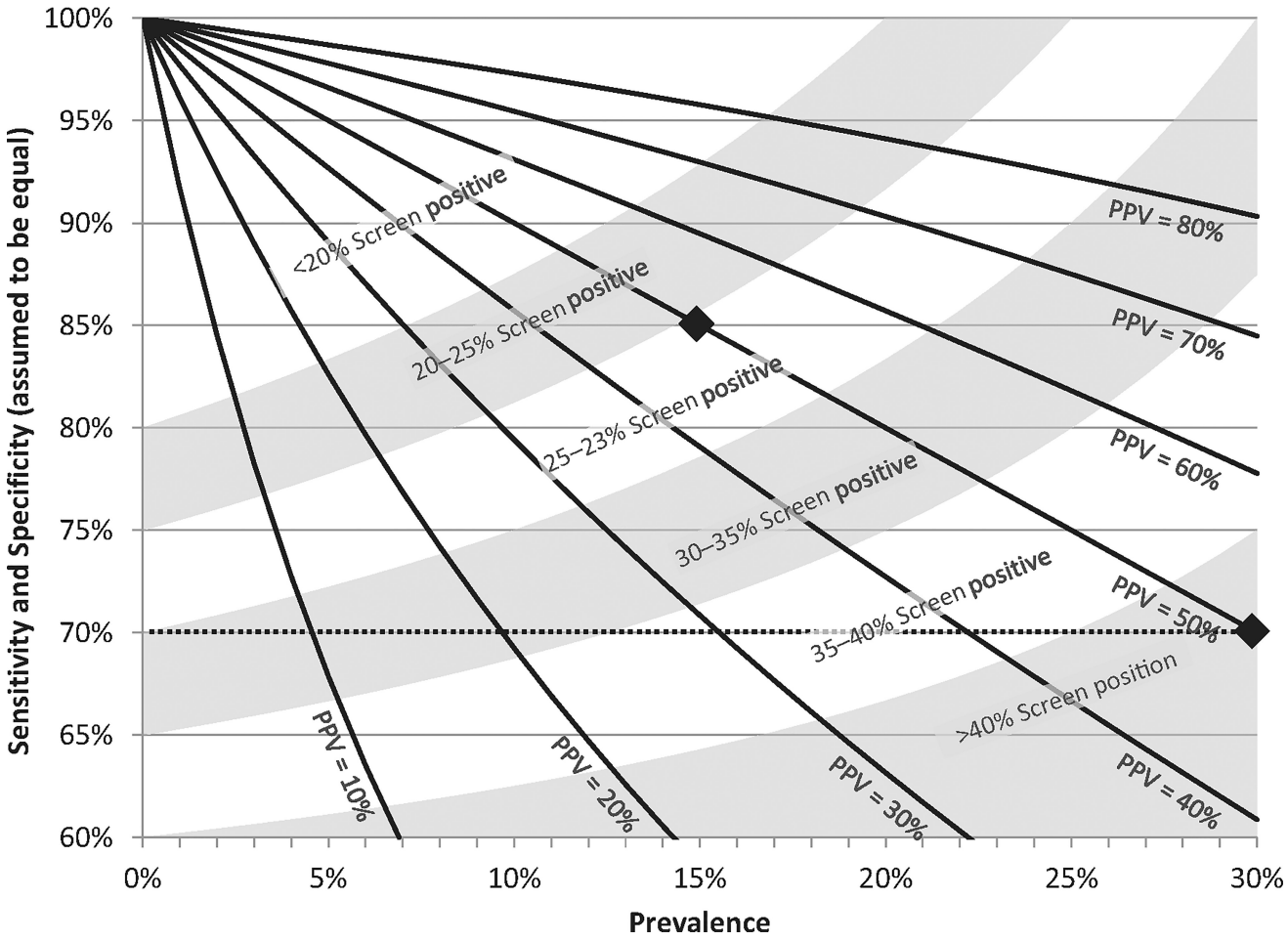

Figure 2.

Positive predictive value (PPV) and proportion who screen positive as a function of sensitivity, specificity, and prevalence.