Diagnostic Efficiency of the CBCL Thought Problems and DSM-Oriented Psychotic Symptoms Scales for Pediatric Psychotic Symptoms

Stephanie Salcedo, MA[1], **Sabeen H. Rizvi, PhD**[1,2], **Lindsey K. Freeman, BA**[3], **Jennifer K. Youngstrom, PhD**[1], **Robert L. Findling, MD, MBA**[4], and **Eric A. Youngstrom, PhD**[1]

[1]Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, North Carolina, U.S.A

[2]Department of Psychology, Gargi College, Delhi University, New Delhi, India

[3]Department of Psychological Sciences, University of Missouri, Columbia, Missouri, U.S.A

[4]Department of Psychiatry, Johns Hopkins University, Maryland, U.S.A

Abstract

Objective—We compared the diagnostic efficiency of the Child Behavior Checklist (CBCL) Thought Problems subscale and the rationally derived DSM-oriented Psychotic Symptoms Scale (DOPSS) to identify clinically concerning psychosis in a multi-site sample of youths seeking outpatient mental health services ($N = 694$).

Method—We operationally defined clinically concerning psychosis as the presence of clinically significant hallucinations or delusions, assessed by the Schedule for Affective Disorders and Schizophrenia (K-SADS) psychosis items.

Results—Both the Thought Problems and DOPSS scores showed significant areas under the curve (AUCs=.65 and .70, respectively), but the briefer DOPSS showed statistically significantly better diagnostic efficiency for any clinically concerning psychosis, but the difference was small enough that it would not be clinically meaningful. The optimal psychosis screening cut-off score (maximizing sensitivity and specificity) was 68.5+ [corresponding diagnostic likelihood ratio (DiLR)=1.59] for the Thought Problems subscale and 1.67+ (DiLR=1.97) for the DOPSS.

Conclusions—Both the CBCL Thought Problems and DOPSS are clinically useful for identifying psychotic symptoms in children, and although the DOPSS showed statistically better

discriminating power, the difference was small so we would not necessarily recommend the DOPSS over standard scoring.

Keywords

Psychotic symptoms; children; evidence-based assessment; CBCL; Thought Problems; ROC

Psychosis can be associated with severe impairment and immense burden for the individuals affected and their families [1,2]. Because psychosis is often associated with a reduced quality of life, early assessment and diagnosis are necessary to treat psychotic symptoms and reduce their negative consequences [3]. However, not all youths who report unusual beliefs or concerning behavior to their parents will have a psychiatric disorder. Psychotic-like experiences among youths are common, with 17% of children ages 9–12 years and 7.5% of adolescents ages 13–18 years reporting psychotic-like symptoms in a meta-analysis [4]; but far fewer developed full-fledged disorders with psychosis. Therefore, there is a need to balance the competing goals of early detection of clinically concerning psychotic symptoms to identify those warranting thorough evaluation and early intervention and avoiding false positives that may pathologize normal behavior and lead to unnecessary treatment. It would be particularly valuable to know the accuracy of widely used instruments for alerting to potential psychosis, as many such measures are widely used in outpatient settings. If one of these were reasonably accurate, that could help clinicians improve detection by gleaning more information from tools they already use, directly addressing concerns about implementation and feasibility, versus needing to add a specialized scale just for the purpose of screening for psychosis [5].

Several rating scales contain items about psychotic symptoms in youth [6–8]. However, less is known about the accuracy of these rating scales for identifying clinically concerning psychotic symptoms. Both evidence-based medicine (EBM) and evidence-based assessment (EBA) [9] advocate using validated tools for assessment as opposed to unaided clinical judgment to achieve diagnostic precision and improve clinical decision making.

The present study compared the discriminative accuracy of two subscales derived from the Child Behavior Checklist (CBCL) [10] the Thought Problems T-score and Lengua's rationally derived psychotic symptoms scale [11] to detect clinically concerning psychotic presentation—transdiagnostically defined—in a pediatric sample. The CBCL is among the most commonly used rating scales for youths; and it assesses a wide array of syndromes, including anxiety, depression, somatic complaints, attention difficulties, and behavior problems, as well as more broad Internalizing, Externalizing, and Total Problems scores [12]. However, the CBCL's standard scales may not be optimally effective at discriminating between different diagnoses [13]. Because the CBCL Thought problems scale combines a wide range of behaviors into a single scale on statistical grounds, it mixes items that might indicate obsessions, autistic behaviors, or several other syndromes in addition to psychosis. To try to get more clinical discrimination, Lengua and colleagues rationally derived an alternate scoring system based on clinician review of content, and then empirically validated it [11]. Although Achenbach subsequently added several DSM-oriented scales to the official scoring system, psychosis was not one of them. However, Lengua et al. developed a brief (4

item) scale for that purpose, which we will refer to the DSM-oriented Psychotic Symptoms Scale (DOPSS). We hypothesize that the DOPSS score will show greater discriminative validity than the Thought Problems score for identifying clinically concerning psychosis because it more narrowly focuses on relevant content. We also compare the performance of both scales to a recent study evaluating the Child and Adolescent Symptom Inventory (CASI-4R) [14] in an independent sample using similar diagnostic methods [15].

Method

Participants and Procedures

These secondary analyses use data from the Assessing Bipolar Disorder: A Community-Academic Blend (ABACAB) [16]. The Institutional Review Board at Case Western Reserve University and Applewood Centers, Incorporated reviewed and approved the protocol for this primary outcome study. Participants were recruited from community mental health centers (with four urban sites) and an outpatient academic medical center with multiple pharmacotherapy studies. Inclusion criteria were (a) youths between the ages of 5 and 18, (b) youths and parents able to communicate orally in English, and (c) both youths and parents present for the assessments. Participants were excluded if youths met criteria for a pervasive developmental disorder or mental retardation [16]. Present analyses focused on youths between ages 6 and 18 years of age (N=694; M=11.20; SD= 3.26) as the norms for the CBCL start at age 6.

Measures

Diagnoses—Both the youths and caregivers were interviewed using the Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL) [17] and mood modules from the Washington University version. The interviewers were rigorously trained before conducting the actual interviews (inter-rater K . 85 at item level). The same clinician interviewed both the youth and caregiver, and interviewers used clinical judgment to determine whether any re-interviews were necessary to determine a diagnosis. A licensed child psychologist or psychiatrist reviewed all the diagnoses, and a consensus review process finalized diagnoses [18]. The KSADS criterion diagnoses were blind to CBCL scores. Clinically concerning psychosis for our study was operationally defined as a score of 3 or more on any of the hallucination and/or a score of 3 or more on the delusion item on the K-SADS, consistent with prior studies (see appendix for K-SADS hallucinations and delusions questions) [19]. Children were included if they met the threshold for clinical psychotic symptoms on the K-SADS. The youths in this group met criteria for any of the following groups of disorders (including diagnoses in the mood spectrum): ADHD, ODD, schizophrenia or any other psychotic disorders, disruptive behavior disorders (ODD, CD), PTSD, adjustment disorder, anxiety, bipolar disorder (types I, II), and unipolar depression (or other mood disorder).

Achenbach Child Behavior Checklist (CBCL) [10]—The CBCL (118-item parent report) assesses a broad range of psychopathology within the past six months. We focused on the CBCL Thought Problem Scale and Lengua's DSM-oriented Psychotic Symptoms Scale [11]. The CBCL Thought Problem Scale contains 15 items assessing obsessive

thoughts, self-harm, hallucinations, nervous twitching, picking parts of body, playing with own sex parts in public and/or too much, compulsions, less need for sleep, storing too many items, strange behavior or ideas, sleep walking or talking, and trouble sleeping. The items are rated from 0 (*not true*) to 2 (*very true or often true*). The CBCL Thought Problems Scale offers a *T*-score based on age and sex norms.

The DOPSS [11] uses four items that assess visual and auditory hallucinations, as well as strange behavior or ideas. We prorated scores if respondents missed one item and only calculated scores for parents who completed at least three of these items (otherwise they were scored as missing). Reliability was acceptable for both the scales (Thought Problem Scale α=.77; DOPSS α=.61–good considering its brevity).

Data Analytic Plan

Analyses used SPSS (Version 20) and R software packages pROC and Aod [20]. We examined baseline demographic and clinical characteristics for potential outliers. We calculated Areas Under the Curve (AUCs) from Receiver Operating Characteristic (ROC) analyses to quantify the diagnostic efficiency of the two CBCL subscales, Thought Problems and DOPSS, for identifying any clinically concerning psychosis. Guidelines by Rice and Harris [21] suggest that AUCs in the mid .50s are considered small, mid .60s are medium (and may provide incremental clinical value), low to mid .70s are large (clinically useful), low .80s are excellent, and high .80s are exceptional under clinically rigorous studies, as well as highly informative.

Next we used Venkatraman's test [22] to compare the ROC curves between the Thought Problems and DOPSS scores, which compares the curves at all the operating points instead of the overall AUC estimates, allowing for more power to detect differences [22]. We also benchmarked both scales against the performance of the CASI psychosis scale in an independent sample [15] using the Hanley & McNeil test [23]. The CASI sample was from the Longitudinal Assessment of Manic Symptoms (LAMS) study, which recruited participants from 10 outpatient mental health centers in the Midwest (for more information on recruitment and eligibility, see [24]. We calculated diagnostic likelihood ratios (DiLRs) from both splitting the sample into quintiles and optimal cut-points for each scale that maximized the combination of sensitivity and specificity from the ROC curves [20]. DiLRs are clinically useful effect sizes that inform the odds of a diagnosis associated with a particular range of scores. DiLRs less than 1 are associated with test scores that indicate lower probability of disorder, whereas scores above 1 are associated with higher probabilities of the disorder [25].

Results

Demographics and Preliminary Analyses

The study sample included 70% African Americans ($n = 488$), 23% Caucasians ($n = 160$), and 2% Hispanics ($n = 13$). Nine percent of our sample had clinically concerning psychosis ($n=64$). Older participants (ages 11–18) were significantly more likely to have clinically concerning psychosis (11% vs. 7%; $p < .05$). Table 1 shows the demographic and clinical

differences between children of these two groups. Youths with any clinically concerning psychosis had significantly more Axis I diagnoses at baseline, and they had higher CBCL Thought Problems and DOPSS scores than youths who did not meet criteria for clinically concerning psychosis (ps<.005). Youths with clinically concerning psychosis were also more likely to be female and White (ps .05). Youths with clinically concerning psychosis were also more likely to have bipolar disorder, PTSD, conduct disorder, and an anxiety disorder (ps<.05). There were no significant differences in age (p >.05; see Table 1).

Diagnostic Efficiency Analyses

Table 2 presents the AUCs for the CBCL thought problem T scores and DOPSS scores in youths with any clinically concerning psychosis. Both subscales performed fairly well in discriminating between youths with clinically concerning psychosis (medium to large effect sizes; AUCs=.65 and .70, ps<.001; see Figure 1). The DOPSS AUC was significantly higher than the Thought Problems AUC (Venkatraman p <.05). However, with the DOPSS AUC only being .05 greater than the Thought Problems AUC, this difference is not necessarily practically significant. Both performed significantly less well than the CASI Psychosis scale did in the Longitudinal Assessment of Manic Symptoms (LAMS) data [15]: CASI AUC=.83, versus the Thought Problems AUC=.65, Hanley & McNeil z=3.93, p=.0001, and the DOPSS AUC=.70, Hanley & McNeil z=2.75, p=.0059.

Diagnostic Likelihood Ratios (DiLRs)

Table 3 reports DiLRs using cut-off scores that optimized the combination of sensitivity and specificity, as well as more fine-grained multi-level DiLRs (divided into four strata) [25] for the CBCL Thought Problems and DOPSS scores. We provided DiLRs for *any clinically concerning psychosis*, as this is more clinically useful. Using optimal cut scores, a Thought Problems T-score of 68.5 (Sensitivity = 61%; Specificity = 62%) or higher results in increased odds (DiLR+ =1.59) of experiencing any clinically concerning psychosis. Using multilevel DiLRs, youths with clinically concerning psychosis were over 3 times more likely (DiLR+ =3.28) to score 83 or higher on the Thought Problems Scale. In contrast, youths with Thought Problems T-score below 59 had decreased odds of experiencing clinical psychosis.

For the DOPSS, an optimal cut score of 1.67 (Sensitivity=64%; Specificity=67%) or higher increased the odds of any clinically concerning psychosis by 1.97. With multilevel DiLRs, youths who score 5 or more on the DOPSS scale have approximately a 3.69 times increase in odds (DiLR+ =3.69) of having clinical psychosis, whereas those who score below 2 have a decrease in odds (DiLR+ =0.53).

Discussion

Our aim was to test and compare the diagnostic efficiency of the Achenbach Thought Problems Scale and DOPSS as screening measures for any clinically meaningful psychosis, defined transdiagnostically, in a pediatric outpatient mental health setting among youths ages 6 to 18. Because the CBCL is widely used, often as part of a core battery, it is helpful to know whether scores on it provide any information alerting about potential psychosis. Both

scales were statistically valid. The DOPSS significantly outperformed the Thought Problems scale, but the difference was small enough that it would not be clinically meaningful. Both scales showed medium to large effect sizes according to Rice and Harris' [21] guidelines, indicating that either the Thought Problems and DOPSS scores could be clinically useful for identifying clinically concerning psychosis. The Thought Problems is routinely available in Achenbach scoring report, whereas the DOPSS would require additional hand scoring. Although the DOPSS showed statistically better discriminating power, the difference was modest in terms of posterior probability estimates for a patient having psychosis under clinically realistic conditions (Table 3). Therefore, we would not necessarily recommend the DOPSS over the standard scoring.

We calculated optimal cut DiLRs, a form of effect size that helps estimate change in odds of diagnosis for an individual, to make the Thought Problems and DOPSS scales easier to apply in clinical settings. Thought Problems *T*-scores of 68.5 or greater were associated with 1.59 times greater odds of reporting any clinically concerning psychosis, whereas DOPSS scores of 1.67 or greater were associated with 1.97 times greater odds of reporting clinically concerning psychosis. Multilevel DiLRs make the test more informative by calculating DiLRs for extreme, intermediate, and low score ranges. High scores on the Thought Problems subscale (83+) were associated with 3.28 times greater odds, whereas high scores on the DOPSS (5+) were associated with 3.69 times greater odds. It is important to note that a positive score does *not* indicate that the individual has psychosis; instead, a positive score indicates that further assessment is needed to establish a diagnosis and determine the best course of treatment. A low score, or low PPV, on the other hand, is more decisive and could be clinically helpful because it can serve to rule out a diagnosis [26]. Another important consideration is that the diagnostic accuracy can change across different settings. Changes in base rate directly affect the positive and negative predictive powers, and changes in the severity of psychosis would directly influence sensitivity. Similarly, settings with more distressed patients and more complex presentations would have higher false alarm rates, and correspondingly lower specificity [27–29]. In contrast, if this measure is used in the general pediatric clinic, we would expect it to become less sensitive (due to milder psychosis being much more common than severe) but possibly more specific (as conditions that might lead to false positives also might become more rare).

The prevalence of clinically meaningful psychotic presentation in our sample (9%) was similar to rates from other studies [19,30], and the minimal exclusion criteria also increase the generalizability of results. The sample was diverse in nature and relatively large. Another strength was that we had an "a priori" design, in which our research question and separate psychosis groups were decided before running the analyses, and diagnoses were blind to the results of the CBCL. To our knowledge, we are the first to use ROC analyses for the DOPSS.

Limitations include that we only used the parent report for our analyses. Future research should examine parent-youth agreement about psychotic symptoms using the Achenbach Youth Self Report or Teacher Report Form scores. In addition, we only included English speakers in this study. Although the inclusion and exclusion criteria were minimal, it is important to replicate and compare these findings in other outpatient and academic settings

[15]. Of note, both performed significantly less well than the CASI Psychosis scale recently did in an independent sample [15].

When assessing for psychotic symptoms, it may be helpful for clinicians to have brief screening tools to better identify cases with clinically concerning psychosis. Although both the Thought Problems and DOPSS scores showed significant and clinically meaningful ability to discriminate clinically concerning psychosis, the four-item scale proposed by Lengua and colleagues (2001) showed better discriminating power, justifying the extra work in hand-scoring the four items. Of note, the CASI Psychosis scale performed even better [15], suggesting that it might be worth switching to it in settings where psychosis might be a particular concern. Screening and offering interventions may be efficient and cost effective for children and adolescents at an early age before disability or dysfunctionality becomes embedded and difficult to reverse. It is important to remember that a positive screen on any of the subscales does not guarantee a psychosis diagnosis. Clinicians should be cautious while interpreting the findings and should always rely on other factors (e.g., base rates of the setting, family history and other potential risk factors, and ideally followed up with a valid reference standard such as a KSADS or other semi-structured approach) to reach a valid diagnosis. Tests like these may contribute to the assessment by raising the index of suspicion or "red flags" triggering thorough assessment for diagnosis [18].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Harrison G, Hopper K, Craig T, Laska E, Siegel C, Wanderling J, Dube KC, Ganev K, Giel R, an der Heiden W, Holmberg SK, Janca A, Lee PW, Leon CA, Malhotra S, Marsella AJ, Nakane Y, Sartorius N, Shen Y, Skoda C, Thara R, Tsirkin SJ, Varma VK, Walsh D, Wiersma D. Recovery from psychotic illness: a 15- and 25-year international follow-up study. Br J Psychiatry. 2001; 178:506–517. [PubMed: 11388966]

2. Wu EQ, Birnbaum HG, Shi L, Ball DE, Kessler RC, Moulis M, Aggarwal J. The economic burden of schizophrenia in the United States in 2002. J Clin Psychiatry. 2005; 66(9):1122–1129. [PubMed: 16187769]

3. Stafford MR, Jackson H, Mayo-Wilson E, Morrison AP, Kendall T. Early interventions to prevent psychosis: systematic review and meta-analysis. BMJ. 2013; 346:f185.doi: 10.1136/bmj.f185 [PubMed: 23335473]

4. Kelleher I, Connor D, Clarke MC, Devlin N, Harley M, Cannon M. Prevalence of psychotic symptoms in childhood and adolescence: a systematic review and meta-analysis of population-based studies. Psychol Med. 2012; 42(9):1857–1863. DOI: 10.1017/s0033291711002960 [PubMed: 22225730]

5. Kelleher I, Harley M, Murtagh A, Cannon M. Are screening instruments valid for psychotic-like experiences? A validation study of screening questions for psychotic-like experiences using in-depth clinical interview. Schizophr Bull. 2011; 37(2):362–369. DOI: 10.1093/schbul/sbp057 [PubMed: 19542527]

6. Kline E, Schiffman J. Psychosis risk screening: A systematic review. Schizophr Res. 2014; 158(1–3):11–18. DOI: 10.1016/j.schres.2014.06.036 [PubMed: 25034762]

7. Thompson E, Kline E, Reeves G, Pitts SC, Bussell K, Schiffman J. Using parent and youth reports from the Behavior Assessment System for Children, Second Edition to identify individuals at clinical high-risk for psychosis. Schizophr Res. 2014; 154(1–3):107–112. DOI: 10.1016/j.schres.2014.02.009 [PubMed: 24630261]

8. Thompson E, Kline E, Reeves G, Pitts SC, Schiffman J. Identifying youth at risk for psychosis using the Behavior Assessment System for Children, Second Edition. Schizophr Res. 2013; 151(1–3): 238–244. DOI: 10.1016/j.schres.2013.09.022 [PubMed: 24119463]

9. Youngstrom EA. A Primer on Receiver Operating Characteristic Analysis and Diagnostic Efficiency Statistics for Pediatric Psychology: We Are Ready to ROC. Journal of pediatric psychology. 2013; doi: 10.1093/jpepsy/jst062

10. Achenbach TM, Rescorla LA. Manual for the ASEBA School-Age Forms and Profiles. University of Vermont, Research Center for Children, Youth, and Families; Burlington, VT: 2001.

11. Lengua LJ, Sadowski CA, Friedrich WN, Fisher J. Rationally and empirically derived dimensions of children's symptomatology: expert ratings and confirmatory factor analyses of the CBCL. J Consult Clin Psychol. 2001; 69(4):683–698. [PubMed: 11550734]

12. Macmann GM, Barnett DW, Lopez EJ. The Child Behavior Checklist/4–28 and related materials: Reliability and validity of syndromal assessment. School Psych Rev. 1993; 22(2):322–333.

13. Jensen PS, Watanabe HK, Richters JE, Roper M, Hibbs ED, Salzberg AD, Liu S. Scales, diagnoses, and child psychopathology: II. Comparing the CBCL and the DISC against external validators. J Abnorm Child Psychol. 1996; 24(2):151–168. DOI: 10.1007/BF01441482 [PubMed: 8743242]

14. Gadow KD, Sprafkin J. Checkmate Plus. Stony Brook; NY: 1994. Child Symptom Inventories Manual.

15. Rizvi SH, Salcedo S, Youngstrom EA, Freeman LK, Group TL, Findling RL. Diagnostic accuracy of the CASI-4R psychosis subscale in pediatric outpatient mental health clinics. under review.

16. Youngstrom EA, Meyers OI, Demeter C, Kogos Youngstrom J, Morello L, Piiparinen R, Feeny NC, Findling RL, Calabrese JR. Comparing diagnostic checklists for pediatric bipolar disorder in academic and community mental health settings. Bipolar Disord. 2005; 7(Special Issue: Pediatric Bipolar Disorder):507–517. [PubMed: 16403176]

17. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. J Am Acad Child Adolesc Psychiatry. 1997; 36(7): 980–988. DOI: 10.1097/00004583-199707000-00021 [PubMed: 9204677]

18. Youngstrom EA, Meyers O, Demeter C, Youngstrom JK, Morello L, Piiparinen R, Feeny N, Calabrese JR, Findling RL. Comparing diagnostic checklists for pediatric bipolar disorder in academic and community mental health settings. Bipolar Disord. 2005; 7(6):507–517. DOI: 10.1111/j.1399-5618.2005.00269.x [PubMed: 16403176]

19. Ulloa RE, Birmaher B, Axelson D, Williamson DE, Brent DA, Ryan ND, Bridge J, Baugher M. Psychosis in a Pediatric Mood and Anxiety Disorders Clinic: Phenomenology and Correlates. J Am Acad Child Adolesc Psychiatry. 2000; 39(3):337–345. [PubMed: 10714054]

20. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011; 12(1):77. [PubMed: 21414208]

21. Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. Law Hum Behav. 2005; 29(5):615–620. DOI: 10.1007/s10979-005-6832-7 [PubMed: 16254746]

22. Venkatraman ES. A permutation test to compare receiver operating characteristic curves. Biometrics. 2000; 56(4):1134–1138. [PubMed: 11129471]

23. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983; 148:839–843. [PubMed: 6878708]

24. Findling RL, Youngstrom EA, Fristad MA, Birmaher B, Kowatch RA, Arnold LE, Frazier TW, Axelson D, Ryan N, Demeter CA, Gill MK, Fields B, Depew J, Kennedy SM, Marsh L, Rowles BM, Horwitz SM. Characteristics of children with elevated symptoms of mania: the Longitudinal

Assessment of Manic Symptoms (LAMS) study. J Clin Psychiatry. 2010; 71(12):1664–1672. DOI: 10.4088/JCP.09m05859yel [PubMed: 21034685]

25. Straus SE, Glasziou P, Richardson WS, Haynes RB. Evidence-based medicine: How to practice and teach it. 4. Churchill Livingstone Elsevier; New York, NY: 2011.

26. Straus SE, Glasziou P, Richardson WS, Haynes RB. Evidence-based medicine: How to practice and teach EBM. 4. Churchill Livingstone; New York, NY: 2011.

27. Zhou X, Obuchowski N, McClish D. Statistical methods in diagnostic medicine. New York, NY: Wiley; 2002. The sensitivity and specificity of clustered binary data; 104–106.

28. Youngstrom EA, Egerton GA, Rizvi SH, Freeman LK, Genzlinger JL, Van Meter A. Improving the global identification of bipolar spectrum disorders: Meta-analysis of the diagnostic accuracy of checklists. Psychol Bull. 2017 epub.

29. Moreira ALR, Van Meter A, Genzlinger J, Youngstrom EA. Review and Meta-Analysis of Epidemiologic Studies of Adult Bipolar Disorder. J Clin Psychiatry. 2017; 78(9):e1259–e1269. DOI: 10.4088/JCP.16r11165 [PubMed: 29188905]

30. Birmaher B, Axelson D, Monk K, Kalas C, Goldstein B, Hickey MB, Obreja M, Ehmann M, Iyengar S, Shamseddeen W, Kupfer D, Brent D. Lifetime psychiatric disorders in school-aged offspring of parents with bipolar disorder: the Pittsburgh Bipolar Offspring study. Arch Gen Psychiatry. 2009; 66(3):287–296. DOI: 10.1001/archgenpsychiatry.2008.546 [PubMed: 19255378]

31. Cohen J. Statistical power analysis for the behavioral sciences. Hilsdale. NJ: Lawrence Earlbaum Associates; 1988. 2

Appendix: K-SADS-PL-W Hallucinations and Delusions Questions

PSYCHOSIS

HALLUCINATIONS	**P**	**C**	**S**	
	0	0	0	No information or N/A
Sometimes children, when they are alone, hear voices, see things, or smell things and they don't quite know where they come from.	1	1	1	Not at all - Absent
Has this happened to you?				
Have you ever seen things that were not there?	2	2	2	Suspected / Possible
When did you?				
What did you see?	**3**	**3**	**3**	Mild. Definitely present but subject is generally aware it is his imagination and usually able to ignore it. Occurs no more than once per week.
What did you hear?				
Has there been anything unusual about the way things sounded?				
How often have you heard these voices (noises)? (smell, feeling, visions) Is it some of the time, only now and then, most of the time, or all of the time?	4	4	4	Moderate. Generally believes in the reality of the hallucinations, but it has little influence on his behavior. (Or) Occurs at least once per week.
What do you think it is?				
Do you think it is your imagination or real?	5	5	5	Severe. Convinced his hallucination is real and significantly effects his actions. i.e.; locks doors to keep pursuers away. (Or) Occurs frequently.
Did you think it was real when you (heard, saw. etc.) it?				
What did you do when you (heard, saw, etc.) it?				

PSYCHOSIS

HALLUCINATIONS	**P**	**C**	**S**	
	6	6	6	Extreme. Actions based on hallucinations have major impact on him or others: Unable to do school work because of constant "conversation." (Or) Occurs most of the time.

	PAST	_____	_____	_____
		P	**C**	**S**

DELUSIONS

	P	**C**	**S**	
Do you know what imagination is? Tell me.				
Sometimes does your imagination play tricks on you?	0	0	0	No information
What kind of tricks? Tell me more about them.				
Do you have any ideas about things that you don't tell anyone because they might not understand?	1	1	1	Definitely not delusional
What are they?	2	2	2	Suspected.
Do you have any secret thoughts? Tell me about them.				
Do you believe in other things that other people don't believe in? Like what?	**3**	**3**	**3**	Mild. Delusion definitely present but at times subject questions his false belief.
Is anybody out to hurt you ?				
Does anybody control your mind or body (like a robot)?	4	4	4	Moderate. Generally has conviction in his false belief.
Is anything happening to your body?				
Do you ever feel the world is coming to an end?				
Do you ever think you are an important or great person? Who?	5	5	5	Severe. Delusion has a significant effect on his actions, e.g., often asks family to forgive his sins, preoccupied with belief that he is a new Messiah
Are you sure that this (... ?) is that way?				
Could there be any other reason for it?				
How do you know that it happens as you say?	6	6	6	Extreme. Actions based on delusions have major impact on him or others, e.g., stops eating because believes food is poisoned
Any other possible explanations?				
Do you enjoy making up stories like this?				
Or is it different from making up stories?				
(you might suggest other possible explanations and see how the subject reacts to them)				

	PAST:	_____	_____	_____
		P	**C**	**S**

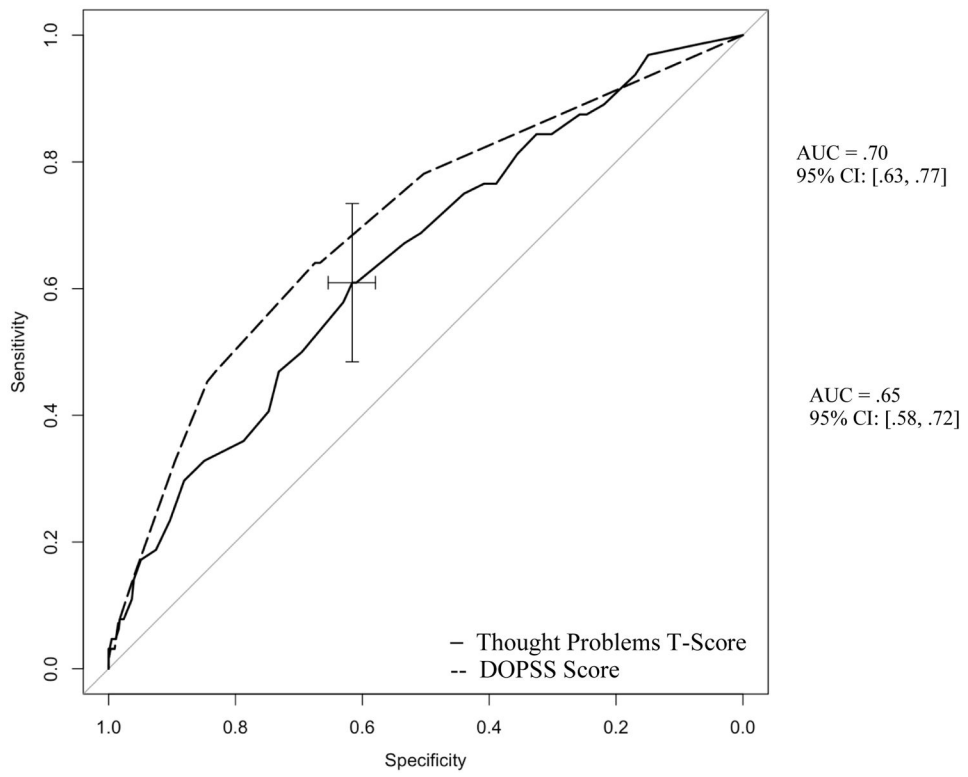| PSYCHOSIS			
HALLUCINATIONS	**P**	**C**	**S**
Do you think it could be your imagination ?			
What did you do about...?			

Figure 1.
Receiver Operating Characteristic (ROC) curves for Thought Problems *T*-Scores and DSM-Oriented Psychotic Symptoms (DOPSS) Scores.

Table 1

Descriptive Statistics for Clinical and Demographic Variables, and Bivariate Tests of Association with Clinically Concerning Psychosis Status

Variable	Any Clinically Concerning Psychosis (n=64)	No Clinically Concerning Psychosis (n=630)	Effect Size
	M (SD) or n (%)	M (SD) or n (%)	
Age	11.91[^] (2.92)	11.12 (3.26)	d=0.24
Female	34 (53%)[*]	246 (39%)	phi=0.08
Race (White %)	19 (30%)[*]	141 (22%)	phi=0.05
Number of Diagnoses	3.31[**] (1.69)[**]	2.58 (1.29)	d=0.54
Bipolar Disorder	19 (30%)[**]	95 (15%)	phi=0.11
Other Mood	23 (36%)	179 (28%)	phi=0.05
PTSD	10 (16%)[*]	49 (8%)	phi=0.08
Substance Use	5 (8%)	29 (5%)	phi=0.05
ADHD	35 (56%)	399 (64%)	phi=−0.05
Conduct Disorder	13 (20%)[*]	72 (12%)	phi=0.08
Oppositional Defiant Disorder	21 (32%)	254 (41%)	phi=−0.04
Anxiety Disorder	26 (42%)[**]	156 (25%)	phi=0.11
Thought Problems T-Score	69.92[***] (10.08)	64.63 (9.63)	d=0.54
DOPSS Score	2.53[***] (2.14)[***]	1.16 (1.54)	d=0.83

Note: Where data points were missing, effect sizes were calculated out of total number of available cases. Diagnoses describe whether the person met criteria for the disorder regardless of comorbidity, so the sum of cases adds to more than the subgroup n, and not all diagnoses tallied were the source of the clinically concerning psychosis. Cohen's d guidelines are 0.2 (small), 0.5 (medium), and 0.8 (large); 0.1 (small), 0.3 (medium), and 0.8 (large) are phi guidelines [31].

[^]
p=.07,

[*]
p<.05,

[**]
p<.005,

[***]
p<.0005, two-tailed, based on t-test (means), Levene's test (SD), or X^2 (proportions); effect sizes are Cohen's d (means) or phi (proportions).

Table 2

Areas Under the Curve (AUC)from Receiver Operating Characteristic analyses for Clinically Concerning Psychosis Status

Index Test	Area Under Curve	95% CI		Difference between AUCs
		Lower	Upper	
Any Clinically Concerning Psychosis				
Thought Problems *T*-Score	.65***	.58	.72	
DOPSS Score	.70***	.63	.77	*p* <.001

Note.

*
p<.05,

**
p<.005,

***
p<.001, two tailed.

Table 3

Different Optimal Threshold and Multi-Level Diagnostic Likelihood Ratios for CBCL Parent Thought Problems and DOPSS for Any *Clinically Concerning Psychosis*

	Thought Problems T-Score					**5% Prevalence**	
Cut Score	**Sensitivity**	**Specificity**	**Level**	**DiLR+**	**DiLR-**	**PPV**	**NPV**
Optimal threshold: 68.5+	.61	.62	.40	1.59	0.63	.08	.97
95% CI:				*(1.37 to 1.88)*	*0.46 to 0.86)*	*.067 to .090*	*.957 to .976*
Multilevel DiLRs			Sample %	DiLR		PPV	
50 to 58	--	--	29	0.52	--	.03	--
59 to 73	--	--	49	1.00	--	.05	--
74 to 82	--	--	20	1.49	--	.07	--
83+	--	--	3	3.28	--	.15	--

	DOPSS Scores					**5% Prevalence**	
Cut Score	**Sensitivity**	**Specificity**	**Level**	**DiLR+**	**DiLR–**	**PPV**	**NPV**
Optimal threshold: 1.67+	.64	.67	.35	1.97	0.53	.08	.98
95% CI:				*1.65 to 2.28*	*0.39 to 0.75*	*.080 to .107*	*.962 to .980*
Multilevel DiLRs			Sample %	DiLR		PPV	
0 to 1	--	--	65	0.53	--	.03	--
2 to 3	--	--	23	1.42	--	.07	--
4	--	--	8	2.81	--	.13	--
5+	--	--	5	3.69	--	.16	--

Note: Optimal threshold=maximum combined sensitivity and specificity. Sensitivity=portion of cases with psychotic condition correctly identified; specificity=portion of cases without without psychotic condition diagnoses correctly identified; level=portion of sample testing above threshold; DiLR+=diagnostic likelihood ratio associated with a positive test, DiLR-=diagnostic likelihood ratio associated with a negative test result; PPV=positive predictive value, or the portion of test positives that have psychotic condition; NPV=negative predictive value, or the portion of test negatives that do not have a psychotic condition. The prevalence rates of 5% represents what might typically be encountered in outpatient mental health, providing a sense of how the instruments might work in practice.