



HHS PUBLIC ACCESS

Author manuscript

J Consult Clin Psychol. Author manuscript; available in PMC 2017 April 01.

Published in final edited form as:

J Consult Clin Psychol. 2016 April ; 84(4): 323–333. doi:10.1037/ccp0000070.

A Randomized Controlled Trial of Cognitive Debiasing Improves Assessment and Treatment Selection for Pediatric Bipolar Disorder

Melissa M. Jenkins, PhD^{1,2} and **Eric A. Youngstrom, PhD³**¹Department of Psychiatry, University of California San Diego, San Diego, CA²Research Services, VA San Diego Healthcare System, San Diego, CA³Departments of Psychology and Neuroscience and Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract

Objective—This study examined the efficacy of a new cognitive debiasing intervention in reducing decision-making errors in the assessment of pediatric bipolar disorder (PBD).

Method—The study was a randomized controlled trial using case vignette methodology. Participants were 137 mental health professionals working in different regions of the US ($M=8.6\pm 7.5$ years of experience). Participants were randomly assigned to a (1) brief overview of PBD (control condition), or (2) the same brief overview plus a cognitive debiasing intervention (treatment condition) that educated participants about common cognitive pitfalls (e.g., base-rate neglect; search satisficing) and taught corrective strategies (e.g., mnemonics, Bayesian tools). Both groups evaluated four identical case vignettes. Primary outcome measures were clinicians' diagnoses and treatment decisions. The vignette characters' race/ethnicity was experimentally manipulated.

Results—Participants in the treatment group showed better overall judgment accuracy, $p < .001$, and committed significantly fewer decision-making errors, $p < .001$. Inaccurate and somewhat accurate diagnostic decisions were significantly associated with different treatment and clinical recommendations, particularly in cases where participants missed comorbid conditions, failed to detect the possibility of hypomania or mania in depressed youths, and misdiagnosed classic manic symptoms. In contrast, effects of patient race were negligible.

Conclusions—The cognitive debiasing intervention outperformed the control condition. Examining specific heuristics in cases of PBD may identify especially problematic mismatches between typical habits of thought and characteristics of the disorder. The debiasing intervention was brief and delivered via the Web; it has the potential to generalize and extend to other diagnoses as well as to various practice and training settings.

Corresponding Author: Dr. Eric A. Youngstrom, Departments of Psychology and Neuroscience and Psychiatry, Center for Excellence in the Research and Treatment of Bipolar Disorder, University of North Carolina, CB #3270, Davie Hall, Chapel Hill, NC 27599-3270. ey@unc.edu.

Conflict of Interest Disclosures:

Dr. Jenkins has no conflicts of interest to disclose.

Keywords

Cognitive Debiasing; Decision-Making; Pediatric Bipolar Disorder

Pediatric bipolar disorder is an extremely challenging diagnosis (Leibenluft, 2011; Marchand, Wirth, & Simon, 2006; Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009). There often are long delays in diagnosing youth with PBD, with symptoms often being misdiagnosed as something else (Marchand et al., 2006). Paradoxically, bipolar also is often a “false positive” diagnosis, clinically assigned and treated when the youth actually has a different disorder (Leibenluft, 2011; Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009). In short, clinicians miss true cases of bipolar disorder and diagnose PBD in many cases where the youth does not actually have the disorder. Thus, the risk for cognitive error is especially high in clinicians’ assessment of bipolar symptomatology. The medical decision-making literature has identified over *thirty* faulty heuristics, biases, and cognitive distortions (Croskerry, 2005; Elstein & Schwartz, 2002). Examining specific heuristics in cases of PBD may identify especially problematic mismatches between typical habits of thought and characteristics of the disorder. Teaching clinicians cognitive debiasing strategies may improve their decision-making and help reduce the frequency of misdiagnosis around PBD.

Several factors challenge diagnostic decisions and reliability of clinical impressions for PBD. First, case presentations appear differently depending on whether a client is experiencing florid mania, severe depression, mixed or a euthymic state (Youngstrom, 2005). Consequently, clinicians can fail to detect hypomania or mania in youth with depression or youth experiencing a euthymic state, neglecting the possibility of PBD in their diagnoses, case conceptualizations, and treatment plans. When this happens, clinicians commit *base-rate neglect*--failing to adequately take into account the prevalence of a particular disease. Epidemiological studies suggest that bipolar is more common than previously thought (Hasin, Goodwin, Stinson, & Grant, 2005; Merikangas, et al., 2007; Merikangas & Pato, 2009), especially in clinical settings (Biederman, et al., 1996; Blader & Carlson, 2007; Geller, et al., 2002; Youngstrom, Youngstrom, & Starr, 2005). Further, a higher proportion of people with depression will actually have bipolar disorder, particularly child populations (Angst, Sellaro, Stassen, & Gamma, 2005). Both of these trends worsen the impact of base rate neglect.

Second, symptom overlap makes it difficult to discern symptoms of bipolar from other likely suspects in pediatric populations. For example, irritability can be a symptom of bipolar, ADHD, unipolar depression, and ODD. Clinicians are vulnerable to *diagnosis momentum* (also known as *diagnostic creep*): the tendency for a particular initial or provisional diagnosis to become established without adequate evidence. Research indicates clinicians overestimate the likelihood of bipolar without adequate evidence (Dubicka, Carlson, Vail, & Harrington, 2008; Jenkins, et al., 2011). They may be framing this judgment quickly on the basis of a few features, and then fail to check whether a sufficient number of criteria are met. Similarly, once the bipolar hypothesis gains momentum, disconfirming evidence may be discounted or ignored. Clinicians, like many people, are

prone to resist modifying or discarding ideas despite further evidence to the contrary (Dawes, Faust, & Meehl, 1989; Meehl, 1954).

Third, due to high comorbidity, clinicians may not see a “typical” uncomplicated bipolar presentation (Biederman, Klein, Pine, & Klein, 1998; Youngstrom, Findling, et al., 2005). For example, individuals with bipolar have high rates of ADHD (Galanter & Leibenluft, 2008; Kessler, et al., 2005; Lewinsohn, Klein, & Seeley, 1995; Singh, DelBello, Kowatch, & Strakowski, 2006). As a result of overlapping symptomatology and high comorbidity rates, clinicians are prone to commit *search satisficing*--when one calls off a search once something is found (i.e., finding something may be satisfactory but not finding everything is suboptimal (see Croskerry, 2002)). In the case of PBD and ADHD, clinicians may “call off the search” when they diagnose one of these disorders, causing them to miss the other disorder for youths who actually meet DSM criteria for both conditions. Clinical “diagnosis as usual” detects one less diagnosis per case on average than found using a more structured interview approach with the same cases (Jensen-Doss, Youngstrom, Youngstrom, Feeny, & Findling, 2014; Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009).

Fourth, given the evidence suggesting frequent overdiagnosis in the U.S. (Dubicka, Carlson, Vail, & Harrington, 2008; Soutullo et al., 2005), clinicians may show exaggerated certainty in their decision-making. This tendency is consistent with *over-confidence bias*, or thinking one knows more than one does. Overconfidence bias can lead to delayed or missed diagnosis and unwarranted interventions (Croskerry, 2002).

Finally, another challenge is the frequent misdiagnosis of minority populations (Garland et al., 2005). For example, African American pediatric populations are more likely to receive diagnoses of conduct disorder or schizophrenia (DelBello, Lopez-Larson, Soutullo, & Strakowski, 2001; Fabrega, Ulrich, & Mezzich, 1993; Kilgus, Pumariega, & Cuffe, 1995), although currently no research indicates that the risk of developing bipolar disorder varies by race/ethnicity (American Psychiatric Association, 2001; Van Meter, Moreira, & Youngstrom, 2011). Race/ethnicity bias is associated with misdiagnosis, delay in diagnosis and treatment, and stigma.

Cognitive Debiasing Strategies as a Possible Intervention to Improve Clinical Diagnosis

At present, there is little empirical information about cognitive debiasing strategies in mental health (Galanter & Patel, 2005). Although these strategies are relatively innovative for psychology in the sense that they are not formally organized or taught as they are presently in other medical areas (Croskerry, 2003), experts have repeatedly recommended training mental health professionals in decision-making (Arkes, 1991; Baker-Ericzen, Jenkins, Park, & Garland, 2014). Judgment does not necessarily improve with additional experience (Lueger, 2002; Spengler et al., 2009), and more experienced clinicians may be at greater risk for biased judgment than novice clinicians (Strohmer & Leierer, 2000). Overall, research needs to identify effective ways to teach mental health professionals how to avoid cognitive pitfalls (Regehr & Norman, 1996). Simply telling individuals to avoid particular biases is not sufficient to change behavior (Arkes, 1991; Kurtz & Garfield, 1978).

Several cognitive debiasing strategies significantly reduce diagnostic error in medical settings (Croskerry, 2003; Gigerenzer & Goldstein, 1996). These strategies could potentially help with PBD. *Developing insight and awareness* involves providing detailed descriptions of known cognitive biases, together with multiple clinical examples illustrating their adverse effects on diagnosis (Croskerry, 2003). In the case of PBD, this would involve education on key heuristics and biases (e.g., base-rate neglect, search satisficing, diagnosis momentum, overconfidence bias and race/ethnicity bias), and using varied clinical scenarios to highlight possible outcomes (e.g., misdiagnosis, delay in diagnosis). Another key strategy, *making the task easier*, entails providing more information about the specific problem to reduce task difficulty and ambiguity (Croskerry, 2003). Here, mood graphs can illustrate different mood states of bipolar disorder and highlight the diagnostic challenges associated with the heterogeneity of the condition. For example, low or depressed mood can lead to base-rate neglect if clinicians fail to assess for past episodes of hypomania and mania. Mood graphs (e.g., Bieling et al., 2003) help both clinicians and patients develop insight and awareness, as well as communicate more clearly.

Considering alternatives is the procedure of establishing forced consideration of alternative possibilities (Croskerry, 2003). Symptom checklists can help clinicians consider alternative diagnoses by decreasing reliance on intuition in clinical problem solving (Ely, Graber, & Croskerry, 2011). For PBD, symptoms checklists can guard against search satisficing, helping clinicians assess for bipolar disorder in cases of comorbid ADHD and vice versa.

Decreasing reliance on memory entails improving the accuracy of judgments through cognitive aids such as mnemonics (Croskerry, 2003). GRAPES is a helpful mnemonic acronym for symptoms more specific to bipolar disorder, standing for Grandiosity, Racing thoughts, changes in Activity, Pressured speech, Elated mood, and decreased need for Sleep. Decreasing reliance on memory protects against several cognitive errors, particularly diagnosis momentum and overconfidence bias (Arkes, 1981).

Cognitive forcing strategies help avoid predictable bias in particular clinical situations (Croskerry, 2003). This strategy may protect against race/ethnicity bias. Having clinicians routinely assess for mood symptoms in suspected cases of schizophrenia and conduct disorder could decrease misdiagnosis of African Americans.

Statistical methods also can decrease judgment biases and errors (Aegisdottir et al., 2006; Croskerry, 2003; Grove, Zald, Lebow, Snitz, & Nelson, 2000). In fact, “actuarial” statistical prediction methods that use Bayes’ theorem to estimate the probability of a diagnosis based on test findings or clinical observations have been shown to be more accurate and efficient for diagnosing PBD (Jenkins, Youngstrom, Washburn, & Youngstrom, 2011; Jenkins, Youngstrom, Youngstrom, Feeny, & Findling, 2011). Clinicians can implement these methods in ways that require no mathematical computation on the part of the clinician (see Jenkins et al., 2011). Unfortunately, these methods are not routinely taught or used in psychological practice; and even these methods require clinical judgment to maximize their clinical utility (Elstein & Schwartz, 2002; Youngstrom, Freeman, & Jenkins, 2009).

More generally, experts recommend opportunities for *metacognition* and *simulation* (Croskerry, 2003). In short, decision-making benefits from stepping back from the current task and examining one's thinking process (i.e., reflecting on one's approach before making a PBD diagnosis or frequently co-occurring condition) as well as engaging in mental rehearsals. The latter involves practice with different clinical scenarios involving PBD and also experience using actuarial methods.

The present study tested the efficacy of a new intervention designed to improve clinical judgment in the assessment of PBD by educating participants about common cognitive pitfalls and training them in recommended debiasing strategies. We hypothesized that participants' diagnostic judgments in the treatment group would show higher overall decision accuracy than participants' judgments in the control condition. Additionally, we hypothesized participants in the treatment group would commit significantly fewer diagnostic errors and show significantly higher diagnostic accuracy than participants in the control group on individual vignettes constructed to test for specific cognitive errors (e.g., base rate neglect, search satisficing, diagnosis momentum). A secondary aim explored the impact of diagnostic decisions on treatment decisions. Specifically, we predicted that employing a faulty heuristic (as demonstrated by suboptimal diagnostic decisions) would affect clinicians' decisions regarding next clinical action, suggesting that heuristics change treatment as well as assessment formulations. An exploratory goal of the study investigated the role of race/ethnicity bias in clinicians' diagnostic decisions, as the vignette design offered an opportunity to change race while holding all other features constant.

Methods

The Institutional Review Board at the University of North Carolina approved this study.

Participants

Participants were 137 mental health providers: 56 in the treatment group and 81 in the control group ($M = 8.6 \pm 7.8$ years of practice experience). Eligible participants were licensed clinicians or currently supervised by a licensed mental health professional and had experience treating pediatric patient populations for psychiatric issues. Table 1 provides information about participant demographics. Table 2 provides details about participants' professional backgrounds. Twenty-three participants reported attending a previous training in the assessment of PBD (e.g., continuing education seminar). Distribution of these participants between the control ($n = 13$) and treatment ($n = 10$) conditions was balanced, $p > .05$. Recruitment took place between 2011 and 2013, ending when the recruitment goal was met. Participants were recruited across the US using study fliers, listserv announcements, and by word-of-mouth. Figure 1 depicts the Trial Flow Diagram.

Materials

Four case vignettes were used. Each vignette had features designed to test for a specific cognitive error associated with the assessment of PBD. Vignette characters were experimentally manipulated to test possible race/ethnicity bias; vignettes described youths

who were African American half of the time and Caucasian half of the time, with all other details the same.

Vignette 1: Base rate neglect—This vignette described a 14-year-old female with current symptoms characteristic of a major depressive episode. Responses that included depression and inquired about hypo/mania were coded as accurate, whereas responses that included depression but missed the possibility of mania/hypomania were coded as “somewhat accurate” because they missed the recommendation to systematically check for history of hypomania or mania when working with depression, particularly early onset depression (American Psychiatric Association, 2013; McClellan, Kowatch, & Findling, 2007). Inaccurate responses missed depression along with consideration of mania symptoms.

Vignette 2: Search satisficing—This vignette described a 7-year-old male with mania and ADHD symptoms. Responses that considered both ADHD and a bipolar disorder were coded as accurate, versus responses that included only ADHD or a bipolar disorder, which were coded as somewhat accurate. Inaccurate responses missed ADHD and bipolar disorder.

Vignette 3: Diagnosis momentum—This vignette portrayed a 10-year-old male with bipolar symptoms, a moderately highly elevated test score on a widely used assessment instrument, and a family history of bipolar disorder. After providing a probable diagnosis and recommending next clinical actions, participants rated the probability of a bipolar diagnosis (from 0 to 100). In an actuarial approach, the combination of assessment findings yields a moderate Bayesian probability of a bipolar diagnosis (i.e., a point estimate of 27% probability).

Vignette 4: Classic mania—This vignette focused on an 11-year-old female with classic symptoms of mania including grandiosity, hypersexuality as an example of disinhibited and risky behavior, psychomotor agitation, and distractibility – meeting criteria for at least four “B Criterion” symptoms in addition to the episodic disturbance of mood (per DSM-IV-TR) (American Psychiatric Association, 2001). Consistent with the episodic nature of mood disorders, her symptoms were described as intermittent. This youth met duration criteria for a manic episode (e.g., 1 week or longer), and experienced impairment as a result of her symptoms. This vignette was similar to a vignette used in Dubicka, Carlson, Vail, & Harrington (2008).

Measures

Demographic and background information—Participants provided demographic information and professional information (e.g., years of clinical experience). See Tables 1 and 2.

Diagnostic accuracy—Participants reported the youth's probable diagnoses after reading each case vignette. These diagnoses were scored on a 3-point Likert scale of (1 = inaccurate diagnosis, 2 = somewhat accurate, and 3 = accurate diagnosis) using a priori criteria

established by expert diagnosticians coding features of the vignette against DSM-IV criteria. Participants could also indicate other diagnoses that they were considering.

Decision-making error status—“Inaccurate” and “somewhat accurate” diagnoses counted as committing a decision-making error; “accurate” responses did not.

Risk estimate accuracy—For the vignette evaluating “diagnosis momentum,” participants’ estimates were compared to an actuarial estimate: the Bayesian posterior probability of the likelihood of bipolar disorder (Jenkins et al., 2011; Youngstrom & Duax, 2005; Youngstrom et al., 2009). The Bayesian estimate was 27% ($\pm 5\%$) based on: clinical setting (e.g., the youth was seen in outpatient clinic, so the recommended starting base rate of PBD is 6%); family history information (e.g., the youth had a second degree relative with bipolar disorder, increasing the odds of a bipolar diagnosis by a factor of 2.5), and a test score of 35 on the Parent General Behavior Inventory (PGBI; which is associated with a diagnostic likelihood ratio of 2.3) (Youngstrom et al., 2004). Estimates of 27 or estimates that fell within $\pm 5\%$ of the Bayesian optimal estimate (from 22 to 32) (Sedlmeier & Gigerenzer, 2001) indicated no decision-making error. Estimates that fell outside of a $\pm 5\%$ range (less than 22 or greater than 32) were classified as committing a decision-making error. For diagnostic accuracy, a 3-point response was 27%, a 2-point response was between 22% and 32%, and a 1-point response was less than 22% or greater than 32%.

Overall diagnostic accuracy—Overall diagnostic accuracy was a composite measure, averaging diagnostic accuracy and risk estimate accuracy across all case vignettes.

Procedure

Study administration was web-based, combining Qualtrics (Qualtrics, 2014) and an automated Powerpoint presentation with narration. Clinicians accessed the study through a private Web address and password. Qualtrics randomly assigned participants to the control condition (brief presentation on PBD) or treatment condition (the brief presentation on PBD + the cognitive debiasing intervention). The intervention consisted of two main parts: education on common cognitive pitfalls in the assessment of PBD (e.g., base-rate neglect; search satisficing; diagnosis momentum; race/ethnicity bias); and, training and tools to avoid cognitive pitfalls when assessing for PBD. The training and tools consisted of reminders to consider the alternatives (e.g., symptom checklists, routinely asking “what else might this be”), metacognition (e.g., examining one's own decision-making during and after assessment, prior to making diagnosis), decrease reliance on memory (e.g., mnemonics to remember bipolar symptoms), actuarial approaches (e.g., diagnostic likelihood ratios and Bayesian reasoning), and simulation (e.g., opportunities to practice decision-making with case information).

After watching their respective presentations, participants completed four case vignette exercises. Qualtrics randomized the presentation order of case vignettes. Participants were instructed to: (i) decide the probable diagnosis from a pull down menu of 26 diagnosis options; and, (ii) recommend next clinical action (e.g., more assessment, psychotherapy, medication, no treatment, other). Participants could select one or two initial treatment

methods (Currin, Schmidt, & Waller, 2007). After clinicians selected a probable diagnosis, they were asked if they were considering any additional diagnoses. This provided participants an opportunity to report any comorbid disorders and provide differential diagnosis. Additional diagnoses were incorporated in coding diagnostic accuracy (as described above).

Data Analytic Procedure

Data were screened to ensure quality and to check standard statistical assumptions. The amount of missing data was small (<2%), and data were excluded listwise (Allison, 2002). Multiple linear regression models tested treatment condition, controlling for years of clinical experience, as predictors of participants' overall performance both in terms of the total number of decision-making errors and the accuracy of diagnostic decisions across all vignettes. Chi-square examined: (1) the association between vignette characters' race/ethnicity and participants' diagnostic accuracy, and (2) associations between decision-making error and practice implications (regardless of condition status). For individual vignettes, logistic regression tested whether treatment status changed the probability of decision-making errors for each vignette. Polytomous Universal Model (PLUM) regression modeled diagnostic accuracy, with years of clinical experience, professional title, and condition status as the IVs. ANCOVA compared risk estimates between treatment groups, controlling for clinical experience. Sensitivity analyses examined whether the results changed if not controlling for years of experience, or also controlling for type of professional.

There was adequate power to detect effects for all primary analyses. A sensitivity analysis using G*Power (Faul, Erdfelder, Buchner & Langm 2009) indicated 80% power to detect effect sizes of $f^2 = .07$ or larger (Cohen, 1988) for the given sample size ($N = 137$) and $\alpha = .05$, 2-tailed. Cohen described effect sizes of $f^2 .02$ as "small," $.15$ as "medium," and $.35$ as "large." Statistical analyses were performed using IBM Statistical Package for the Social Sciences (SPSS) Version 21.0. We reported effect sizes as phi for χ^2 with 1 *df*, Nagelkerke R^2 for multiple *df* logistic and polytomous regressions, R^2 for multiple regressions, and Cohen's d when comparing means.

Results

Sample Characteristics and Randomization Check

The treatment and control groups showed no significant differences, $p > .05$, on any demographic or professional variables with the exception of their primary professional activities, $\chi^2 (5) = 12.29$, $p = .03$. See Tables 1 and 2.

Overall Diagnostic Accuracy

Clinicians in the intervention condition made significantly fewer errors across all vignettes ($M = 1.5 \pm 1.2$ versus $M = 3.0 \pm 0.9$), with a large effect size, $R^2 = .36$, $p < .001$ and Cohen's $d = 1.67$. Intervention remained significant when controlling for experience; experience and professional type did not contribute to prediction of errors.

For diagnostic accuracy, assignment to the training condition made a significant unique contribution even after controlling for years of experience, $R^2 = .36, p < .001$. Average accuracy was 9.9 ± 1.7 in the training condition versus 7.6 ± 1.2 in the control, $d = 1.08, p < .0005$. Years of clinical experience or professional type did not make a significant unique contribution, $ps > .18$.

Sensitivity to Possible Hypomania or Mania

Using a stringent definition, 23% in the treatment group committed a decision-making error, versus 69% in the control group, $\chi^2(1) = 29.11, phi = .46, p < .001$. In terms of diagnostic accuracy, the overall chi-square for the PLUM regression was significant, $\chi^2(3) = 34.04, p < .001$, with a Nagelkerke $R^2 = .28$. Treatment significantly improved diagnostic accuracy in the expected direction, $B = -2.23(1), p < .001$. Years of clinical experience and professional type did not make significant unique contributions, $ps > .05$.

For next action, 82% of participants chose assessment, 69% therapy, 26% medication, and 3% no further clinical action necessary. When participants considered the possibility of hypo/mania in their diagnostic reasoning, they were significantly more likely to recommend further assessment, $\chi^2(1) = 8.04, phi = .24, p < .05$. There were no significant associations between decision making error status and participants' likelihood to recommend therapy, $p > .05$, or medication, $p > .05$. Chi-squared revealed no significant association between decision-making error and antidepressant medication, $p = .08$.

Detecting Comorbid Conditions

In the treatment group, 29% committed a decision-making error, versus 59% in the control group, $\chi^2(1) = 12.83, phi = .31, p < .001$. The overall chi-square for the PLUM regression was significant, $\chi^2(3) = 11.66, p = .009$, with a Nagelkerke $R^2 = .11$. Debiasing treatment significantly improved diagnostic accuracy, $B = -1.26(1), p = .001$; years of clinical experience did not make a significant unique contribution, $p > .05$, nor did professional title, $p > .05$.

For this vignette, 82% participants selected more assessment, 45% medication, and 38% therapy. When participants failed to consider a comorbid diagnosis of bipolar disorder or ADHD, they were significantly more likely to recommend medication or a medication referral, $\chi^2(1) = 13.09, phi = .31, p < .001$. When participants detected the possibility of a comorbid condition, they were less likely to recommend therapy, $\chi^2(1) = 4.08, phi = .17, p > .05$, and more likely to recommend additional assessment $\chi^2(1) = 17.08, phi = .35, p > .001$.

Risk Estimates and Diagnosis Momentum

Participants in the treatment condition showed significantly lower diagnosis momentum, $p < .001$, even when controlling for years of experience, $F(2,128) = 8.39, p < .001, R^2 = .12$. In the control group, participants' estimates of the probability of a bipolar diagnosis ($M = 46.4, SD = 20.7$) were significantly higher than 27% for this case ($t(80) = 8.44, p < .001$). Control group participants' estimates also were significantly higher than the "near miss" upper bound of 32% ($27\% + 5\%$), $p < .001$.

The treatment group's estimates ($M = 34.2$, $SD = 15.4$) tended to be higher than the Bayesian estimate as well, $t(55) = 3.50$, $p < .05$. However, the treatment group's average was not significantly different than the "close" limit, $t(55) = 1.07$, $p = .29$. The cognitive debiasing intervention lead to less cognitive error in synthesizing the case information, $X^2(1) = 49.36$, $phi = .60$, $p < .001$, based on logistic regression predicting estimates in the "close" range. There was no significant association between decision-making error and clinical actions, $p > .05$; 93% of all participants chose therapy, 48% assessment, and 28% medication.

Classic Mania

The intervention reduced decision-making errors (61% versus 82% in the controls), $\chi^2(1) = 7.17$, $phi = .23$, $p = .007$. When controlling for years of experience and professional type, the overall model was significant, $\chi^2(3) = 8.58$, $p = .035$, Nagelkerke $R^2 = .07$, and the intervention still predicted fewer errors, $B = -.98(1)$, $p = .005$.

For next clinical action, 87% selected more assessment, 70% therapy, and 26% medication. When participants failed to detect bipolar disorder they were less likely to recommend medication, $\chi^2(1) = 11.09$, $phi = .28$, $p < .005$, and more likely to recommend therapy, $\chi^2(1) = 14.07$, $phi = .32$, $p < .001$.

Race/Ethnicity Bias

The vignette characters' race/ethnicity was not associated with participants' diagnostic accuracy, largest $\chi^2 = .28$, all $p > .05$. Effect sizes for the race/ethnicity variable were all small, $d = .01$ to $.28$, where $.2$ is conventionally considered "small."

Discussion

The goal of the present study was to see whether a brief training intervention could yield measurable improvement in the accuracy of clinical diagnostic assessment in practicing clinicians working with a standardized set of clinical vignettes. The study used a randomized trial design to test the effects of cognitive debiasing training on clinical decision-making about PBD and frequently co-occurring conditions. We selected pediatric bipolar disorder as the focus because (a) the diagnosis has been controversial, (b) the research base has changed rapidly in this area (e.g., more than 90% of 7700+ hits in PubMed were published in the last 15 years), (c) there are big differences in training around the topic (Dubicka et al., 2008), leading to pockets of very high or low diagnosis rates, and low overall accuracy comparing clinical diagnoses to research interviews (Jensen-Doss, Youngstrom, Youngstrom, Feeny, & Findling, 2014; Rettew et al., 2009) and (d) we could identify multiple cognitive heuristics likely to play a role in evaluating youths with mood and behavioral issues. Randomization assigned participating mental health professionals in a balanced manner in terms of professional training and experience, and the training intervention produced statistically significant and large effect sizes improving most outcome measures.

We measured overall decision accuracy in two ways: total decision-making errors and diagnostic accuracy ratings across all four vignettes. In both cases, the treatment group performed significantly better than the control group. Participants who received the cognitive debiasing intervention committed fewer decision-making errors and generated

more accurate diagnoses for multiple vignettes with various case presentations. Along with providing empirical support for cognitive debiasing interventions in mental health, our findings highlight the complex task clinicians regularly face when presented with new case information. Decades of research document problems associated with clinical judgment (Aegisdottir et al., 2006; Grove et al., 2000; Spengler et al., 2009); the present study offers a promising new approach for improving clinical judgment and equipping clinicians with relatively simple yet effective strategies to decrease the likelihood of cognitive error.

Despite the superiority of the cognitive debiasing intervention, participants in both conditions appeared to struggle the most with the vignette depicting a youth with classic mania. This outcome is surprising given that this vignette, in many ways, was the most straightforward: The portrayed youth met all DSM-IV diagnostic criteria. Also, all participants were primed for PBD just prior to responding to case vignettes by virtue of the automated PowerPoint presentation. It is possible clinicians are prone to seeking more benign diagnoses or explanations for certain symptoms and behaviors (e.g., adjustment disorder), perhaps hesitant to diagnose mania in younger patient populations (i.e., the vignette character was 11 years old). This is concerning given the substantial delay in the community between onset of symptoms and diagnosis for individuals with bipolar illness (Marchand et al., 2006). More research is needed to better understand how the intervention can be tailored to further reduce diagnostic error around more classic presentations of bipolar illness.

A second goal of the present study involved testing the potential influence of decision accuracy on clinician's recommended treatment approaches. Similar to many psychiatric disorders, accurate diagnosis of PBD is important before starting treatment (Kowatch et al., 2005; Weller, Danielyan, & Weller, 2004). Suboptimal diagnostic decisions were significantly associated with different treatment and clinical recommendations, particularly when participants missed comorbid conditions, failed to detect the possibility of hypomania or mania in depressed youths, and misdiagnosed classic manic symptoms. The vignette depicting more classic mania symptoms was diagnostically challenging for clinicians regardless of intervention condition status; but participants were more likely to recommend medication when participants detected bipolar disorder. In this case, an accurate diagnosis aligned with recommended treatment guidelines in the literature (McClellan, Kowatch, Findling, & Work Group on Quality, 2007). Further, for the youth presenting with bipolar and ADHD symptoms, participants who failed to consider a comorbid condition were more likely to recommend medication. In instances where clinicians detected the ADHD but not the bipolar and recommended medication, the treatment plan would be at odds with the recommendations of current practice parameters, which suggest focusing on stabilization of mood first, followed by augmentation with medication addressing residual ADHD symptoms if necessary (McClellan et al., 2007).

In contrast, participants who accurately detected both conditions were more likely to select more assessment, which would be helpful in determining the optimal sequence of interventions – including in this case key decisions about whether to focus first on ADHD symptoms or mood issues. Interestingly, cognitive error, particularly search satisficing, appears associated with premature treatment decisions, while more accurate judgment leads

to a more conservative and methodical approach. In other words, when one identifies “warning signs” of a co-occurring condition (e.g., sleep disturbance or episodic mood and energy disturbance in a case of ADHD), it may appropriately slow down the diagnostic process, allowing for more comprehensive evaluation of alternative or comorbid conditions prior to initiating active treatment (e.g., medication) (Youngstrom, Choukas-Bradley, Calhoun, & Jensen-Doss, 2014). Taken together, faulty heuristics appear to affect clinicians’ treatment formulations and may have important implications for patient care, especially in light of research indicating diagnostic accuracy may be an important precursor to successful treatment (Jensen-Doss & Weisz, 2008).

Contrary to study hypotheses, participants’ diagnoses were not affected by the race/ethnicity of the vignette character. Specifically, clinicians’ rates of bipolar diagnoses did not vary as a function of the youth being African American or Caucasian, with all other case information held equal by design. It is worth noting that in addition to describing the youth's race/ethnicity in the case vignettes, we included a picture to make race/ethnicity more salient to study participants.

Our non-significant race/ethnicity findings are intriguing given the tendency in the community for African American youths to be misdiagnosed with conduct disorder or schizophrenia (DelBello et al., 2001; Fabrega et al., 1993; Kilgus et al., 1995). It is possible that the misdiagnosis of African American youth is more complex than the mechanisms probed in the vignette, which manipulated the racial/ethnic label and appearance of the case. Recent ethnographic studies reveal differences in how minority populations describe mood symptoms, often emphasizing behavioral issues versus emotional problems (Carpenter-Song, 2009). Parental beliefs about causes of children's behavior also vary as a function of race/ethnicity, influencing how parents conceptualize and describe their children's problems to clinicians (Yeh et al., 2005). For example, when a family's description of the presenting problem focuses on behavior, clinicians formulate a hypothesis focused on a disruptive behavior disorder and fail to explore mood issues systematically as a competing hypothesis. Thus, cultural differences in description of the presenting problem may trigger availability heuristics, confirmation bias, and search satisficing to produce the discrepant assessment and treatment choice outcomes. The mere experimental manipulation of youths as Caucasian or African American in the present study unlikely captured this phenomenon, although it does suggest that perceptual biases or implicit racism are unlikely to be major factors in misdiagnosis. Additional mixed-methods research could elucidate the role, forms and effects of bias (Carpenter-Song, Whitley, Lawson, Quimby, & Drake, 2011).

Limitations

Several limitations warrant attention. First, external validity of studies with case vignettes can be limited because vignettes may not fully reflect complex interpersonal situations and contextual pressures present in real life diagnosis. Case vignettes made it possible to sample a broader group of clinicians, and it also made it possible to experimentally manipulate both the treatment and the case's race while holding other clinical features constant. A second limitation was the restricted number of cognitive errors and debiasing techniques the study could feasibly investigate (cf. Croskerry, 2002). We compensated for this limitation by

carefully selecting heuristics based on a thorough literature review and conceptualizing which heuristics were likely to be commonplace (Galanter & Patel, 2005). By focusing on heuristics likely to operate in more cases, it is possible to generate good coverage with a smaller number of targets (i.e., Pareto's 80:20 Rule, the "Law of the Vital Few"). Additional research needs to examine other possible heuristics as well as to test a broader array of debiasing strategies. Third, randomization resulted in an unequal number of participants in the treatment and control conditions. Future investigations could benefit from urn randomization or other blocking methods that would promote balanced assignment between arms or within strata (e.g., Project MATCH Research Group, 1997). Additionally, a substantial number of individuals were excluded from participation, with 17 declining and 107 initially viewing the website but not continuing in the study. Due to the online data collection methods employed, it was not possible to examine the impact of study attrition and potential selection bias, such as characteristics of excluded participants. These individuals showed initial interest by viewing the study website but provided limited or no demographic or professional characteristic information. An unknown portion also may have experienced technical difficulties if they were using an older internet browser or version of Adobe Flash Player™. Thus, we know very little about the subset of individuals who chose not to participate.

Further, the possibility of self-selection bias and priming warrant attention. Regarding the former, participants may reflect a subset of clinicians particularly eager to learn new assessment methods, potentially decreasing generalizability of study findings. However, the same mechanism may also have resulted in diminished effect sizes if the participants in the control condition were more assessment savvy than the average clinician. Similarly, all participants were primed for childhood mood disorders regardless of group assignment, potentially increasing participants' sensitivity to bipolar disorder in their diagnostic decisions. Again, the effect of this priming would be to increase the sensitivity of control participants to bipolar features, akin to enhancing the rate of placebo response in a treatment study, thus attenuating treatment effects. The fact that the debiasing strategies continued to show statistically significant and large treatment effects attests to the promise of the approach. It will be important for future work to examine the extent to which initial treatment effects are retained over time (e.g., recency effects and forgetting). These research questions are intriguing and also crucial in future development, testing, and dissemination of debiasing interventions for mental health professionals.

Additional limitations in terms of generalizability include the small number of male and African American participants, the low representation of clinicians from the Northeast, and a high percentage of participants indicating a cognitive behavioral therapy theoretical orientation, all of which may differ from the general population of practicing clinicians.

Future Directions

Overall, findings showed that a brief (<30 minute) intervention could be delivered via the Web to practicing clinicians. The debiasing intervention produced medium to large gains in the accuracy of diagnostic and treatment decisions across multiple vignettes. Results also

were impressive for showing improvement around a complicated and contentious diagnosis – pediatric bipolar disorder.

Future work should focus on cognitive debiasing for improving mental health decision-making in general. Research needs to examine the effectiveness of online teaching resources for mental health practitioners as well as the best media for presenting information to audiences. Several models for Web-based learning courses (e.g., Trauma Focused Cognitive Behavioral Therapy *Web*, Medical University of South Carolina, <http://tfcbt.musc.edu/>) are paving the way for scaling up and disseminating mental health interventions (Kazdin & Blase, 2011). When evaluating the effectiveness of web-based interventions, it is important to measure clinicians' adoption of strategies into practices. The present study demonstrated this with vignettes, but future studies should examine whether this translates into change with subsequent cases. Also, it may be beneficial to explore deeper the reasons behind error-prone behavior in clinical diagnosis. This line of research could inform future theory regarding decision-making and mental health practice. Strategies provided in the cognitive debiasing intervention are likely to generalize to other aspects of patient care, including the assessment of other mental disorders and treatment decisions. Thus, although the current study concentrated on PBD, these strategies may increase clinicians' awareness of common cognitive-based errors and teach them new ways of thinking. Further examination of generalizability to other challenging psychiatric diagnostic and treatment decisions will be important.

Implications for Practice, Training, and Policy

Present findings are necessarily preliminary, as they are based on a single sample, a limited set of vignettes, and a small range of cognitive heuristics, as well as not having data about long term changes in behavior. However, these findings are similar to results of other debiasing interventions in medicine (e.g., Crosskerry, 2013), and the effect sizes were medium or large under what could be considered the assessment analogy of an “efficacy” study with highly controlled conditions. The approach illustrates the potential value of practicing application of the strategies to cases, even vignettes, rather than reading about heuristics and debiasing as abstract principles (Davidow & Levenson, 1993; Galanter & Patel, 2005). This is consistent with current thinking in active learning (Tomcho & Foels, 2012). It suggests that teaching these methods in graduate assessment classes, and then creating opportunities to continue honing them in group supervision, case conferences, workshops, online continuing education exercises, or other venues that foster repeated application and discussion of principles and methods may be worth exploring as ways of extending dissemination, implementation, and retention of gains in practice (Collins, Leffingwell, & Belar, 2007).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Dr. E. Youngstrom has received travel support from Bristol-Myers Squibb and consulted with Lundbeck, Otsuka, Western Psychological Services, and Pearson about assessment.

Role of Funding Source: This material is based upon work supported by the North Carolina Translational and Clinical Sciences Institute: NIH grant#: 1UL1TR001111. The funding source did not play a role in the study design, in the collection, analysis, and interpretation of data, in the writing of the manuscript, or in the decision to submit the paper for publication.

References

- Aegisdottir S, White MJ, Spengler PM, Maugherman AS, Anderson LA, Cook RS, et al. The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *Counseling Psychologist*. 2006; 34:341. doi: 10.1177/0011000005285875.
- Allison, PD. Missing data. Sage; Thousand Oaks, CA: 2002.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders (4th--Text Revision ed.). Author; Washington, DC: 2001.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed.. Author; Washington, DC: 2013.
- Arkes HR. Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology*. 1981; 49:323–330. doi: 10.1037/0022-006X.49.3.323. [PubMed: 7276321]
- Arkes HR. Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*. 1991; 110:486–498. doi: 10.1037/0033-2909.110.3.486.
- Baker-Ericzen MJ, Jenkins MM, Park S, Garland AF. Clinical decision-making in community children's mental health: Using innovative methods to compare clinicians with and without training in evidence-based treatment. *Child & Youth Care Forum*. 2014; 14:1–25. doi: 10.1007/s10566-014-9274-x.
- Bieling PJ, MacQueen GM, Marriot MJ, Robb JC, Begin H, Joffe RT, Young LT. Longitudinal outcome in patients with bipolar disorder assessed by life-charting is influenced by DSM-IV personality disorder symptoms. *Bipolar Disorders*. 2003; 5:14–21. doi:10.1034/j.1399-5618.2003.00014.x. [PubMed: 12656933]
- Bornstein BH, Emler AC. Rationality in medical decision making: a review of the literature on doctors' decision-making biases. *Journal of Evaluation in Clinical Practice*. 2001; 7:97–107. doi: 10.1046/j.1365-2753.2001.00284.x. [PubMed: 11489035]
- Carpenter-Song E. Caught in the psychiatric net: Meanings and experiences of ADHD, pediatric bipolar disorder and mental health treatment among a diverse group of families in the United States. *Culture, Medicine and Psychiatry*. 2009; 33:61–85. doi:10.1007/s11013-008-9120-4.
- Carpenter-Song E, Whitley R, Lawson W, Quimby E, Drake RE. Reducing disparities in mental health care: Suggestions from The Dartmouth-Howard Collaboration. *Community Mental Health Journal*. 2011; 47:1–13. doi: 10.1007/s10597-009-9233-4. [PubMed: 19662528]
- Collins FL Jr, Leffingwell TR, Belar CD. Teaching evidence-based practice: Implications for psychology. *Journal of Clinical Psychology*. 2007; 63:657–670. [PubMed: 17551942]
- Cook B, McGuire T, Miranda J. Measuring trends in mental health care disparities, 2000–2004. *Psychiatric Services*. 2007; 58:1533–1540. doi: 10.1176/ps.2007.58.12.1533. [PubMed: 18048553]
- Croskerry P. Achieving quality in clinical decision making: cognitive strategies and detection of bias. *Academic Emergency Medicine*. 2002; 9:1184–1204. doi: 10.1197/aemj.9.11.1184. [PubMed: 12414468]
- Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*. 2003; 78:775. doi: 10.1097/00001888-200308000-00003. [PubMed: 12915363]
- Croskerry P. The theory and practice of clinical decision-making. *Canadian Journal of Anesthesia*. 2005; 52:1–8. doi: 10.1007/BF03023077. [PubMed: 15625248]

- Croskerry P. From mindless to mindful practice—Cognitive bias and clinical decision making. *New England Journal of Medicine*. 2013; 368:2445–2448. doi: 10.1056/NEJMp1303712. [PubMed: 23802513]
- Currin L, Schmidt U, Waller G. Variables that influence diagnosis and treatment of the eating disorders within primary care settings: A vignette study. *International Journal of Eating Disorders*. 2007; 40:257–262. doi: 10.1002/eat.20355. [PubMed: 17167756]
- Davidow J, Levinson EM. Heuristic principles and cognitive bias in decision making: Implications for assessment in school psychology. *Psychology in the Schools*. 1993; 30:351–361.
- Dawes R, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science*. 1989; 243:1668–1674. doi: 10.1002/eat.20355. [PubMed: 2648573]
- DeBello MP, Lopez-Larson MP, Soutullo CA, Strakowski SM. Effects of race on psychiatric diagnosis of hospitalized adolescents: A retrospective chart review. *Journal of Child and Adolescent Psychopharmacology*. 2001; 11:95–103. doi:10.1089/104454601750143528. [PubMed: 11322750]
- Dubicka B, Carlson GA, Vail A, Harrington R. Prepubertal mania: diagnostic differences between US and UK clinicians. *European Child and Adolescent Psychiatry*. 2008; 17:153–161. doi: 10.1007/s00787-007-0649-5. [PubMed: 17876503]
- Eddy DM. Probabilistic reasoning in clinical medicine: Problems and opportunities. *Judgment Under Uncertainty: Heuristics and Biases*. 1982; 17:249–267. doi: 10.1017/CBO9780511809477.019.
- Elstein AS, Schwartz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ*. 2002; 324:729–732. doi: 10.1136/bmj.324.7339.729. [PubMed: 11909793]
- Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. *Academic Medicine*. 2011; 86:307–313. doi: 10.1097/ACM.0b013e31820824cd. [PubMed: 21248608]
- Fabrega H, Ulrich R, Mezzich JE. Do Caucasian and Black adolescents differ at psychiatric intake? *Journal of the American Academy of Child and Adolescent Psychiatry*. 1993; 32:407–413. doi: 10.1136/bmj.324.7339.729. [PubMed: 8444771]
- Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods*. 2009; 41:1149–1160. doi: 10.3758/BRM.41.4.1149. [PubMed: 19897823]
- Galanter CA, Leibenluft E. Frontiers between attention deficit hyperactivity disorder and bipolar disorder. *Child and Adolescent Psychiatric Clinics of North America*. 2008; 17:325–346. doi: 10.1016/j.chc.2007.11.001. [PubMed: 18295149]
- Galanter CA, Patel VL. Medical decision making: a selective review for child psychiatrists and psychologists. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*. 2005; 46:675–689. doi: 10.1111/j.1469-7610.2005.01452.x.
- Garland AF, Lau AS, Yeh M, McCabe KM, Hough RL, Landsverk JA. Racial and ethnic differences in utilization of mental health services among high-risk youths. *American Journal of Psychiatry*. 2005; 162:1336–1343. doi: 10.1176/appi.ajp.162.7.1336. [PubMed: 15994717]
- Gigerenzer G, Goldstein DG. Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*. 1996; 103:650–669. doi: 10.1037/0033-295X.103.4.650. [PubMed: 8888650]
- Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*. 2000; 12:19–30. doi: 10.4088/JCP.v66n1001. [PubMed: 10752360]
- Hirschfeld RM, Lewis L, Vornik LA. Perceptions and impact of bipolar disorder: how far have we really come? Results of the national depressive and manic-depressive association 2000 survey of individuals with bipolar disorder. *Journal of Clinical Psychiatry*. 2003; 64:161–174. doi: 10.4088/JCP.v64n0209. [PubMed: 12633125]
- Jenkins MM, Youngstrom EA, Washburn JJ, Youngstrom JK. Evidence-based strategies improve assessment of pediatric bipolar disorder by community practitioners. *Professional Psychology: Research and Practice*. 2011; 42:121–129. doi: 10.1037/a0022506.

- Jenkins MM, Youngstrom EA, Youngstrom JK, Feeny NC, Findling RL. Generalizability of evidence-based assessment recommendations for pediatric bipolar disorder. *Psychological Assessment*. 2012; 24:269–281. doi: 10.1037/a0025775. [PubMed: 22004538]
- Jensen-Doss A, Weisz JR. Diagnostic agreement predicts treatment process and outcomes in youth mental health clinics. *Journal of Consulting and Clinical Psychology*. 2008; 76:711–722. doi: 10.1037/0022-006X.76.5.711. [PubMed: 18837589]
- Jensen-Doss A, Youngstrom E, Youngstrom J, Feeny N, Findling R. Predictors and moderators of agreement between clinical and research diagnoses for children and adolescents. *Journal of Consulting and Clinical Psychology*. 2014; 82:1151–1162. doi: 10.1037/a0036657. [PubMed: 24773574]
- Joseph MF, Youngstrom EA, Soares JC. Antidepressant-coincident mania in children and adolescents treated with selective serotonin reuptake inhibitors. *Future Neurology*. 2009; 4:87–102. doi: 10.2217/14796708.4.1.87. [PubMed: 19884978]
- Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on Psychological Science*. 2011; 6:21–37. doi: 10.1177/1745691610393527. [PubMed: 26162113]
- Kessler RC, Berglund P, Demler O, Jin R, Walters EE. Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*. 2005; 62:593–602. doi: 10.1001/archpsyc.62.6.593. [PubMed: 15939837]
- Kilgus MD, Pumariega AJ, Cuffe SP. Influence of race on diagnosis in adolescent psychiatric inpatients. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1995; 34:67–72. doi: 10.1097/00004583-199501000-00016. [PubMed: 7860460]
- Kowatch RA, Fristad M, Birmaher B, Wagner KD, Findling RL, Hellander M. Treatment guidelines for children and adolescents with bipolar disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2005; 44:213–235. doi: 10.1097/00004583-200503000-00006. [PubMed: 15725966]
- Kurtz RM, Garfield SL. Illusory correlation: A further exploration of Chapman's paradigm. *Journal of Consulting and Clinical Psychology*. 1978; 46:1009–1015. doi: 10.1037/0022-006X.46.5.1009. [PubMed: 701539]
- Leibenluft E. Severe mood dysregulation, irritability, and the diagnostic boundaries of bipolar disorder in youths. *American Journal of Psychiatry*. 2011; 168:129–142. doi: 10.1176/appi.ajp.2010.10050766. [PubMed: 21123313]
- Lish JD, Dime-Meenan S, Whybrow PC, Price RA, Hirschfeld RM. The National Depressive and Manic-depressive Association (DMDA) survey of bipolar members. *Journal of Affective Disorders*. 1994; 31:281–294. doi: 10.1016/0165-0327(94)90104. [PubMed: 7989643]
- Lueger RJ. Practice-informed research and research-informed psychotherapy. *Journal of Clinical Psychology*. 2002; 58:1265–1276. doi: 10.1002/jclp.10110. [PubMed: 12357441]
- Marchand WR, Wirth L, Simon C. Delayed diagnosis of pediatric bipolar disorder in a community mental health setting. *Journal of Psychiatric Practice*. 2006; 12:128–133. doi: 10.1097/00131746-200603000-00011. [PubMed: 16728912]
- McClellan J, Kowatch R, Findling RL, Work Group on Quality. Practice parameter for the assessment and treatment of children and adolescents with bipolar disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2007; 46:107–125. doi: 10.1097/01.chi.0000242240.69678.c4. [PubMed: 17195735]
- Meehl, PE. *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. University of Minnesota Press; Minneapolis: 1954.
- Meehl PE. The cognitive activity of the clinician. *American Psychologist*. 1960; 15:19–27. doi: 10.1037/h0041744.
- Nierenberg AA, Ostacher MJ, Calabrese JR, Ketter TA, Marangell LB, Miklowitz DJ, Sachs GS. Treatment-resistant bipolar depression: A STEP-BD equipose randomized effectiveness trial of antidepressant augmentation with lamotrigine, inositol, or risperidone. *American Journal of Psychiatry*. 2006; 163:210–216. doi: 10.1176/appi.ajp.163.2.210. [PubMed: 16449473]

- Pacchiarotti I, Bond DJ, Baldessarini RJ, Nolen WA, Grunze H, Licht RW, Vieta E. The International Society for Bipolar Disorders (ISBD) Task Force Report on Antidepressant Use in Bipolar Disorders. *American Journal of Psychiatry*. 2013 doi: 10.1176/appi.ajp.2013.13020185.
- Project MATCH Research Group. Matching Alcoholism Treatments to Client Heterogeneity: Project MATCH Posttreatment Drinking Outcomes. *Journal of Studies on Alcohol*. 1997; 58:7–29. doi: 10.15288/jsa.1997.58.7. [PubMed: 8979210]
- Qualtrics. 2014. Provo, Utah, USA Patent No
- Regehr G, Norman GR. Issues in cognitive psychology: implications for professional education. *Academic Medicine*. 1996; 71:988. doi: 10.1097/00001888-199609000-00015. [PubMed: 9125988]
- Rettew DC, Lynch AD, Achenbach TM, Dumenci L, Ivanova MY. Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research*. 2009; 18:169–184. doi: 10.1002/mpr.289. [PubMed: 19701924]
- Sedlmeier P, Gigerenzer G. Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology General*. 2001; 130:380–400. doi: 10.1037/0096-3445.130.3.380. [PubMed: 11561916]
- Soutullo CA, Chang KD, Diez-Suarez A, Figueroa-Quintana A, Escamilla-Canales I, Rapado-Castro M, Ortuno F. Bipolar disorder in children and adolescents: international perspective on epidemiology and phenomenology. *Bipolar Disorders*. 2005; 7:497–506. doi: 10.1111/j.1399-5618.2005.00262.x. [PubMed: 16403175]
- Spengler PM, White MJ, Ægisdóttir S, Maugherman AS, Anderson LA, Cook RS, Rush JD. The meta-analysis of clinical judgment project: Effects of experience on judgment accuracy. *The Counseling Psychologist*. 2009; 37:350–399. doi: 10.1177/0011000006295149.
- Strohmer DC, Leierer SJ. Modeling rehabilitation counselor clinical judgment. *Rehabilitation Counseling Bulletin*. 2000; 44:3. doi: 10.1177/003435520004400102.
- Tomcho TJ, Foels R. Meta-analysis of group learning activities: Empirically based teaching recommendations. *Teaching of Psychology*. 2012; 39:159–169. doi: 10.1177/0098628312450414.
- Van Meter AR, Moreira A, Youngstrom EA. Meta-analysis of epidemiologic studies of pediatric bipolar disorder. *Journal of Clinical Psychiatry*. 2011; 72:1250–1256. doi: 10.1348/000711200160345. [PubMed: 21672501]
- Weller E, Danielyan A, Weller R. Somatic treatment of bipolar disorder in children and adolescents. *Psychiatric Clinics of North America*. 2004; 27:155–178. doi: 10.1016/S0193-953X(03)00116-3. [PubMed: 15062636]
- Youngstrom EA, Birmaher B, Findling RL. Pediatric bipolar disorder: validity, phenomenology, and recommendations for diagnosis. *Bipolar Disorders*. 2008; 10:194–214. doi: 10.1111/j.1399-5618.2007.00563.x. [PubMed: 18199237]
- Youngstrom EA, Choukas-Bradley S, Calhoun CD, Jensen-Doss A. Clinical guide to the Evidence-Based Assessment approach to diagnosis and treatment. *Cognitive and Behavioral Practice*. 2014; 22:20–35. doi: 10.1016/j.cbpra.2013.12.005.
- Youngstrom EA, Duax J. Evidence based assessment of pediatric bipolar disorder, part 1: Base rate and family history. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2005; 44:712–717. doi: 10.1097/01.chi.0000162581.87710.bd. [PubMed: 15968241]
- Youngstrom EA, Findling RL, Calabrese JR, Gracious BL, Demeter C, Bedoya DDP, Price M. Comparing the diagnostic accuracy of six potential screening instruments for bipolar disorder in youths aged 5 to 17 years. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2004; 43:847–858. doi: 10.1097/01.chi.0000125091.35109.1e. [PubMed: 15213586]
- Youngstrom EA, Freeman AJ, Jenkins MM. The assessment of children and adolescents with bipolar disorder. *Child and Adolescent Psychiatric Clinics of North America*. 2009; 18:353–390. doi: 10.1016/j.chc.2008.12.002. [PubMed: 19264268]

Public Health Significance Statement

Study findings increase understanding of clinicians' cognitive vulnerabilities; they model a new approach for improving clinical decision-making. Greater awareness of faulty heuristics and using cognitive debiasing strategies improve clinicians' diagnostic reasoning and result in more accurate treatment decisions.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

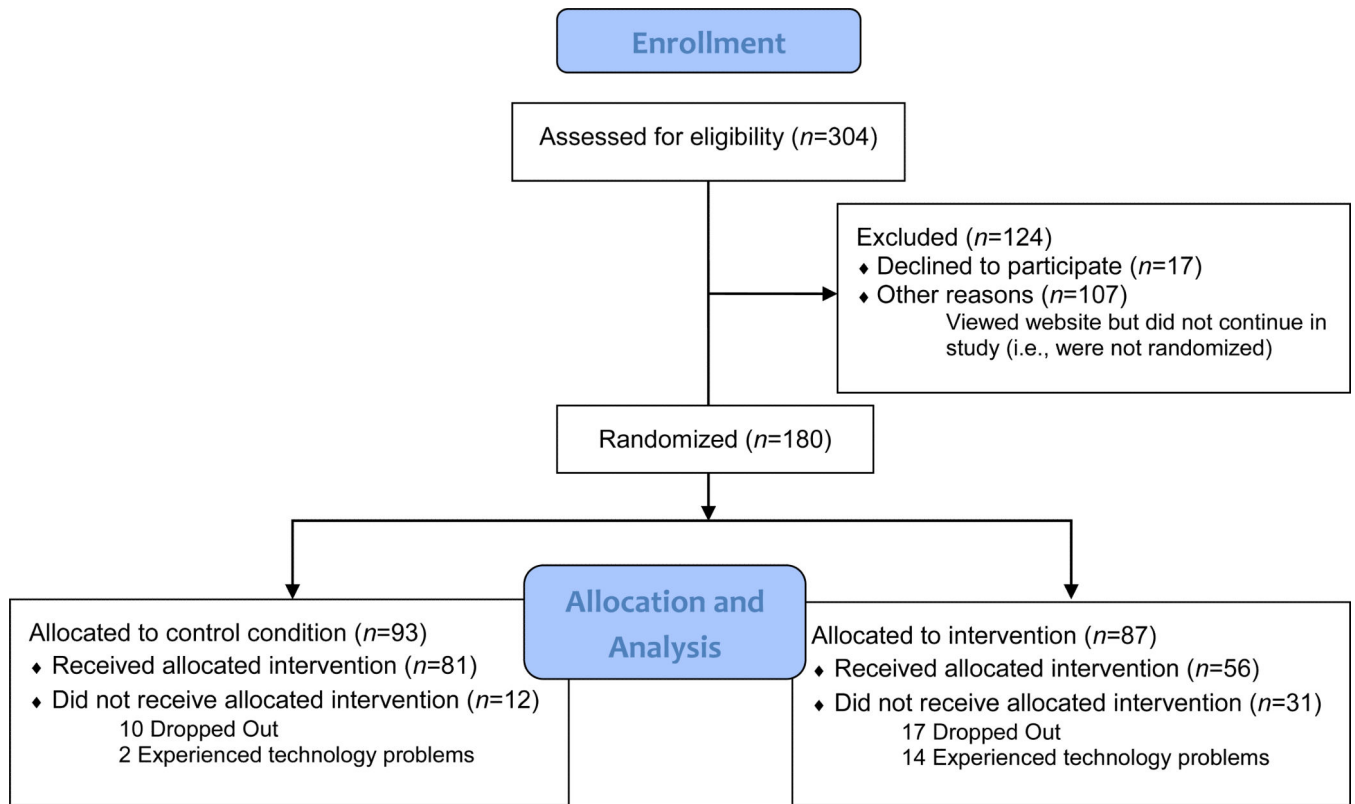


Figure 1.
CONSORT flow diagram.

Table 1

Participant Demographics

Characteristic	Control Group (<i>n</i> = 81) <i>n</i> (%)	Treatment Group (<i>n</i> = 56) <i>n</i> (%)	Total (<i>N</i> = 137) <i>n</i> (%)
Female Gender	69 (85)	45 (80)	81 (83)
Age			
< 30	34 (42)	28 (50)	62 (45)
31-45	37 (46)	17 (30)	54 (39)
45-65	10 (12)	10 (18)	20 (15)
>65	0 (0)	1 (2)	1 (1)
Race/ethnicity			
Asian	8 (10)	7 (13)	15 (11)
African American	2 (2.5)	2 (4)	4 (3)
Caucasian	61 (75)	42 (75)	103 (75)
Latino	8 (10)	4 (7)	12 (9)
Other	2 (2.5)	1(2)	3 (2)
Current Geographical Location in US			
Southern	30 (37)	22 (39)	39 (28)
Northeast	9 (11)	6 (11)	15 (11)
Western	20 (25)	11 (20)	31 (23)
Midwest	22 (27)	17 (30)	39 (28)
Years of experience	8.07 (SD=7.04)	9.32 (SD=8.26)	8.6 (SD=7.49)
Professional Title:			
Psychiatry	20 (25)	10 (18)	30 (22)
Psychology	38 (47)	27 (48)	65 (47)
Clinical	28 (35)	18 (32)	38 (34)
School	7 (9)	6 (11)	13 (10)
Counseling	3 (4)	3 (5)	6 (4)
Social Work	10 (12)	10 (18)	20 (15)
MFT	5 (6)	4 (7)	9 (7)
Counselor	8 (10)	5 (9)	13 (10)
Unlicensed, supervised professionals	27 (33)	18 (32)	45 (33)

Note. Tx = treatment. MFT = Marriage and Family Therapist. χ^2 tests indicate that the treatment and control groups did not significantly differ on any demographic factors, all $p > .05$.

Table 2

Participant Characteristics

Characteristic	Control Group (<i>n</i> = 81) <i>n</i> (%)	Tx Group (<i>n</i> = 56) <i>n</i> (%)	Total (<i>N</i> = 137) <i>n</i> (%)
Psychology Professional	38 (47)	27 (48)	65 (47)
Primary theoretical orientation			
Cognitive Behavioral	36 (44)	24 (43)	60 (44)
Family Systems	6 (7)	5 (9)	11 (8)
Humanistic	3 (4)	0 (0)	3 (2)
Integrative	6 (7)	5 (9)	11 (8)
Psychoanalytic	1 (1)	2 (4)	3 (2)
Combined	18 (22)	10 (18)	28 (20)
Other	2 (3)	2 (4)	4 (3)
Primary clinical setting			
Community Mental Health	17 (21)	12 (21)	29 (21)
University Setting	9 (11)	9 (16)	18 (13)
Hospital (inpatient and outpatient)	21 (26)	13 (23)	34 (25)
School	17 (21)	8 (14)	25 (18)
Private Practice	2 (3)	3 (5)	5 (4)
Residential	2 (3)	1 (2)	3 (2)
Other	2 (3)	1 (2)	3 (2)
Combined	11 (14)	8 (14)	19 (14)
Primary professional activity			
Clinical service	69 (85)	37 (66)	106 (77)
Research	3 (4)	5 (9)	8 (6)
Administration	1 (1)	3 (6)	4 (3)
Teaching	0 (0)	3 (5)	3 (2)
Other	0 (0)	2 (4)	2 (2)
Combined	8 (10)	6 (11)	14 (10)
Trained in evidence-based practice	77 (95)	52 (93)	129 (94)