HHS Public Access

Author manuscript

Depress Anxiety. Author manuscript; available in PMC 2020 August 01.

Published in final edited form as:

Depress Anxiety. 2019 August ; 36(8): 744–752. doi:10.1002/da.22900.

Integrating Evidence-based Assessment into Clinical Practice for Pediatric Anxiety Disorders

Joseph F. McGuire, Ph.D.[1,2], **Nicole E. Caporino, Ph.D.**[3], **Sophie A. Palitz, M.A.**[4], **Philip C. Kendall, Ph.D. ABPP**[4], **Anne Marie Albano, Ph.D.**[5], **Golda Ginsburg, Ph.D.**[6], **Boris Birmaher, M.D.**[7], **John T. Walkup, M.D.**[8], **John Piacentini, Ph.D.**[2]

[1]Department of Psychiatry and Behavioral Science, Johns Hopkins University School of Medicine

[2]Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles

[3]Department of Psychology, American University

[4]Department of Psychology, Temple University

[5]New York State Psychiatric Institute and Columbia University Medical Center

[6]University of Connecticut Health Center

[7]University of Pittsburgh School of Medicine

[8]Lurie Children's Hospital and Northwestern University Feinberg School of Medicine

Abstract

Background: Although evidence-based assessments are the cornerstone of evidence-based treatments, it remains unknown whether incorporating evidence-based assessments into clinical practice enhances therapists' judgment of therapeutic improvement. This study examined whether

Corresponding Author: Joseph F. McGuire, Ph.D., Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, Maryland. 1800 Orleans St, Baltimore, MD 21287. jfmcguire@jhmi.edu.

Financial Disclosures

Ms. Palitz has nothing to disclose. Dr. McGuire has received support from the Tourette Association of America, American Academy of Neurology, the Brain Research Foundation, and NIMH. He receives royalties from Elsevier, and serves as a consultant for Bracket and Luminopia. Dr. Caporino has received an honorarium from NICHD and grant funding from the American Psychological Foundation and the Anna-Maria and Stephen Kellen Foundation. Dr. Kendall has received research support from NIMH. He has received honoraria from professional societies for speaking at conventions. He has received royalties from Guildford Press, Ericsson, Workbook Publishing (his spouse's employment), and Oxford University Press. Dr. Ginsburg has received support from NIMH and from the US Department of Education/Institute of Education Sciences. Dr. Albano has received royalties from Oxford University Press for the Anxiety Disorders Interview Schedule, Child and Parent Versions. She has received an Editor's Honorarium from the American Psychological Association. Dr. Birmaher has received grants from NIMH and receives royalties from APA Press, UptoDate, and Random House for previous publications. Dr. Walkup has received research support from the Tourette Syndrome Association of America and the Hartwell Foundation. He has received honoraria and travel expenses for speaking engagements and meetings sponsored by the Tourette Association of America. He has received royalties from Guilford Press and Oxford University Press for multi-author books published about Tourette's syndrome and from Wolters Kluwer for CME activity on childhood anxiety. He has served as an unpaid advisor to the Anxiety Disorders Association of America and the Trichotillomania Learning Center. He has served as a paid speaker for the Tourette Syndrome--Center for Disease Control and Prevention outreach educational programs, the American Academy of Child and Adolescent Psychiatry, and the American Psychiatric Association. Dr. Piacentini has received grant or research support from NIMH, Pfizer Pharmaceuticals through the Duke University Clinical Research Institute CAPTN Network, the Tourette Association of America, and the TLC Foundation for BFRBs. He is a co-author of the Child OCD Impact Scale-Revised (COIS-R), the Child Anxiety Impact Scale (CAIS), the Parent Tic Questionnaire (PTQ), and the Premonitory Urge for Tics Scale (PUTS) assessment tools, all of which are in the public domain therefore no royalties are received. He has received royalties from Guilford Press and Oxford University Press. He has served on the speakers' bureau of the Tourette Association of America, and the International Obsessive Compulsive Disorder Foundation.

the inclusion of youth-and parent-reported anxiety rating scales improved therapists' judgment of treatment response and remission compared to the judgment of treatment-masked IEs after (a) weekly/biweekly acute treatment and (b) monthly follow-up care.

Methods: 436 youth received cognitive-behavioral therapy (CBT), medication, CBT with medication, or pill placebo through the Child/Adolescent Anxiety Multimodal Study (CAMS). Participants and parents completed the following anxiety scales at pre-treatment, post-treatment, and follow-up: Screen for Childhood Anxiety and Related Disorders (SCARED) and Multidimensional Anxiety Scale for Children (MASC). IEs rated anxiety on the Clinical Global Impression of Severity (CGI-S) and Improvement (CGI-I) at post-treatment and follow-up. Therapists rated anxiety severity and improvement using scales that paralleled IE measures.

Results: Fair-to-moderate agreement was found between therapists and IEs after acute treatment (κ=.38-.48), with only slight-to-fair agreement found after follow-up care (κ=.07-.33). Optimal algorithms for determining treatment response and remission included the combination of therapists' ratings and the parent-reported SCARED after acute (κ=.52-.54) and follow-up care (κ=.43-.48), with significant improvement in the precision of judgments after follow-up care (p<. 02-.001).

Conclusion: Therapists are good at detecting treatment response and remission, but the inclusion of the parent-report SCARED optimized agreement with IE ratings—especially when contact was less frequent. Findings suggest that utilizing parent-report measures of anxiety in clinical practice improves the precision of therapists' judgment.

Keywords

assessment; anxiety disorders; children; adolescents; evidence-based assessment

Anxiety disorders are characterized by disproportionate distress to a triggering event that elicits uncontrollable worries, avoidance of anxiety-provoking situations, and a myriad of safety behaviors (American Psychiatric Association, 2013; King, Gullone, & Ollendick, 1990). These conditions predominantly onset in childhood or adolescence (Keller et al., 1992; Regier, Rae, Narrow, Kaelber, & Schatzberg, 1998), and affect up to 20% of youth (Merikangas et al., 2010; Shaffer et al., 1996). Pediatric anxiety is associated with significant impairment (Langley et al., 2014), limited functioning (Swan & Kendall, 2016), and a poor quality of life (Stevanovic, 2013). As pediatric anxiety represents a strong risk factor for severe psychopathology in adulthood (Pine, Cohen, Gurley, Brook, & Ma, 1998), the effective treatment of anxiety in youth can minimize morbidity across the lifespan.

The two evidence-based treatments for pediatric anxiety are pharmacotherapy and cognitive-behavior therapy (CBT; Walkup et al., 2008). Evidence-based pharmacotherapy for anxiety primarily consists of selective serotonin reuptake inhibitors (SSRIs; Ipser, Stein, Hawkridge, & Hoppe, 2009). Meanwhile, evidence-based CBT for pediatric anxiety consists of multiple components, such as psychoeducation about anxiety, cognitive restructuring, homework assignments, and exposure-based behavioral experiments (Kendall, Chu, Gifford, Hayes, & Nauta, 1999; Podell, Mychailyszyn, Edmunds, Puleo, & Kendall, 2010). Collectively, CBT and SSRIs are associated with a 55%−80% treatment response and 20%−68% diagnostic remission rate in pediatric clinical trials, as assessed by independent evaluators (IEs) masked

to treatment condition and using clinician-administered diagnostic interviews and rating scales (Ginsburg et al., 2011; Walkup et al., 2008). As aspirations in clinical practice are to achieve diagnostic remission for anxious youth, it is important to implement assessment tools that accurately portray both therapeutic improvement and remission in an efficient manner.

However, integrating clinical trials outcome measures to monitor improvement in clinical practice is challenging for multiple reasons. Attempts have been made to overcome these challenges using clinician-administered scales like the Pediatric Anxiety Rating Scale (PARS; Research Units on Pediatric Psychopharmacology Anxiety Study Group, 2002), which have benchmarks of treatment response and clinical remission (Caporino et al., 2013; Johnco, Salloum, Lewin, & Storch, 2015). However, the PARS has not been widely adopted in clinical practice due in part to its time-intensive nature and training needed for administration. In comparison to interviews (Silverman & Albano, 1996) and clinician-rated scales (Research Units on Pediatric Psychopharmacology Anxiety Study Group, 2002), youth-and parent-rated scales offer several advantages. These measures are brief, cost-effective, require minimal training to administer and interpret, and remove potential clinician bias. Therefore, youth-and parent-rated anxiety scales may be an ideal tool to incorporate into evidence-based measurement in clinical practice.

Several youth-and parent-rated anxiety scales are available that have good reliability and validity (see Silverman & Ollendick, 2005 for a review). Two such youth-and parent-report anxiety scales that have demonstrated strong psychometric properties also have benchmarks for determining treatment response and remission that exhibit good agreement with IE ratings: the Screen for Childhood Anxiety and Related Disorders (SCARED)-Child-and Parent-Report and the Multidimensional Anxiety Scale for Children (MASC) Parent-Report (Caporino et al., 2017; Palitz et al., 2018). Caporino and colleagues (2017) found that a 55% reduction on the SCARED-Parent total score and 50% reduction on the SCARED-Child total score optimally predicted treatment response. There were slight differences between the optimal percent reduction that predicted treatment response between children and adolescents on the SCARED-Parent (55% reduction in total score for adolescents and 60% reduction in total score for children) and SCARED-Child (65% reduction for adolescents and 50% reduction in total score for children). Meanwhile, a SCARED-Parent total score cutoff of 10 and a SCARED-Child total score cutoff of 12 optimally predicted clinical remission. Similarly, there were slight differences between total score cutoffs that predicted clinical remission for children and adolescents on the SCARED-Parent (total score cutoff of 9 for adolescents and 12–13 for children) and SCARED-Child (total score cutoff of 7 for adolescents and 5 for children). For the MASC-Parent Report, Palitz et al. 2018 used the same sample as Caporino et al. 2017 and found that a 35% reduction in the total score optimally predicted treatment response, and a total score cutoff of 42 optimally predicted clinical remission (Palitz et al., 2018).

A key aspect of evidence-based treatment is the systematic assessment of symptom severity and therapeutic improvement. The SCARED and MASC reports offer benchmarks for identifying treatment response and remission, but it is unknown whether using these scales in clinical practice improves therapists' judgment of treatment response and remission.

Accordingly, we examined the agreement between therapists' ratings and treatment-masked IE ratings of treatment response and remission in a multisite clinical trial for pediatric anxiety disorders (Walkup et al., 2008). Next, we evaluated algorithms that combined therapist ratings, parent responses, and youth responses using the SCARED and MASC benchmarks after acute weekly/biweekly treatment (i.e., post-treatment) and monthly follow-up care (i.e., a 6-month follow-up). We were interested in identifying the combination of ratings that maximized agreement with objective IE ratings. Finally, we explored factors associated with agreement between optimal algorithms and IE ratings.

METHODS

Participants

Participants were youth enrolled in the Child/Adolescent Multimodal Study (CAMS; Walkup et al., 2008) who completed all measures of interest before and after treatment (N = 436). Briefly, inclusion criteria for the trial included: (1) 7–17 years of age; (2) a primary diagnoses of Separation Anxiety Disorder, Generalized Anxiety Disorder, and/or Social Phobia with substantial impairment; (3) an estimated IQ > 80; and (4) be medication free or on a stable dose of a stimulant medication (Walkup et al., 2008). Youth were excluded from participation if any of the following were present: (1) a co-occurring psychiatric condition that warranted immediate treatment not provided within the study (e.g., bipolar disorder, eating disorder, etc.); (2) an unstable medical condition; (3) school refusal due to anxiety; (4) an acute risk to themselves or others; (5) non-responsive to two adequate trials of SSRIs or CBT; and (6) psychoactive medications other than stable dose of stimulants (see Walkup et al., 2008). Participants consisted of 325 children (7–12 years old) and 111 adolescents (13–17 years of age), with an even gender distribution. Participants were mostly White (n = 352, 80.7%) and non-Hispanic (n = 388, 89%). Most youth had two or more anxiety disorders (n = 343, 78.7%), with only 21.3% presenting with a single anxiety disorder (Separation Anxiety Disorder: n = 15, 3.4%; Generalized Anxiety Disorder: n = 30, 6.9%; and Social Phobia: n 48, 11%; Kendall et al., 2010).

Measures

Screen for Child Anxiety Related Emotional Disorders – Child and Parent Versions (SCARED-Child/Parent; Birmaher et al., 1997). The SCARED-Child/Parent are parallel youth-and parent-report measures of the frequency of common anxiety symptoms over 41 items. Items are rated on a 3-point Likert-type scale and summed to produce a total score (range: 0–82). The SCARED-Child/Parent has good psychometric properties in youth and is publically available (Birmaher et al., 1999; Birmaher et al., 1997).

Multidimensional Anxiety Scale for Children – Parent Version (MASC-Parent; Wood, Piacentini, Bergman, McCracken, & Barrios, 2002). The MASC-Parent is a 39-item parent-report measure of the frequency of anxiety symptoms. Items are rated on a 4-point Likert-type scale and summed to produce a total score (range: 0–117). The MASC-Parent has good psychometric properties in youth (Baldwin & Dadds, 2007; Wood et al., 2002), and is an accurate predictor of diagnostic status (Villabø, Gere, Torgersen, March, & Kendall, 2012).

Clinical Global Impression of Improvement (CGI-I) and Severity (CGI-S) Scales (Guy, 1976). The CGI-I and CGI-S are single-item clinician-rated measures of global improvement from baseline and global rating of anxiety severity, respectively. The CGI-I is commonly used as the primary outcome measure in clinical trials of pediatric anxiety (Birmaher et al., 2003; Liebowitz et al., 2002; Storch et al., 2015; Walkup et al., 2008). *Treatment-masked IEs* rated perceived improvement on a 7-point scale that ranged from "Very much improved" (1) to "Very much worse" (7), with a rating of "Much improved" or "Very much improved" indicating a positive treatment response. The same IEs used the CGI-S to rate anxiety severity on a 7-point scale that ranged from "Not at all ill" (1) to "Extremely ill" (7). Consistent with benchmarking studies (Caporino et al., 2013; Caporino et al., 2017; Palitz et al., 2017), ratings of "Not at all ill" or "Borderline ill" designated clinical remission. To ensure reliable ratings, IEs were trained to a pre-specified reliability criterion and continually monitored through weekly on-site supervision and biweekly cross-site calibration conference calls during the trial.

Treating therapists (CBT therapist or child psychiatrist) completed parallel CGI-I and CGI-S ratings using the same conventions as described above. For participants who received both weekly CBT and biweekly pharmacotherapy, CBT therapists' ratings were used due to their greater level of therapeutic contact with study participants.

Procedures

Procedures were approved by the Institutional Review Board at each of the six data collection sites. After parental consent and youth assent were obtained, a pre-treatment assessment was completed that included the IE-rated CGI-S and youth-and parent-report scales (SCARED-Child/Parent and MASC-Parent). Youth who met study eligibility criteria were randomly assigned to receive one of four treatments: placebo, manual-based CBT (Coping Cat program), medication (sertraline), or CBT plus medication. Full details regarding the study protocol and procedures have been described elsewhere (Compton et al., 2010). Following the completion of the 12-week course of acute treatment (i.e., post-treatment) and monthly follow-up care (i.e., 6-month follow-up), youth-and parent-report scales (SCARED-Child/Parent and MASC-Parent), IE ratings (CGI-I and CGI-S) and therapist-rated CGI-S and CGI-I scales were administered.

Analytic plan

There were no missing data from participants during acute treatment ($N = 436$), but 52 cases (11%) were missing data at 6-month follow-up. Data were determined to be missing completely at random (Little's MCAR, $p > .05$) and addressed using multiple imputation with pooled estimates. For increased measurement precision, the optimal benchmarks described above were used for the SCARED-Child/Parent and MASC-Parent based on participants' age (Caporino et al., 2017; Palitz et al., 2017). Chi-square and kappa statistics evaluated agreement between IEs' ratings (CGI-I and CGI-S) and: (1) therapists' ratings (CGI-I Therapist, CGI-S Therapist); and (2) the combination of therapists' ratings and response/remission benchmarks on the anxiety measures (SCARED-Child, SCARED-Parent, MASC-Parent). Response and remission algorithms that combined measures utilized an "and" rule for categorization (i.e., treatment response was needed on all measures

included in the algorithm to be considered a treatment responder). Kappa was interpreted using the following conventions: values ≤ 0 as no agreement, .01–.20 as none to slight, .21–.40 as fair, .41–.60 as moderate, .61–.80 as substantial, and .81–1.00 as almost perfect agreement (Landis & Koch, 1977). Z-scores were calculated to compare kappa obtained between the therapist rating and the optimal algorithm, as well as algorithms that included only youth or parent report. Significance values set at $p < .05$ for these z-score comparisons (Fleiss, 1981). Finally, chi-squares and t-tests explored differences in demographic (e.g., age, gender, race, ethnicity) and clinical characteristics (e.g., anxiety severity on rating scales, the presence of a co-occurring internalizing disorder or externalizing disorder) between youth who were classified the same or differently between IE and optimal algorithm approaches.

RESULTS

Agreement between IE ratings and therapist ratings after acute treatment.

At post-treatment, there was moderate agreement between IE and therapist ratings of treatment response ($\kappa = .48$, see Table 1), but only fair agreement on ratings of remission ($\kappa = .38$, see Table 2).

Algorithms that optimize therapist agreement with IE ratings after acute treatment.

Table 1 displays the agreement between IE ratings of treatment response and algorithms that used therapist, parent, and youth responses. Overall, these algorithms produced fair-to-moderate agreement with IE ratings ($\kappa = .36$-$.54$). Algorithms that included youth-report exhibited significantly less agreement ($\kappa = .41$) relative to algorithms that included only parent-report ($\kappa = .52$-$.54$; $z = 2.24$, $p < .03$). Optimal agreement was observed when using a combination of the therapist-rated CGI-I and treatment response benchmark on the SCARED-Parent. Although exhibiting slightly greater agreement, this combined approach did not significantly improve agreement of treatment response classifications in comparison to therapist-rated CGI-I ratings alone ($z = 1.03$, $p = .30$).

Table 2 displays the agreement between IE ratings of remission and remission algorithms. Findings were largely similar to treatment response analyses, with fair-to-moderate agreement between IE ratings and combined-informant algorithms ($\kappa = .35$-$.45$). Although algorithms that included youth-report only exhibited less agreement ($\kappa = .37$) compared to algorithms that included parent-report only ($\kappa = .41$-$.45$), these differences were not statistically significant ($z = 0.62$–1.26, $p = .21$-$.54$). Optimal agreement was achieved when using a combination of the therapist-rated CGI-I and clinical remission benchmark on the SCARED-Parent. Although exhibiting slightly greater agreement, this combined approach did not significantly improve agreement of remission classifications in comparison to therapist-rated CGI-S ratings alone ($z = 1.11$, $p = .27$).

Factors associated with agreement when using optimal algorithm after acute treatment.

Despite good agreement between IE ratings and the optimal treatment response algorithm (76.1% agreement), disagreement was present ($n = 104$ cases, 23.9%). Nine youth were classified as treatment responders using the optimal response algorithm but not by IE ratings

(i.e., a false positive), and 95 youth were classified as responders by IE ratings but not the optimal response algorithm (i.e., a false negative). Cases that produced disagreement had a lower pre-treatment MASC-Parent total scores (M = 62.24, SD = 15.34) compared to cases that exhibited classification agreement (M = 65.76, SD = 14.64, t = −2.12, p < .04, d = .24). No other differences were found on pre-treatment demographic (p = .12–1.00) or clinical characteristics (p = .06-.43).

Similarly, there was good agreement between IE remission ratings and the optimal remission algorithm (73.4% agreement), but some disagreement was present (n = 116, 26.6%). Twenty-two youth were classified as in remission using the optimal remission algorithm but not the IE rating (i.e., a false positive), and 94 youth were classified as in remission by the IE rating but not the optimal remission algorithm (i.e., a false negative). Cases that produced classification disagreement had lower pre-treatment CGI-S scores (M = 4.88, SD = 0.65) compared to cases that exhibited classification agreement (M = 5.08, SD = 0.75, t = −2.57, p < .01, d = .28). There were no other differences on pre-treatment demographic (p = .21-.70) or clinical characteristics (p = .32-.89).

Agreement between IE ratings and therapist ratings after follow-up care.

At 6-month follow-up, when participants were receiving only monthly maintenance care, there was only slight agreement between IE and therapist ratings of treatment response (κ = .07, see Table 3), but fair agreement on ratings of remission (κ = .33, see Table 4).

Algorithms that optimize therapist agreement with IE ratings after follow-up care.

Table 3 displays the agreement between IE ratings of treatment response and algorithms using therapist, parent, and child responses. Overall, these algorithms produced fair-to-moderate agreement with IE ratings (κ = .27-.43). Similar to acute treatment, algorithms that included youth-only reports exhibited less agreement (κ = .27) than algorithms that included parent-only reports (κ = .36-.43). While differences in agreement between algorithms that used only the SCARED-Child and only the MASC-Parent were not significant (z = 1.29, p = .20), algorithms that only used the SCARED-Parent did exhibited significantly greater agreement with IE ratings compared to the algorithms that used only the SCARED-Child (z = 2.33, p < .02). Optimal agreement was observed when using a combination of the therapist-rated CGI-I and treatment response benchmark on the SCARED-Parent. This optimal algorithm significantly improved the agreement of treatment response classifications with IE ratings in comparison to therapist-rated CGI-I ratings alone (z = 4.15, p < .001).

Table 4 displays the agreement between IE ratings of clinical remission and remission algorithms. These algorithms produced fair-to-moderate agreement with IE ratings (κ = .42–.48). In comparison to prior assessments, algorithms that included youth-only reports exhibited similar agreement (κ = .44) to algorithms that included parent-only reports (κ = .44-.48) and were not significantly different from one another (z = 0.00–0.67, p = .50–1.00). Optimal agreement was achieved using a combination of the therapist-rated CGI-I and remission benchmark on the SCARED-Parent. This algorithm significantly improved the agreement of remission classifications with IE ratings in comparison to therapist-rated CGI-S ratings alone (z = 2.41, p < .02).

Factors associated with agreement when using optimal algorithms after follow-up care.

Despite overall agreement between IE ratings and the optimal treatment response algorithm (75.2% agreement), some disagreement was present (n = 108 cases, 24.8%). Thirty-one youth were classified as responders using the treatment response algorithm but not the IE ratings, and 77 youth were classified as responders by the IE but not the response algorithm. There were no differences between cases that exhibited agreement and disagreement on pre-treatment demographic (p = .37–.97) or clinical characteristics (p = .36–.99).

There was also good agreement between IE ratings and the optimal remission algorithm (73.6% agreement), with some disagreement present (n = 115 cases, 26.4%). Twenty-nine youth were classified as in remission using the optimal remission algorithm but not the IE rating, and 86 youth were classified as in remission by the IE rating but not the remission algorithm. No differences were found between cases that exhibited agreement and disagreement on pre-treatment demographic (p = .41–.97) or clinical characteristics (p = .28–.90).

Discussion

Given the importance of evidence-based measurement in clinical practice, this report identified usable algorithms for determining treatment response and clinical remission that maximized agreement with objective IE ratings after 12 weeks of acute treatment (i.e., post-treatment) and monthly follow-up care (i.e., 6-month follow-up). Multiple treatment response and remission algorithms that combined therapist, parent, and youth responses were considered. First, similar to reports in pediatric OCD (Lewin, Peris, De Nadai, McCracken, & Piacentini, 2012), therapists exhibited good agreement with IE ratings after acute treatment; although the level of agreement dropped during follow-up care when therapeutic contact was less frequent. While acute treatment involved weekly CBT sessions and/or biweekly pharmacotherapy visits, maintenance care involved only monthly visits (Piacentini et al., 2014). Thus, greater agreement during acute treatment is likely connected to greater therapeutic contact. Across response and remission ratings, therapists' ratings exhibited good sensitivity for detecting IE-rated response and remission (i.e., 58%−95%), which suggests that therapists are adept at perceiving improvement among their patients. However, therapists' ratings lacked specificity (i.e., correctly identifying non-responders, 11%−65%). The combined algorithm of therapist-rated CGI and SCARED-Parent improved agreement with objective IE ratings of response and remission. Although this algorithm had slightly lower sensitivity than therapist ratings alone (53%−76%), it had improved specificity (73%−94%), positive predictive value (i.e., rate at which response/remission identified by the algorithm was also identified by IEs, 83%−95%), and negative predictive value (i.e., rate at which non-response/remission identified by the algorithm was also identified by IEs, 52%−69%) in most cases. Thus, incorporating the SCARED-Parent improved therapists' precision in detecting treatment response and remission, especially after monthly follow-up care.

Second, when examining factors associated with classification agreement, cases that exhibited disagreement had lower pre-treatment anxiety severity on the MASC-Parent and CGI-S compared to cases that exhibited agreement after acute treatment. Lower pre-

treatment anxiety may be associated with disagreement for several reasons. For instance, treatment response cutoffs were based on percent reductions in total score (compared to cutoff scores for remission). It can be more challenging to achieve greater percent reductions in the presence of a lower pre-treatment score. As such, these findings highlight that therapist should be cautious when determining treatment response and remission when pre-treatment anxiety severity is low. However, no other participant characteristics were associated with disagreement after acute or follow-up care, nor were lower pre-treatment scores on the SCARED-Child/Parent associated with disagreement. Thus, while therapists should be cautious, the utilization of evidence-based measures appears to optimize the precision of therapists' assessment of treatment progress across a range of clinical presentations.

Despite study strengths (e.g., multi-site design, rigorous methodology, multiple treatment conditions, reliable IEs), a few limitations exist. First, although drawn from geographically-dispersed university clinics, there may be questions concerning the generalizability of findings to a broader population of clinic-referred youth. However, we note that minimal differences on internalizing problem measures have been reported between youth who present at community-based clinics and youth referred to university-based clinics (Southam-Gerow, Weisz, & Kendall, 2003). Second, participants in this study were mostly non-Hispanic White. Although demographic factors did not influence classification agreement, future studies should replicate these findings in other racial and ethnic groups. Third, study therapists had experience treating anxious youth and received ongoing supervision, which may have increased precision in detecting clinically meaningful change during acute treatment. As this level of clinical experience and oversight would not likely be found in regular clinical practice, community practitioner ratings in routine care might exhibit less agreement with treatment-masked IE ratings. Thus, the inclusion of the SCARED-Parent (and/or other parent-report benchmarks of response and remission) may further improve the precision of therapists' assessment of treatment response and remission. Future research should investigate moderators of therapist agreement with treatment-masked IE ratings to determine whether there are specific patient and/or therapist characteristics associated with greater agreement with masked IE ratings. Finally, the present sample was used to identify benchmarks of response and remission for the SCARED and MASC. Future research is needed to replicate these algorithms in independent samples.

Conclusion

In summary, this report is the first to offer evidence that using empirically-derived benchmarks improves the precision of therapists' judgment of treatment response and remission. Specifically, incorporating parent reported anxiety (especially the SCARED-Parent) with therapists' global ratings of improvement optimized agreement with an external gold-standard metric of treatment response and remission. Although these findings may be counterintuitive due to the internalized nature of these symptoms, the present results argue strongly for the inclusion of parent-reports in regular clinical practice for both CBT therapists and child psychiatrists. While this study occurred primarily in university-based anxiety disorder clinics, these findings bear considerable importance wherever pediatric anxiety is being treated using evidence-based practice. Indeed, the inclusion of parent-

reported anxiety scales in therapists' judgments of anxiety improvement may lead to more accurate treatment-related decision making, resulting in better clinical outcomes, and may bring about more efficient use of mental health resources. The inclusion of evidence-based assessments is important in all cases of pediatric anxiety. Moreover, it may be most pertinent for newer therapists who have less experience treating pediatric anxiety. For instance, less experienced therapists may over and/or underestimate youth's therapeutic improvement due to limited hands-on experience and/or supervision, which could result in early treatment discontinuation or a prolonged treatment course. Similarly, these findings are highly relevant for clinicians providing medication management (i.e., child psychiatrists, nurse practitioners, pediatricians, etc.) who have limited therapeutic contact with pediatric patients due to the briefer and less frequent nature of monthly medication management visits. Given their brevity, parent-report measures should be implemented within standard clinical care to aid practitioners in detecting clinical levels of anxiety and efficiently determining when clinically meaningful response and remission has been achieved. Although this report focused on the MASC and SCARED child-and parent-report, there are other rating scales such as the Spence Children's Anxiety Scale (Spence, Barrett, & Turner, 2003) or the MASC-Second Edition (March, 2013) that have good psychometric properties and should be examined in a parallel fashion as done here.

Acknowledegements:

This research was supported in part by grants (U01 MH064089 to Dr. Walkup; U01 MH64092 to Dr. Albano; U01 MH64003 to Dr. Birmaher; U01 MH63747 to Dr. Kendall; U01 MH64107 to Dr. March; U01 MH64088 to Dr. Piacentini; and U01 MH064003 to Dr. Compton) from the National Institute of Mental Health (NIMH). Sertraline and matching placebo were supplied free of charge by Pfizer. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIMH or NIH.

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

Baldwin JS, & Dadds MR (2007). Reliability and validity of parent and child versions of the multidimensional anxiety scale for children in community samples. Journal of the American Academy of Child & Adolescent Psychiatry, 46, 252–260. [PubMed: 17242629]

Birmaher B, Axelson DA, Monk K, Kalas C, Clark DB, Ehmann M, … Brent DA. (2003). Fluoxetine for the treatment of childhood anxiety disorders. Journal of the American Academy of Child & Adolescent Psychiatry, 42, 415–423. [PubMed: 12649628]

Birmaher B, Brent DA, Chiappetta L, Bridge J, Monga S, & Baugher M (1999). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): A replication study. Journal of the American Academy of Child & Adolescent Psychiatry, 38, 1230–1236. [PubMed: 10517055]

Birmaher B, Khetarpal S, Brent D, Cully M, Balach L, Kaufman J, & Neer SM (1997). The Screen for Child Anxiety Related Emotional Disorders (SCARED): Scale construction and psychometric characteristics. Journal of the American Academy of Child & Adolescent Psychiatry, 36, 545–553. [PubMed: 9100430]

Caporino NE, Brodman DM, Kendall PC, Albano AM, Sherrill J, Piacentini J, … Walkup J T. (2013). Defining treatment response and remission in child anxiety: signal detection analysis using the pediatric anxiety rating scale. Journal of the American Academy of Child & Adolescent Psychiatry, 52, 57–67. [PubMed: 23265634]

Caporino NE, Sakolsky D, Brodman DM, McGuire JF, Piacentini J, Peris TS, … Kendall PC (2017). Establishing Clinical Cutoffs for Response and Remission on the Screen for Child Anxiety-Related

Emotional Disorders (SCARED). Journal of the American Academy of Child & Adolescent Psychiatry.

Compton SN, Walkup JT, Albano AM, Piacentini JC, Birmaher B, Sherrill JT, … Waslick BD (2010). Child/adolescent anxiety multimodal study (CAMS): Rationale, design, and methods. Child and Adolescent Psychiatry and Mental Health, 4, 1. [PubMed: 20051130]

Fleiss JL (1981). Statistical methods for rates and proportions (2nd ed.): John Wiley & Sons.

Ginsburg GS, Kendall PC, Sakolsky D, Compton SN, Piacentini J, Albano AM, … March J (2011). Remission after acute treatment in children and adolescents with anxiety disorders: findings from the CAMS. Journal of Consulting & Clinical Psychology, 79, 806–813. [PubMed: 22122292]

Guy W (1976). Clinical Global Impression Scale. Rockville, MD: National Institute of Mental Health.

Ipser JC, Stein DJ, Hawkridge S, & Hoppe L (2009). Pharmacotherapy for anxiety disorders in children and adolescents. Cochrane Database Systematic Review, CD005170.

Johnco CJ, Salloum A, Lewin AB, & Storch EA (2015). Refining clinical judgment of treatment response and symptom remission identification in childhood anxiety using a signal detection analysis on the Pediatric Anxiety Rating Scale. Journal of Child & Adolescent Psychopharmacology, 25, 674–683. [PubMed: 26579629]

Keller MB, Lavori PW, Wunder J, Beardslee WR, Schwartz CE, & Roth J (1992). Chronic course of anxiety disorders in children and adolescents. Journal of the American Academy of Child & Adolescent Psychiatry, 31, 595–599. [PubMed: 1644719]

Kendall PC, Chu B, Gifford A, Hayes C, & Nauta M (1999). Breathing life into a manual: Flexibility and creativity with manual-based treatments. Cognitive & Behavioral Practice, 5, 177–198.

Kendall PC, Compton SN, Walkup JT, Birmaher B, Albano AM, Sherrill J, … Gosch E (2010). Clinical characteristics of anxiety disordered youth. Journal of Anxiety Disorders, 24, 360–365. [PubMed: 20206470]

King NJ, Gullone E, & Ollendick TH (1990). Childhood anxiety disorders and depression: Phenomenology, comorbidity, and intervention issues. Scandinavian Journal of Behaviour Therapy, 19, 59–70.

Landis JR, & Koch GG (1977). The measurement of observer agreement for categorical data. Biometrics, 159–174. [PubMed: 843571]

Langley AK, Falk A, Peris T, Wiley JF, Kendall PC, Ginsburg G, … Piacentini J (2014). The Child Anxiety Impact Scale: Examining parent-and child-reported impairment in child anxiety disorders. Journal of Clinical Child & Adolescent Psychology, 43, 579–591. [PubMed: 23915200]

Lewin AB, Peris TS, De Nadai AS, McCracken JT, & Piacentini J (2012). Agreement between therapists, parents, patients, and independent evaluators on clinical improvement in pediatric obsessive-compulsive disorder. Journal of Consulting & Clinical Psychology, 80, 1103–1107. [PubMed: 22963592]

Liebowitz MR, Turner SM, Piacentini J, Beidel DC, Clarvit SR, Davies SO, … Sallee FR (2002). Fluoxetine in children and adolescents with OCD: a placebo-controlled trial. Journal of the American Academy of Child & Adolescent Psychiatry, 41, 1431–1438. [PubMed: 12447029]

March JS (2013). Multidimensional Anxiey Scale for Children (2nd ed). Toronto, Ontario, Candada: Multi-Health Systems.

Merikangas KR, He JP, Burstein M, Swanson SA, Avenevoli S, Cui L, … Swendsen J (2010). Lifetime prevalence of mental disorders in U.S. adolescents: Results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). Journal of the American Academy of Child & Adolescent Psychiatry, 49, 980–989. [PubMed: 20855043]

Palitz SA, Caporino NE, McGuire JF, Piacentini J, Albano AM, Birmaher B, … Kendall P C. (2018). Defining Treatment Response and Remission in Youth Anxiety: A Signal Detection Analysis with the Multidimensional Anxiety Scale for Children. Journal of the American Academy of Child & Adolescent Psychiatry, 57, 418–427. [PubMed: 29859557]

Piacentini J, Bennett S, Compton SN, Kendall PC, Birmaher B, Albano AM, … Ginsburg G (2014). 24-and 36-week outcomes for the Child/Adolescent Anxiety Multimodal Study (CAMS). Journal of the American Academy of Child & Adolescent Psychiatry, 53, 297–310. [PubMed: 24565357]

Pine DS, Cohen P, Gurley D, Brook J, & Ma Y (1998). The risk for early-adulthood anxiety and depressive disorders in adolescents with anxiety and depressive disorders. Archives of General Psychiatry, 55, 56–64. [PubMed: 9435761]

Podell JL, Mychailyszyn M, Edmunds J, Puleo CM, & Kendall PC (2010). The Coping Cat Program for anxious youth: The FEAR plan comes to life. Cognitive and Behavioral Practice, 17, 132–141.

Regier DA, Rae DS, Narrow WE, Kaelber CT, & Schatzberg AF (1998). Prevalence of anxiety disorders and their comorbidity with mood and addictive disorders. British Journal of Psychiatry Suppl, 34, 24–28.

Research Units on Pediatric Psychopharmacology Anxiety Study Group. (2002). The Pediatric Anxiety Rating Scale (PARS): Development and psychometric properties. Journal of the American Academy of Child & Adolescent Psychiatry, 41, 1061–1069. [PubMed: 12218427]

Shaffer D, Fisher P, Dulcan MK, Davies M, Piacentini J, Schwab-Stone ME, … Bird HR (1996). The NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3): Description, acceptability, prevalence rates, and performance in the MECA study. Journal of the American Academy of Child & Adolescent Psychiatry, 35, 865–877. [PubMed: 8768346]

Silverman WK, & Albano AM (1996). The Anxiety Disorders Interview Schedule for DSM-IV -Child and Parent versions. San Antonio, TX: Graywind Publications.

Silverman WK, & Ollendick TH (2005). Evidence-based assessment of anxiety and its disorders in children and adolescents. Journal of Clinical Child and Adolescent Psychology, 34(3), 380–411. [PubMed: 16026211]

Southam-Gerow MA, Weisz JR, & Kendall PC (2003). Youth with anxiety disorders in research and service clinics: Examining client differences and similarities. Journal of Clinical Child & Adolescent Psychology, 32, 375–385. [PubMed: 12881026]

Spence SH, Barrett PM, & Turner CM (2003). Psychometric properties of the Spence Children's Anxiety Scale with young adolescents. Journal of anxiety disorders, 17(6), 605–625. [PubMed: 14624814]

Stevanovic D (2013). Impact of emotional and behavioral symptoms on quality of life in children and adolescents. Quality of Life Research, 22, 333–337. [PubMed: 22437546]

Storch EA, Salloum A, King MA, Crawford EA, Andel R, McBride NM, & Lewin AB (2015). A Randomized Controlled Trial In Community Mental Health Centers of Computer-Assisted Cognitive Behavioral Therapy Versus Treatment As Usual For Children with Anxiety. Depression & Anxiety, 32, 843–852. [PubMed: 26366886]

Swan AJ, & Kendall PC (2016). Fear and missing out: Youth anxiety and functional outcomes. Clinical Psychology: Science and Practice, 23, 417–435.

Villabø M, Gere M, Torgersen S, March JS, & Kendall PC (2012). Diagnostic efficiency of the child and parent versions of the Multidimensional Anxiety Scale for Children. Journal of Clinical Child & Adolescent Psychology, 41, 75–85. [PubMed: 22233247]

Walkup JT, Albano AM, Piacentini J, Birmaher B, Compton SN, Sherrill JT, … Kendall PC (2008). Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. New England Journal of Medicine, 359, 2753–2766. [PubMed: 18974308]

Wood JJ, Piacentini JC, Bergman RL, McCracken J, & Barrios V (2002). Concurrent validity of the anxiety disorders section of the anxiety disorders interview schedule for DSM-IV: Child and Parent versions. Journal of Clinical Child & Adolescent Psychology, 31, 335–342. [PubMed: 12149971]

Table 1.

Agreement of Treatment Response Between Independent Evaluators and Therapist Ratings Using Benchmarks after Acute Treatment

		Optimal Percent Reduction Benchmarks				
Treatment Response Classification	**Kappa**	**X^2**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
CGI-I Therapist	.48*	100.15*	82%	65%	81%	67%
CGI-I Therapist + SCARED-C Total Score	.41*	90.89*	59%	88%	90%	54%
CGI-I Therapist + SCARED-P Total Score	**.54***	**147.18***	**66%**	**94%**	**95%**	**60%**
CGI-I Therapist + MASC-P Total Score	.52*	137.49*	66%	92%	94%	60%
CGI-I Therapist + SCARED-C Total Score + SCARED-P Total Score	.38*	97.39*	49%	97%	97%	51%
CGI-I Therapist + SCARED-C Total Score + MASC-P Total Score	.36*	86.64*	46%	97%	96%	50%

Note: Algorithm that optimizes therapist agreement with IE ratings after acute treatment is presented in boldface type. Sensitivity = the proportion of true positives, probability of detecting a treatment response, Specificity = the proportion of true negatives, probability of detecting a non-response, Positive Predictive Value (PPV) = the rate at which a treatment response identified by the algorithm is also an IE-rated treatment response, Negative Predictive Value (NPV) = the rate at which a non-response identified by the algorithm is also rated as a non-response by an IE, SCARED = Screen for Child Anxiety Related Emotional Disorders, MASC = Multidimensional Anxiety Scale for Children, CGI-I = Clinical Global Impression of Improvement, CGI-I Therapist = Clinical Global Impression of Improvement rated by treating therapist, C = Child, P = Parent.

* $p<.001$

Table 2.

Agreement of Clinical Remission Between Independent Evaluators and Therapist Ratings Using Benchmarks at after Acute Treatment

		Optimal Total Score Cutoff Benchmarks				
Clinical Remission Classifications	**Kappa**	X^2	**Sensitivity**	**Specificity**	**PPV**	**NPV**
CGI-I Therapist	.38*	65.43*	58%	50%	71%	69%
CGI-I Therapist + SCARED-C Total Score	.37*	76.20*	44%	92%	83%	66%
CGI-I Therapist + SCARED-P Total Score	**.45***	**99.59***	**53%**	**91%**	**83%**	**69%**
CGI-I Therapist + MASC-P Total Score	.41*	81.32*	51%	89%	79%	68%
CGI-I Therapist + SCARED-C Total Score + SCARED-P Total Score	.38*	86.20*	41%	96%	89%	66%
CGI-I Therapist + SCARED-C Total Score + MASC-P Total Score	.35*	71.10*	39%	94%	85%	65%

Note: Algorithm that optimizes therapist agreement with IE ratings after acute treatment is presented in boldface type. Sensitivity = the proportion of true positives, probability of detecting a treatment response, Specificity = the proportion of true negatives, probability of detecting a non-response, Positive Predictive Value (PPV) = the rate at which a treatment response identified by the algorithm is also an IE-rated treatment response, Negative Predictive Value (NPV) = the rate at which a non-response identified by the algorithm is also rated as a non-response by an IE, SCARED = Screen for Child Anxiety Related Emotional Disorders, MASC = Multidimensional Anxiety Scale for Children, CGI-I = Clinical Global Impression of Improvement, CGI-I Therapist = Clinical Global Impression of Improvement rated by treating therapist, C = Child, P = Parent.

*$p<.001$

Table 3.

Agreement of Treatment Response Between Independent Evaluators and Therapist Ratings Using Benchmarks after Maintenance Care

			Optimal Percent Reduction Benchmarks			
Treatment Response Classification	**Kappa**	X^2	**Sensitivity**	**Specificity**	**PPV**	**NPV**
CGI-I Therapist	.07**	4.33**	95%	11%	75%	43%
CGI-I Therapist + SCARED-C Total Score	.27*	34.79*	68%	63%	84%	41%
CGI-I Therapist + SCARED-P Total Score	**.43***	**86.64***	**76%**	**73%**	**89%**	**52%**
CGI-I Therapist + MASC-P Total Score	.36*	60.50*	75%	66%	86%	48%
CGI-I Therapist + SCARED-C Total Score + SCARED-P Total Score	.30*	62.91*	52%	90%	94%	40%
CGI-I Therapist + SCARED-C Total Score + MASC-P Total Score	.28*	52.06*	51%	88%	92%	39%

Note: Algorithm that optimizes therapist agreement with IE ratings after maintenance care is presented in boldface type. Sensitivity = the proportion of true positives, probability of detecting a treatment response, Specificity = the proportion of true negatives, probability of detecting a non-response, Positive Predictive Value (PPV) = the rate at which a treatment response identified by the algorithm is also an IE-rated treatment response, Negative Predictive Value (NPV) = the rate at which a non-response identified by the algorithm is also rated as a non-response by an IE, SCARED = Screen for Child Anxiety Related Emotional Disorders, MASC = Multidimensional Anxiety Scale for Children, CGI-I = Clinical Global Impression of Improvement, CGI-I Therapist = Clinical Global Impression of Improvement rated by treating therapist, C = Child, P = Parent.

* $p<.001$,

** $p<.05$

Depress Anxiety. Author manuscript; available in PMC 2020 August 01.

Table 4.

Agreement of Clinical Remission Between Independent Evaluators and Therapist Ratings Using Benchmarks after Maintenance Care

			Optimal Total Score Cutoff Benchmarks			
Clinical Remission Classifications	**Kappa**	***X*²**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
CGI-I Therapist	.33*	47.50*	71%	61%	71%	62%
CGI-I Therapist + SCARED-C Total Score	.44*	98.92*	56%	90%	89%	61%
CGI-I Therapist + SCARED-P Total Score	**.48***	**108.14***	**65%**	**84%**	**85%**	**65%**
CGI-I Therapist + MASC-P Total Score	.44*	90.30*	64%	82%	82%	63%
CGI-I Therapist + SCARED-C Total Score + SCARED-P Total Score	.44*	106.18*	53%	94%	92%	60%
CGI-I Therapist + SCARED-C Total Score + MASC-P Total Score	.42*	97.76*	51%	94%	91%	59%

Note: Algorithm that optimizes therapist agreement with IE ratings after maintenance care is presented in boldface type. Sensitivity = the proportion of true positives, probability of detecting a treatment response, Specificity = the proportion of true negatives, probability of detecting a non-response, Positive Predictive Value (PPV) = the rate at which a treatment response identified by the algorithm is also an IE-rated treatment response, Negative Predictive Value (NPV) = the rate at which a non-response identified by the algorithm is also rated as a non-response by an IE, SCARED = Screen for Child Anxiety Related Emotional Disorders, MASC = Multidimensional Anxiety Scale for Children, CGI-I = Clinical Global Impression of Improvement, CGI-I Therapist = Clinical Global Impression of Improvement rated by treating therapist, C = Child, P = Parent.

*$p<.001$