Is Treatment Working? Detecting Real Change in the Treatment of Child and Adolescent Depression

Freda F. Liu, PhD,
Molly C. Adrian, PhD

University of Washington School of Medicine and Seattle Children's Hospital, Seattle, WA.

Abstract

Objective: Despite advances in evidence-based treatments for youth depression in recent decades, overall treatment effects are modest at best, with 30% to 50% of youth being nonresponders. Practice parameters consistently recommend systematic assessment and routine monitoring of depressive symptoms, or measurement-based care (MBC), to enhance youth depression treatment. However, the literature offers few guidelines on how to use assessment results to inform care decisions or to detect real and clinically meaningful change. Thus the current study produced reliable change indices (RCIs) per Jacobson and Truax for two commonly used standardized assessments of youth depression (ie, Patient Health Questionnaire–9 items, Modified for Adolescents [PHQ-9A], the Short Moods and Feelings Questionnaire [SMFQ]).

Method: The study sample (N = 1,738) consisted of youths 6 to 18 years old seen in a child and adolescent psychiatry clinic of a regional pediatric medical center who completed at least one of the target depression measures. We examined the factor structure and internal reliability for the PHQ-9A, and calculated RCIs for patients with a depression-related diagnosis for both measures.

Results: Analyses confirmed a one-factor solution and adequate internal consistency ($\alpha = .86$) for the PHQ-9A. All measures yielded acceptable test-retest reliabilities ($r > 0.75$) and RCIs that equate to clinical practice recommendations of using reliable changes scores of 7, 6, and 8 for the PHQ-9A, the SMFQ–Child Report, and the SMFQ–Parent Report, respectively.

Conclusion: Psychometric validation of the PHQ-9A and these RCIs are timely and significant contributions to the treatment of youth depression, by facilitating effective use of MBC—a critical evidence-based strategy for improving treatment outcomes.

Keywords

child and adolescent depression; measurement-based care (MBC); reliable change index (RCI); PHQ-9A; SMFQ

Depression during childhood and adolescence is a critical public health problem. Population-based 1-year prevalence during childhood is low, with less than 1% of youth affected and no differences between boys and girls. During adolescence, the 1-year

Correspondence to Freda F. Liu, 4800 Sand Point Way NE, Seattle, WA 98105; freda.liu@seattlechildrens.org.

prevalence increases, reaching 4% to 5%, with twice as many girls experiencing depression compared to boys. This is approximately the same prevalence as adult depression. By the time youths reach 18 years of age, the cumulative probability of having an episode of clinical depression is 20%.[1] Furthermore, the onset of depression during childhood and adolescence is associated with recurrence later in life.[2] Although 60% to 90% of depressive episodes remit within 1 year, 50% to 70% of youth develop subsequent depressive episodes within 5 years, and few adults show complete symptom and function recovery between episodes.[3] These patterns hold true even for subthreshold depressive symptoms, with similar negative outcomes observed in adulthood.[4] Taken together, depression among youth is an enduring and recurrent disorder.

Because of the established recurrence and chronicity of depression and its impact on attainment of developmental milestones, including educational and social impairments,[2,5] intervening with youth depression is crucial. Treatment for depression during childhood and adolescence continues to improve, with psychotherapy, antidepressant medication (selective serotonin reuptake inhibitors [SSRIs] and serotonin-norepinephrine reuptake inhibitors [SNRIs]), and their combination as well-established evidence-based practices. There have been 2 published meta-analyses of cognitive-behavioral therapy (CBT) based on 35 trials in children and adolescents[6] and 11 in adolescents only,[7] and a larger, more inclusive meta-analysis of youth psychotherapy.[8] All meta-analyses reported that CBT was an effective treatment, but the effect sizes were modest and all below 0.30, which is the lower limit for a moderately effective treatment.[9] Similarly, a meta-analysis of the treatment effects of 36 randomized double-blind trials of SSRI/SNRI with youth less than 18 years of age yielded an effect size of 0.20 when compared to placebo treatment.[10] To contextualize the modest effect sizes of current treatments, approximately 30% to 50% of adolescents who receive any treatment for depressive disorders do not respond.[11] Because of this substantial variability in response to intervention, practice parameters for the treatment of youth depression recommend systematic assessment and routine monitoring of depressive symptoms, or measurement-based care (MBC), to better monitor treatment response and adjust interventions accordingly.[12,13]

Measurement-Based Care

Measurement-based care—the practice of routinely collecting patient-reported outcome measures to inform clinical decision making—has long been shown to effectively enhance the treatment of depression, originally among adult and later in child and adolescent populations across diverse settings (eg, primary care, community mental health).[14-16] In recent years, the field has seen rapidly growing interest as well as regulatory pressures to integrate MBC into routine clinical practice.[17-19] As a transtheoretical and transdiagnostic approach, MBC is an established strategy to effectively improve response and remission rates in youth receiving mental health treatment, which has broad reach and clinical utility.[14] Unfortunately, the scholarly literature has lagged behind in terms of providing guidelines on how to use assessment results to inform clinical decision making, and on an even more fundamental level, how measures can be used to detect real and clinically meaningful change.

Significant Reliable Change

In a recent article, Langkaas *et al.* offer a useful discussion about the different kinds of change that one can assess in MBC.[20] Relevant to our discussion is the distinction they make between observed and detected change. Specifically, observed change refers to the difference that a clinician sees when comparing scores on repeated administrations of a measure (eg, a 3-point difference on the 9-item Patient Health Questionnaire [PHQ-9] from baseline to reassessment). Detected change refers to a difference in scores that is statistically significant (beyond what can be attributed to measurement error). This is also known as reliable change.[21] Jacobson and Traux developed a statistical method for identifying reliable change between two time-points of assessment. This method allowed for the calculation of the minimum change score needed to detect a reliable change, which became known as a reliable change index (RCI). The RCI takes into account the standard error of measurement, which is derived from the standard deviation and test-retest reliability of the measure, and calculates the minimum difference required to be certain, with 95% confidence, that the detected difference is reliable and real beyond what can be attributed to measurement error (see equations 1 and 2).

Despite gaining popularity in the early 1990s as a means of identifying change in psychotherapy for individual patients (in contrast to significant group differences reported in randomized control trials), detecting reliable change has not been well integrated into the practice of MBC. Because the majority of behavior measures were not originally developed for monitoring change, few have published RCIs. Yet without RCIs, clinicians are left to guess whether observed improvements or deteriorations in symptoms are in fact real. Without knowing whether an observed difference is a reliable change, the benefits of routine measurement is greatly diminished. For example, if a clinician administered the PHQ-9 at intake and then again 1 month into treatment and observed that the patient's score had decreased from 15 to 12, without knowing whether this 3-point difference was a reliable change, the clinician would not be able to draw conclusions about whether the patient was improving or symptoms were staying relatively the same (where observed change is likely measurement error).

Moreover, Jacobson and Truax originally introduced the RCI to be used in conjunction with established clinical cutoffs to identify clinically significant change—when reliable change brings a score from the clinical range down to the non-clinical range.[21] Although many well-validated clinical assessment instruments have identified clinical cut-offs, without corresponding RCIs, clinicians would also be unable to determine from their MBC data what constitutes a clinically significant change. Therefore, the availability of appropriate RCIs for commonly used youth depression monitoring measures is crucial for the practice of MBC and improving the quality of depression care for youth.

RCIs for Child/Adolescent Depression Monitoring Measures

Of the brief depression monitoring measures appropriate for youth, only one—PHQ-9—has published RCIs, but only for adult populations.[22-25] In 2004, Lowe *et al.* first published an RCI of 5 points for the PHQ-9; however, this was a study of late-life depression,

and the mean age of participants was 71 years. More recent studies with adults receiving collaborative care[23] or substance use treatment[24] and adults having severe depression[25] yielded RCIs ranging from 5 to 8 points. These differing RCIs clearly demonstrate that RCIs are sensitive to sample variability. They are population specific, such that samples with significantly different standard deviations would yield different RCIs. In other words, RCIs derived for adult populations cannot be assumed to be applicable for youth. To bridge this crucial gap in the literature, the current study aims to derive RCIs for two commonly used clinical instruments for assessing child and adolescent depression symptoms (ie, PHQ-9 Modified for Adolescents [PHQ-9A] and the Short Moods and Feelings Questionnaire [SMFQ]) suitable for MBC. In addition, because the version of the PHQ-9 that is modified for adolescents does not have published psychometrics, another aim of the current study is to provide psychometric validation for the PHQ-9A. This would provide coverage for children as young as 6 and up to 18 years with at least one measure of depressive symptoms with self- or parent-report that can be used in MBC to support clinical decision-making. This archival study was approved by the institutional review board of the authors' home institution.

METHOD

Sample

Data for the current study was pulled from electronic health records (EHR) of patients who received care at a specialty outpatient child and adolescent psychiatry clinic of a regional pediatric tertiary care center in the Pacific Northwest. The overall study sample (N = 1,738) includes families with youths 6 to 18 years of age (mean = 12.89, SD = 3.26), who has at least one completed target depression measure (ie, PHQ-9A, SMFQ–Child Report, or SMFQ–Parent Report) in the electronic measurement feedback system (MFS). The sample consisted of 51.8% female patients. Of the total sample, 8.5% identified as Hispanic, 71.1% white, 4.8% Asian, 2.2% black, 6% multiple races, 7.2% other, and 8.9% declined to respond. The vast majority of these patients (98.2%) preferred to receive care in English. Most patients (78.5%) were covered by commercial insurance, 13.7% Medicaid, 2.6% other government-funded insurance (ie, military or VA), 0.7% self-pay or hospital financial aid, and 4.6% did not have insurance information.

A depression diagnosis was indicated when the patient had a medical record billing code that included one of the following diagnostic codes: depressive disorder NOS (ICD9 code 311; ICD 10: F32.9), major depressive disorder, single episode (ICD9 code 296.2); major depressive disorder, recurrent (ICD9 code: 296.3, ICD10 F33.2); persistent depressive disorder (ICD9 309.0; ICD 10= F34.1), or adjustment disorder with depressed mood (ICD9 309; F43.21). Although 4.7% of our sample did not have billing diagnostic information, one-third of our sample (33.3%, n = 578) met criteria for any depressive disorder diagnosis with the following in each diagnostic category: 29.8% (n = 490) had a depressive disorder diagnosis, 5.1% (n = 91) a recurrent MDD diagnosis, 10.0% (n = 166) a single-episode MDD diagnosis, 2.8% (n = 47) persistent depressive disorder, and 1.1% (n = 18) met criteria for adjustment disorder with depressed mood; 33.2% were diagnosed with more than one depressive disorder.

Procedures

The MBC was delivered via a cloud-based measurement feedback system (MFS), which was implemented in late 2012.[26] Families initiating care at the clinic were all expected to complete an intake form and baseline measures via the MFS. Parents of patients seeking services for internalizing issues (eg, depression and anxiety) were automatically assigned the SMFQ-Parent Report along with an anxiety screening measure to complete in the MFS. Following the initial visit, measure assignment and frequency of administration were entirely up to clinicians' choice, with 7 days being the default frequency for most measures in the MFS. The clinic's guidelines of care for depression indicated that clinicians should administer the SMFQ-Child Report to youth 8 to 11 years old and the PHQ-9A to youth 12 years and older, during initial and follow-up clinic visits. Data extraction involved first pulling all available data on completed administrations of PHQ-9A and SMFQ Child and Parent Reports from the MFS, then pulling administrative data on demographics and diagnostic information from the EHR to match, and finally restricting the age range to those appropriate for our target depression measures (ie, out-of-range administrations were not included, such as 11- or 19-year-olds' completion of the PHQ-9A). This data extraction process yielded 379 patient records with one or more completed PHQ-9A, 342 with one or more SMFQ-Child Reports, and 1,547 with one or more completed SMFQ–Parent Reports. Many patients completed one or more administrations of 2 of these measures (PHQ-9A and SMFQ–Child Report = 224; PHQ-9A and SMFQ–Parent Report = 239; SMFQ–Child and Parent Reports = 186), and some (n = 119) completed all 3 measures.

Measures

PHQ-9A.—The PHQ-9A distills *DSM-IV* depression diagnostic criteria into a brief self-report tool for adolescents and is a recommended screening tool for depression in adolescents 12 to 18 years of age.[27] The tool asks adolescents to rate the frequency of symptoms over the past 2 weeks on a 4-point Likert scale (0 = not at all; 1 = several days, 2 = more than half the days, 3= nearly every day). It is nearly identical to the original PHQ-9 for adults, with the exception of item 7, which alters the language to be developmentally appropriate (ie, asks about concentration difficulties related to "schoolwork and watching TV" as opposed to "work and reading the newspaper"). The diagnostic validity of the PHQ-9 in an adolescent sample was established with 442 adolescents participating in a study on depression outcomes in primary care and enrolled in a large health care delivery system.[16] To the best of our knowledge, no previous study has reported RCIs for the PHQ-9 (or PHQ-9A) with an adolescent sample. Moreover, psychometric properties of the PHQ-9A with the developmentally adjusted item 7 has not been previously reported. Thus for the current study, we examined factor structure and internal reliability for the PHQ-9A in addition to standard deviation and test-retest reliability in service of deriving an RCI.

Short Moods and Feelings Questionnaire (SMFQ).—The SMFQ is a 13-item scale designed to measure depressive symptoms in children and adolescents 8 to 17 years of age. There are parallel parent and child versions, in which respondents use a 3-point Likert rating scale (0 = not true, 1 = sometimes, 2 = true) for symptoms in the past 2 weeks. SMFQ items are derived from *DSM-IV* criteria for major depression and dysthymia.[2] The SMFQ has been shown to have high internal reliability (Cronbach's α = 0.84–0.90),[28,29] and results of

exploratory factor analysis suggest that it is a unifactorial scale.[28,30] High correlations have been found between scores from the SMFQ, the Children's Depression Inventory (CDI), and the Diagnostic Interview Schedule for Children (DISC) depression scale.[28,31] Higher scores indicate more depressive symptoms. Rhew *et al.*, in a large community sample, evaluated predictive properties of cutoff scores using receiver operator characteristics (ROC).[29] At a score of 10 as a cutoff point, the area under the ROC curve (AUC) = 0.86 for the combined child and parent report, sensitivity was 0.76 and specificity was 0.78. A child-report cutoff of 4 resulted in sensitivity of 0.66 and specificity of 0.61. A parent-report cutoff of 4 resulted in sensitivity and specificity of 0.66. In contrast, McKenzie *et al.* examined the SMFQ in a study with 5,769 American and Australian school children (10–15 years of age) and found that using a cutoff score of 7 in the child self-report led to a sensitivity of 0.77 and specificity of 0.68.

Analysis

Psychometric Validation of PHQ-9A.—A maximum likelihood factor analysis was conducted to confirm the one-factor structure of the PHQ-9A consistent with the original PHQ-9. Cronbach α was calculated to examine the internal reliability of the scale.

Deriving RCIs.—Consistent with Jacobson and Truax,[21] we calculated the RCIs with the following formulas, where S_E is the standard error of measurement, s_1 is the standard deviation of baseline assessment, and r_{xx} is the test-retest reliability of the measure.

$$\text{RCI} = 1.96 \times \sqrt{2(S_E^2)}$$

(1)

$$S_E = s_1\sqrt{1 - r_{xx}}$$

(2)

Test-retest reliability for each measure was calculated by identifying the subsample that had completed the first and second administration of each measure within 1 week. With our clinic-based sample, longer lag times would introduce too much treatment effect for a reasonable estimate of test-retest reliability. Given that the RCI is a tool for detecting real change in treatment, it made sense to derive this index from a sample that was clinically affected by the disorder targeted for treatment—in this case, depression. Therefore, only the subsample of patients who had ever had a depression-related diagnosis contributed to the standard deviation calculation of each measure. Analyses for the PHQ-9A, SMFQ–Child Report, and SMFQ–Parent Report were conducted separately.

Finally, because RCIs are population specific and can be affected by sample characteristics (i.e., standard deviation), we also compared standard deviations of these measures (at baseline) by race and ethnicity among patients with depression-related diagnosis to ensure that the RCIs derived would be reasonably stable across the whole sample. Measurement equivalence by race and ethnicity has been previously demonstrated to some extent for

both the PHQ-9 and the Moods and Feelings Questionnaire.[32,33] We hypothesized that there would not be significantly different standard deviations of baseline measures by race and ethnicity.

RESULTS

PHQ-9A Validation

As expected, the maximum likelihood factor analysis confirmed a one-factor solution with no eigenvalues above 1 after the first factor, which accounts for 47.7% of the variance [$\chi^2(27) = 130.36$, $p < .001$]. Table 1 provides eigenvalues and Figure 2 scree plot. Test of internal reliability yielded a Cronbach α of 0.86.

Descriptive Statistics

Table 2 lists the distribution characteristic of measure scores by depressive disorder diagnosis status. The Levene test of variance equality and independent sample t tests, respectively, confirmed that depressed youth has significantly more variability and higher scores than nondepressed youth on both self-report measures, the PHQ-9A and SMFQ–Child Report (Table 2). Consistent with the epidemiology of depressive disorders, youth with a depression-related diagnosis were significantly older [$t(1525) = -20.07$, $p < .01$] and more likely to be female [$\chi^2(1) = 41.54$, $p < .01$] than those without those diagnoses. There were no significant differences between depressed and nondepressed youths in terms of race, ethnicity, and insurance status.

RCIs

Our calculations yielded acceptable test-retest reliabilities ($r > 0.75$) for all 3 measures. These correlation statistics, along with standard deviations, subsample sizes, and RCIs, are all reported in Table 3. Specifically, we found RCIs of 7.15 for the PHQ-9A, 6.45 for the SMFQ–Child Report, and 7.91 for the SMFQ–Parent Report. For clinical decision-making, we would recommend using RCIs of 7, 6, and 8, for the PHQ-9A, the SMFQ–Child Report, and the SMFQ–Parent Report, respectively, given that observed changes will be in whole numbers. To be conservative, a clinician may elect to use an RCI of 7 for the SMFQ–Child Report.

As expected, there were no significant differences in standard deviations of baseline measures by race (white versus racial minority) or ethnicity (Hispanic versus non-Hispanic) among patients with a depression-related diagnosis. Unfortunately, we were not able to effectively examine differences between racial subgroups, given that minority subgroups in the current sample were quite small (<10% each). Nonetheless, the lack of significant differences by race and ethnicity gives us more confidence in the stability of the RCIs derived for the current sample as a whole.

DISCUSSION

Measurement-based care is a recommended best practice for the treatment of depression in children and adolescents. Despite this recommendation, there are few guidelines for how

to use measure results to inform treatment decisions. One critical determination is whether the observed change in repeated measures over time constitutes a reliable improvement or deterioration. Using data collected through routine clinical implementation of MBC, we were able to calculate RCIs for the PHQ-9A and both child- and parent-report forms of the SMFQ. To the best of our knowledge, this is the first reporting of RCIs for child and adolescent depression measures. These RCIs greatly increase the utility of MBC to enhance the treatment of child and adolescent depression. Specifically, the RCIs not only make it possible for clinicians to identify reliable detected change on repeated measures, but also determine clinically significant change when the RCI is used in conjunction with clinical cutoffs.

Measurement-based care is particularly important for depression treatment, given that many youth do not achieve remission of symptoms even with evidence-based treatments, and clinical judgment regarding treatment outcomes without objective measurement tends to be poor.[34] Understanding reliable change for youth on depressive symptom measures is a first step in ensuring that we are matching patients with optimal interventions. Clinicians need to have reliable and valid change data to know when an augmentation or tapering of treatment is needed. In 2011, Gunlicks-Stoessel and Mufson evaluated the timing and amount of change needed to determine whether adolescents would respond to treatment. Their research found that an early assessment (4 weeks) predicted remission better than later assessments (8 or 12 weeks) and a 16% or greater change in Hamilton Rating Scale for Depression was associated with future remission.[35] Gunlicks-Stoessel *et al.* later found that when youth did not meet this level of change, increasing the frequency of psychotherapy proved to have greater impact on depressive symptom.[36]

This study has a few limitations. The retrospective medical record sampling approach used in this archival study favors external over internal validity. In other words, although our data and findings are likely highly representative of conditions in real-world clinical practice, these data were not as clean and complete as they would have been, had they been collected specifically for the purposes of the current study. Similarly, we acknowledge limitations related to the lack of a structured diagnostic interview to confirm depression diagnoses and demographic characteristics of the study participants. In addition, there was some missingness, albeit relatively few, on most demographic variables. Also worth noting is the substantial decrease in sample size between patients with any completed measures and patients having repeated measures within 1 week of baseline assessment (consistent with general trends in MBC implementation).[37] Moreover, because the second administration of the measure was not completed for the purpose of calculating test-retest reliability, the data actually include a range of intervals (2–7 days) between baseline and first repeated administration. In addition, given that our sample consists of patients and families receiving treatment in a specialty mental health clinic, some early intervention effects may be confounding our test-retest reliability and mildly inflating the RCIs.

It is worth noting that this sample of patients and families who have completed measures in the MFS is less diverse in terms of demographic characteristics compared to the overall clinic patient population. A previous study with part of this sample (2012–2013 data) revealed significant MFS access and MBC participation disparities by insurance status.[26] It

is possible that the RCIs reported here would differ if patients had completed depression measures in the MFS at a more equitable rate, and that current results are representative of only those patients and families who were able to reliably access the MFS and participate in MBC.

In conclusion, the goal of this study was to provide developmentally informed RCIs for two commonly used measures of youth depression symptoms in clinical practice. These RCIs, derived with an ambulatory specialty child and adolescent psychiatry sample, provides a useful tool for clinicians' MBC practice. Along with the psychometric validation of the PHQ-9A, these RCIs are a timely and significant contribution to the treatment of child and adolescent depression, to facilitate the meaningful and effective use of MBC as a crucial step in improving outcomes for a common and debilitating condition of youth depression.

Acknowledgments

REFERENCES

1. Hankin BL, Abramson LY, Moffitt TE, Silva PA, McGee R, Angell KE. Development of depression from preadolescence to young adulthood: emerging gender differences in a 10-year longitudinal study. J Abnorm Psychol. 1998;107:128–140. [PubMed: 9505045]

2. Thapar A, Collishaw S, Pine DS, Thapar AK. Depression in adolescence. Lancet. 2012;379:1056–1067. [PubMed: 22305766]

3. Fava GA, Ruini C, Belaise C. The concept of recovery in major depression. Psychol Med. 2007;37:307–317. [PubMed: 17311684]

4. Fergusson DM, Horwood LJ, Ridder EM, Beautrais AL. Subthreshold depression in adolescence and mental health outcomes in adulthood. Arch Gen Psychiatry. 2005;62:66–72. [PubMed: 15630074]

5. Fletcher JM. Adolescent depression: diagnosis, treatment, and educational attainment. Health Econ. 2008;17:1215–1235. [PubMed: 18157910]

6. Weisz JR, McCarty CA, Valeri SM. Effects of psychotherapy for depression in children and adolescents: a meta-analysis. Psychol Bull. 2006;132:132–149. [PubMed: 16435960]

7. Klein JB, Jacobs RH, Reinecke MA. Cognitive-behavioral therapy for adolescent depression: a meta-analytic investigation of changes in effect-size estimates. J Am Acad Child Adolesc Psychiatry. 2007;46:1403–1413. [PubMed: 18049290]

8. Weisz JR, Kuppens S, Ng MY, et al. What five decades of research tells us about the effects of youth psychological therapy: a multilevel meta-analysis and implications for science and practice. Am Psychol. 2017;72:79–117. [PubMed: 28221063]

9. Sullivan GM, Feinn R. Using effect size-or why the p value is not enough. J Grad Med Educ. 2012;4:279–282. [PubMed: 23997866]

10. Locher C, Koechlin H, Zion SR, et al. Efficacy and safety of selective serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors, and placebo for common psychiatric

disorders among children and adolescents: a systematic review and meta-analysis. JAMA Psychiatry. 2017;74:1011–1020. [PubMed: 28854296]

11. TADS Team. Fluoxetine, cognitive–behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents with Depression Study (TADS) randomized controlled trial. J Am Med Assoc. 2004;292:807–820.

12. Birmaher B, Brent D. Practice parameter for the assessment and treatment of children and adolescents with depressive disorders. J Am Acad Child Adolesc Psychiatry. 2007;46:1503–1526. [PubMed: 18049300]

13. Lewandowski RE, Acri MC, Hoagwood KE, et al. Evidence for the management of adolescent depression. Am Acad Pediatrics. 2013;132:e996–e1009.

14. Gondek D, Edbrooke-Childs J, Fink E, et al. Feedback from outcome measures and treatment effectiveness, treatment efficiency, and collaborative practice: a systematic review. Admin Policy Mental Health. 2016;43:325–343.

15. Bickman L, Kelley SD, Breda C, de Andrade AR, Riemer M. Effects of routine feedback to clinicians on mental health outcomes of youths: results of a randomized trial. Psychiatr Serv. 2011;62:1423–1429. [PubMed: 22193788]

16. Richardson LP, Ludman E, McCauley E, et al. Collaborative care for adolescents with depression in primary care: a randomized clinical trial. JAMA. 2014;312:809–816. [PubMed: 25157724]

17. Fortney JC, Unützer J, Wrenn G, et al. A tipping point for measurement-based care. Psychiatr Serv. 2016;68:179–188. [PubMed: 27582237]

18. Patient Protection and Affordable Care Act, 42 U.S.C. § 18001 (2010).

19. The Joint Commission's New Outcome Measures Standards, January 1, 2018.

20. Langkaas TF, Wampold BE, Hoffart A. Five types of clinical difference to monitor in practice. Psychotherapy. 2018;55:241. [PubMed: 30179031]

21. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991;59:12. [PubMed: 2002127]

22. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the Patient Health Questionnaire-9. Med Care. 2004;1194–1201. [PubMed: 15550799]

23. McMillan D, Gilbody S, Richards D. Defining successful treatment outcome in depression using the PHQ-9: a comparison of methods. J Affect Disord. 2010;127:122–129. [PubMed: 20569992]

24. Delgadillo J. Depression and anxiety symptoms: measuring reliable change in alcohol and drug users. Adv Dual Diagn. 2012;5:102–114.

25. Griffiths CA, Griffiths LJ. Recovery and reliable change rates for patients scoring severe on depression, anxiety or impaired functioning in a psychological therapies service: IAPT. Ment Health Rev. 2015;20:28–35.

26. Liu FF, Cruz RA, Lyon AR. Mind the gap: considering disparities in implementing measurement-based care. J Am Acad Child Adolesc Psychiatry. 2019;58:459–461. [PubMed: 30926072]

27. Siu AL. Screening for depression in children and adolescents: US Preventive Services Task Force recommendation statement. Ann Intern Med. 2016;164:360–366. [PubMed: 26858097]

28. Angold A, Costello EJ, Messer SC, Pickles A. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. Int J Methods Psychiatr Res. 1995.

29. Rhew IC, Simpson K, Tracy M, et al. Criterion validity of the Short Mood and Feelings Questionnaire and one-and two-item depression screens in young adolescents. Child Adolesc Psychiatry Ment Health. 2010;4:8. [PubMed: 20181135]

30. Sharp C, Goodyer IM, Croudace TJ. The Short Mood and Feelings Questionnaire (SMFQ): a unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7- through 11-year-old children. J Abnorm Child Psychol. 2006;34:365–377.

31. Kuo ES, Stoep AV, Stewart DG. Using the short mood and feelings questionnaire to detect depression in detained adolescents. Assessment. 2005;12:374–383. [PubMed: 16244118]

32. Galenkamp H, Stronks K, Snijder MB, Derks EM. Measurement invariance testing of the PHQ-9 in a multi-ethnic population in Europe: the HELIUS study. BMC Psychiatry. 2017;17:349. [PubMed: 29065874]

33. Banh MK, Crane PK, Rhew I, et al. Measurement equivalence across racial/ethnic groups of the mood and feelings questionnaire for childhood depression. J Abnorm Child Psychol. 2012;40:353–367. [PubMed: 21996979]

34. Hannan C, Lambert MJ, Harmon C, et al. A lab test and algorithms for identifying clients at risk for treatment failure. J Clin Psychol. 2005;61:155–163. [PubMed: 15609357]

35. Gunlicks-Stoessel M, Mufson L. Early patterns of symptom change signal remission with interpersonal psychotherapy for depressed adolescents. Depress Anxiety. 2011;28:525–531. [PubMed: 21721071]

36. Gunlicks-Stoessel M, Mufson L, Bernstein G, et al. Critical decision points for augmenting interpersonal psychotherapy for depressed adolescents: a pilot SMART. J Am Acad Child Adolesc Psychiatry. 2018.

37. Lewis CC, Boyd M, Puspitasari A, et al. Implementing measurement-based care in behavioral health: a review. JAMA Psychiatry. 2019;76:324–335. [PubMed: 30566197]

Clinical Vignette

Given that reliable change indices (RCIs) are largely unavailable for measures commonly used in measurement-based care (MBC), and that readers are likely unfamiliar with their use in clinical practice, we provide a brief case vignette to demonstrate how the RCIs reported in the current study can be used to support care decisions.

Mikayla is a 14-year-old African American young woman currently in the 9th grade. She was referred for frequent tearfulness, low mood, and significant avolition. Mikayla is typically a strong student and very involved in club soccer. Recently, however, Mikayla has been missing assignments, not participating in class, and talking about giving up on her dreams of playing soccer in college after she endured a stress fracture early in the season. Mikayla's Nine-Item Patient Health Questionnaire Modified for Adolescents (PHQ9A) scores are displayed in Figure 1. At intake, it was 18 out of 30, clearly above the clinical cut-off of 11 recommended for adolescents.

Mikayla and her parents elected to try a course of cognitive-behavioral therapy (CBT) first. She completed another PHQ9A two weeks into treatment, and her score increased by 3 points, now in the moderately severe range. When reviewing her measure results with her clinician, Mikayla asked whether her depression was getting worse. Given that the change score is not greater than 7 point (the RCI), Mikayla's clinician told her that a few points up or down does not necessarily mean that things are changing in meaningful ways and that they should continue to track her symptoms to see how things progress. By week 4, Mikayla's PHQ9A score was down to 14, which is a 7-point drop from her previous score of 21, which is a reliable decrease in symptoms.

Mikayla continues CBT for 8 more weeks maintaining gains but without further reliable symptom reduction. In week 12, Mikayla's clinician collaboratively reviews treatment progress data with Mikayla and her parents, and together they decide to begin augmenting treatment with an SSRI and Mikayla went home with a prescription. By week 14, Mikayla's score is down to the clinical cutoff (11), though it was not a reliable change compared to her Week 12 score. In week 16, a 2-point increase is observed consistent with her subjective experience of limited notable change in her symptoms. Mikayla is encouraged to continue to take her medicine and practice her CBT skills. By Week 18, Mikayla's PHQ9A score is no longer in the clinically significant range. Although this was not a reliable change from her Week 16 score, it is reliably different from Week 12, Week 6, and when Mikayla first entered treatment. Mikayla continues in treatment for another month. As her PHQ9A scores remain in the normative range and are reliably improved from the beginning of treatment, her clinician confidently concludes that Mikayla's symptoms have improved in a clinically significant manner and that her depression is now in remission and she is ready to graduate from care.
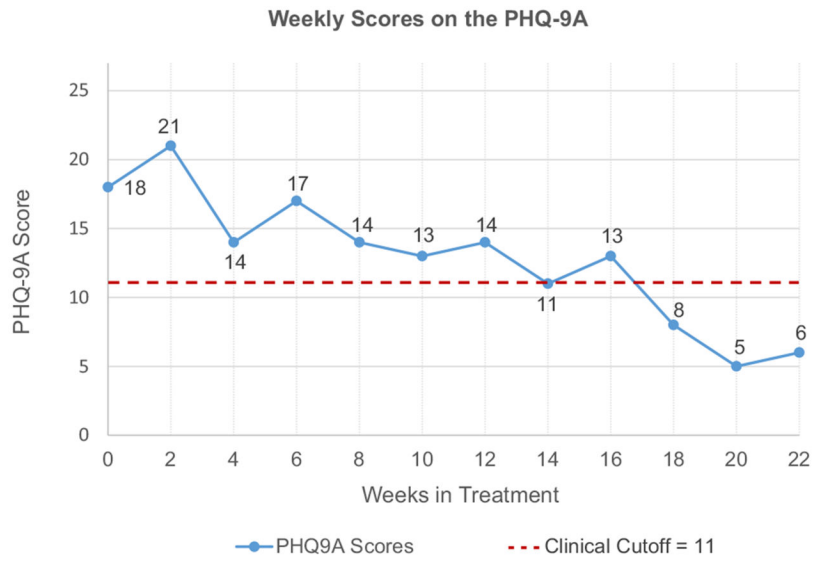
FIGURE 1.

Nine-Item Patient Health Questionnaire Modified for Adolescents (PHQ-9A) Scores Plotted Over Time for Clinical Case Vignette
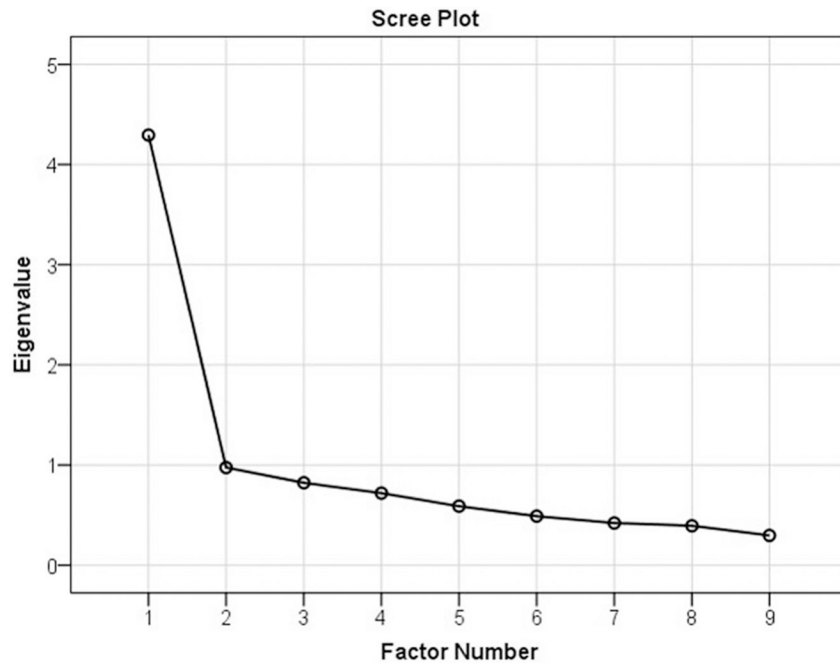
FIGURE 2.
Scree Plot for Factor Analysis

TABLE 1

Eigenvalues

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	4.30	47.72	47.72
2	0.98	10.83	58.56
3	0.82	9.13	67.69
4	0.72	7.99	75.68
5	0.59	6.54	82.22
6	0.49	5.44	87.66
7	0.42	4.68	92.34
8	0.39	4.37	96.71
9	0.30	3.30	100.00

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 2

Descriptives

Measures	Any Depressive Disorder			No Depressive Disorder			Levene Test	t Test
	n	Mean	SD	n	Mean	SD	F	t
PHQ-9A	253	11.99	6.38	119	7.38	5.37	6.26[*]	−7.26[***]
SMFQ-C	206	11.90	6.78	130	7.41	5.71	8.00[**]	−6.27[***]
SMFQ-P	441	8.69	6.05	1,004	8.40	6.19	1.16	− 0.82

Note: n here does not include 4.7% of the sample that is missing diagnostic information. PHQ-9A = Patient Health Questionnaire–9 items, Modified for Adolescents; SMFQ-C = Short Mood and Feelings Questionnaire–Child Report; SMFQ-P = Short Mood and Feelings Questionnaire–Parent Report.

[*]
p < .05

[**]
p < .01

[***]
p < .001.

TABLE 3

Statistics for Reliable Change Indices

	Baseline SD of Depressed Patients (n)	Test-Retest Reliability (n)	RCI	Practical RCI for Clinical Use
PHQ-9A	6.38 (253)	0.84 *** (47)	7.15	7
SMFQ-Child	6.78 (206)	0.88 *** (53)	6.45	6 or 7 (conservative)
SMFQ-Parent	6.05 (441)	0.78 *** (98)	7.91	8

Note: PHQ-9-A = 9 Item Patient Health Questionnaire, Modified for Adolescents; RCI = reliable change indices; SMFQ = Short Moods and Feelings Questionnaire.

***
p < .001.