

UCLA

UCLA Previously Published Works

Title

Interpreting Progress Feedback to Guide Clinical Decision-Making in Children's Mental Health Services

Permalink

<https://escholarship.org/uc/item/97q9b825>

Journal

Administration and Policy in Mental Health and Mental Health Services Research, 43(2)

ISSN

0894-587X

Authors

Tsai, Katherine H
Moskowitz, Andrew L
Brown, Todd E
[et al.](#)

Publication Date

2016-03-01

DOI

10.1007/s10488-015-0630-4

Peer reviewed



HHS Public Access

Author manuscript

Adm Policy Ment Health. Author manuscript; available in PMC 2017 February 09.

Published in final edited form as:

Adm Policy Ment Health. 2016 March ; 43(2): 199–206. doi:10.1007/s10488-015-0630-4.

Interpreting Progress Feedback to Guide Clinical Decision-Making in Children’s Mental Health Services

Katherine H. Tsai, Andrew L. Moskowitz, Todd E. Brown, Alayna L. Park, and Bruce F. Chorpita

Department of Psychology, University of California, Los Angeles, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095, USA

Abstract

Measurement feedback systems (MFSs) can help improve clinical outcomes by enhancing clinical decision-making. Unfortunately, limited information exists to guide the use and interpretation of data from MFSs. This study examined the amount of data that would provide a reasonable and reliable prediction of a client’s rate of symptomatology in order to help inform clinical decision-making processes. Results showed that use of more data predicted greater levels of accuracy. However, there were diminishing returns on the ability for additional data to improve predictive accuracy. Findings inform efforts to develop guidelines on the interpretation of data from MFSs.

Keywords

Evidence-based technology; Clinical reasoning; Feedback; Treatment outcomes; Monitoring

Introduction

Due to the high prevalence of mental health disorders among child and adolescent populations, developing and implementing mental health services with known efficacy and effectiveness is of paramount importance (Chorpita et al. 1998; U.S. Department of Health and Human Services 1999). Thus, over the past 40 years, hundreds of evidence-based treatment (EBT) protocols have emerged to address a myriad of mental health concerns for youths (Chorpita et al. 2011). When tested within the context of randomized clinical trials (RCTs), EBTs have typically been found to outperform usual care services (Weisz et al. 2006). However, despite the wealth of scientific evidence indicating the success of EBTs within research settings (Weisz and Kazdin 2010), endeavors to disseminate and implement such interventions within service systems have encountered numerous challenges (Gold et al. 2006).

One commonly cited challenge is that youth participants in RCTs may not be representative of the youth population that exists within community-based service systems (Garland et al. 2003; Nelson and Steele 2006). When compared with youths who receive treatment within community mental health contexts, youths who receive treatment in research contexts tend

to live in less disadvantaged environments, exhibit lower levels of symptom severity, and present with fewer comorbid conditions (Southam-Gerow et al. 2003). In order to address some of this added case complexity, emerging research has proposed the utility of restructuring the architecture of existing EBTs to allow for real-time adaptations that can better accommodate the diverse demands of community mental health systems (Chorpita et al. 2005; Lyon et al. 2014). For instance, one set of researchers taught providers common therapeutic techniques across standard EBTs and provided them with flexible flowcharts to guide the selection and sequencing of these techniques (Weisz et al. 2012). Yet, the EBT adaptations inherent to this treatment approach calls for the use of feedback on client treatment progress to inform and guide clinical decisions made during treatment delivery to ensure the effectiveness of those adaptations.

Given these emerging treatment approaches, as well as broader evidence indicating that use of feedback may positively impact treatment outcomes in adult and youth populations (Bickman et al. 2011; Knaup et al. 2009; Reese et al. 2009), researchers have posited that use of measurement feedback systems (MFSs) may serve as a viable venue for improving clinical decision making and enhancing the effectiveness of care (Kelley and Bickman 2009; Lambert 2005). MFSs supply providers with real-time data about a client's treatment progress through the use of comprehensive and routinely administered batteries of standardized measures (Bickman 2008a). This real-time information can then be used by providers to refine their interventions to their individual client's specific needs as well as to evaluate the success of their interventions. Information from MFSs has been suggested as a means to assist providers in identifying clients whose status is declining or not improving, and accordingly, prompt them to adapt their treatment plan in a manner that optimizes outcomes (Reese et al. 2009). For instance, feedback has been found to help providers improve the allocation of appropriate levels of support, such that providers increased the frequency of sessions when feedback indicated poor client performance and decreased the frequency of sessions when clients were performing well (Lambert et al. 2001).

Despite findings supporting the promise of MFSs for improving clinical practice, adoption of MFSs within community-based settings has been slow. For instance, researchers have noted various organizational barriers, such as opposition by policy-holders regarding state-wide outcome measurement, a lack of prioritization of adoption of MFSs by funding agencies, and the prohibitive costs of such an endeavor (Bickman 2008b; Landes et al. 2015). Additionally, when attempting to incorporate MFSs into their practices, providers have voiced concerns regarding difficulties in interpreting and applying client-level data stemming from routine feedback (Callaly et al. 2006). For example, one might infer that a steady linear decline in symptomatology may suggest that a provider's treatment plan is effective and thus, the provider should continue with his/her plan. However, treatment response can also occur in other patterns (Hayes et al. 2007), which may make it more difficult for the provider to determine a singular "best" course of action. To expand upon this example, in treatment for post-traumatic stress disorder, a client's level of emotional arousal is expected to increase and fluctuate as the client confronts trauma and learns the appropriate skills to reduce anxiety (Gilboa-Schechtman and Foa 2001). In light of such intricacies, providers have noted that additional support and training in data literacy is necessary to incorporate this client-level outcome data into their clinical work (Callaly et al. 2006).

Current approaches to qualifying the status of a client's treatment progress have mainly focused on exploring the discrepancies between observed values or statuses (e.g., client's reported level of symptomatology) and expected or benchmark values (e.g., clinical level of symptomatology as indicated by the literature; Regan et al. 2013). Examining these discrepancies can help providers discern whether or not their clients are progressing as the literature would expect and, accordingly, whether or not providers should adapt their treatment approach. For instance, one study found that providers who received weekly feedback produced better clinical outcomes relative to providers who did not receive any feedback (Lambert et al. 2001). Expected values of client progress were determined by a reliable change index (RCI) and clinical significance cutoff score, which the researchers created using Jacobson and Truax's (1991) algorithm. The researchers then assigned colored progress markers (e.g., red is indicative that the client is falling below expected progress markers) by creating a graph that compared the client's response to treatment using their initial and change scores (i.e., observed scores) with the RCI and clinical cutoff scores that were created using Jacobson and Truax's (1991) algorithm (i.e., expected scores).

This approach may appear theoretically sound; however, there are a number of treatments and outcomes for which expected values or benchmarks for "good" progress or non-clinical status do not exist. Furthermore, little is known regarding how one best estimates a client's current treatment status (observed value), which is necessary to compare with expected values in order to determine whether "adequate" treatment progress is being made. To this end, the current study aims to examine how many assessments, or *data points*, would be sufficient for providers to formulate a reasonably reliable estimate of their client's current clinical status using a MFS, so as to inform how best to utilize real-time data in these decision making contexts.

Although some researchers suggest the minimum use of two or three data points to discern the effects of treatment (Hayes et al. 1999; Kazdin 2003), the current literature lacks specific guidelines to help providers interpret and apply data from MFSs. Thus, we investigated how well differing numbers of data points predicted subsequent outcome scores. It is also important to consider the balance between the benefits derived through use of additional data in predicting treatment outcomes with the costs associated with acquiring that additional data (Hayes et al. 1999). Thus, we also examined the incremental gain in the strength of the relationship between predicted and observed outcome scores derived from including sequentially more data points.

Given that increasing the amount of data included in prediction models correspondingly improves trend line stabilization (Hayes et al. 1999), it was hypothesized that the inclusion of additional data points would allow for an increasingly more accurate representation of a client's current clinical status. However, it was also hypothesized that there would be diminishing returns as more data were used to predict a client's current clinical status given their potential irrelevance as the earlier data become more distant in time. This study aimed to provide a preliminary examination of the relations of data volume and recency in the estimation of clinical status in a measurement feedback context. Results can potentially inform efforts to develop guidelines on the interpretation and application of data from MFSs to assist therapists with clinical decision-making.

Methods

The present study utilized data collected as part of an ongoing longitudinal treatment study examining the effectiveness of mental health treatments for children and adolescents in Los Angeles County. Families participating in the trial were recruited from among those referred to three different school- and clinic-based community mental health agencies in the county. Prior to any data collection, participating caregivers reviewed and signed institutional review board-approved consent forms, then completed an initial assessment. Eligibility criteria necessitated that children and adolescents have clinically-elevated problem levels in the areas of anxiety, depression, conduct, and/or trauma.

Study Participants

The current study used data from a subsample of families ($n = 61$) from whom weekly assessment data were collected throughout the first 21 weeks of treatment and who were still in treatment at the final assessment point for this study. Within this subsample, youths were predominantly male (59 %) and averaged 9.36 years ($SD = 2.83$) in age. The majority of youths were Hispanic or Latinoa (69 %), followed by individuals of mixed (13 %), African American (13 %), and Caucasian (3 %) race/ethnicity, respectively. One youth (2 %) did not indicate a racial/ethnic category. Caregivers, who elected to complete weekly study assessments, were primarily female (87 %) and averaged 35.96 years ($SD = 9.46$) in age. Families had an average of 4.30 dependents ($SD = 1.59$), with the majority of families (87 %) reporting an annual income of <\$40,000.

Measures

Brief Problem Checklist (BPC)—The BPC (Chorpita et al. 2010) is a 12-item, youth- and caregiver-report measure that assesses youths' internalizing (e.g., "Too fearful or anxious") and externalizing (e.g., "Destroys things belonging to his/her family or others") problems. Administered on a weekly basis, this measure supplies feedback to providers about their client's clinical progress throughout the course of treatment. The BPC was developed by applying item response theory to items adapted from the youth self-report (YSR; Achenbach and Rescorla 2001) and the Child Behavior Checklist (CBCL; Achenbach and Rescorla 2001), to produce youth and caregiver BPC versions, respectively. The caregiver BPC asked informants to rate the youth's behaviors on a three-point Likert scale, ranging from 0 (*not true*) to 2 (*very true*). Internalizing (six items, score range 0–12), externalizing (six items, score range 0–12), and total problems (12 items, score range 0–24) scores can be computed using the sum of all relevant items, such that higher scores indicated more problems. Chorpita et al's (2010) study of 184 children found that the BPC yielded good internal consistency estimates ($\alpha_{\text{internalizing}} = .83$; $\alpha_{\text{externalizing}} = .81$; $\alpha_{\text{total}} = .82$) and test-retest reliability ($r_{\text{internalizing}} = .76$; $r_{\text{externalizing}} = .78$; $r_{\text{total}} = .76$). This study used the caregiver-reported Total Problem scores, which were collected over the phone throughout the duration of their youth's treatment episode.

Statistical Analyses

Our first aim was to examine how many data points would allow service providers to make a reasonable approximation of a client's current clinical status. Current clinical status was

defined as a target assessment point, which varied from the 17th to 21st data point following the baseline assessment. To examine the relationship between predicted clinical status and caregiver-reported ratings of children's problem behaviors, a series of ordinary least squares (OLS) regressions for each individual were fit to predict BPC scores at the time point immediately preceding the target assessment from time in log days, consistent with the approach of Chorpita et al. (2010). Based on these prediction equations, predicted scores were calculated for the target assessment. Predicted and client-reported scores at the target assessment for all clients were then correlated to determine their relationship. In total, 18 models were estimated utilizing different quantities of data ranging from 20 to 3 data points. For reference, data are numbered in descending order relative to the target assessment point. For instance, in the model incorporating the use of the maximum number of data points, we estimated the BPC Total Score for the target time point (21 weeks *after baseline*) from an OLS regression predicting BPC Total Scores from Time for the 20 data points that immediately preceded the target data point (i.e., all data up to the target assessment point of 21 weeks). Similarly, in the model that utilized the minimum number of data points as suggested by previous literature (Hayes et al. 1999), we estimated the BPC Total Score for the target time point (21 weeks *after baseline*) from an OLS regression predicting BPC Total Scores from Time for the three data points immediately preceding the target time point. These point estimates were then correlated with client-reported BPC Total Problem scores from the target assessment points to evaluate how well predicted values matched client-reported values.

The second aim of this study was to determine the incremental gains derived from including additional data points in prediction equations. To this end, change scores between subsequent correlations of predicted and client-reported values were calculated (e.g., difference between correlation coefficients when incorporating 3 vs. 4 data points to examine target assessment at 21 weeks from baseline). These change scores were used to examine relative gains in accuracy of predictions when using varying numbers of data points to predict BPC Total Scores. Positive change scores indicate increased accuracy.

Results

Bivariate correlations were conducted to examine the relationship between observed BPC Total Problem scores and point estimates were calculated using the individual prediction equations. Results demonstrated a significant relationship ($p < .001$) between each observed BPC Total Problem score and all associated point estimates. Correlation coefficients ranged from .65 (three data points predicting target assessment of 20 weeks) to .93 (16 data points predicting target assessment of 17 weeks). Correlations within this range are generally considered to be strong (Cohen 1992). The correlation between predicted and observed values when the target assessment was 20 weeks from baseline was lowest when incorporating three data points ($r = .73$), peaked when incorporating 11 data points ($r = .90$), and then tapered off until all data points were included ($r = .85$). When the target assessment was 19 weeks from baseline, the correlation between predicted and observed values was lowest when incorporating three data points ($r = .65$) and then demonstrated a relatively steady increase until all data points (i.e., 19 data points) were incorporated ($r = .90$). Similarly, when the target assessment was 18 weeks from baseline, the correlation between

predicted and observed values was lowest when incorporating three data points ($r = .67$) and then demonstrated a relatively steady increase until all data points (i.e., 18 data points) were incorporated ($r = .90$). When the target assessment was 17 weeks from baseline, the correlation between predicted and observed values was lowest when incorporating three data points ($r = .85$) and fluctuated until it peaked when all data points (i.e., 17 data points) were incorporated ($r = .92$). Lastly, when the target assessment was 16 weeks from baseline, the correlation between predicted and observed values was lowest when incorporating three data points ($r = .76$) and demonstrated a relatively steady increase until all data points (i.e., 16 data points) were incorporated ($r = .93$).

To examine relative gains in accuracy of predictions, we evaluated change in the correlation of predicted and observed BPC Total scores when an additional point of datum was added to the regression. Results show that across all models, the largest positive incremental gain in correlation coefficients occurred between the inclusion of data points three and four. The correlation coefficients increased by .12, .08, .11, .04, and .06 when examining models predicting target assessment point 20, 19, 18, 17, and 16, respectively. For the entire listing of incremental gain scores, see Table 1.

Discussion

This study aimed to examine the number of data points that would provide a reasonable and reliable prediction of a client's current level of symptomatology. By answering this question, we hoped to inform efforts for developing guidelines regarding the amount of data that would be ideal for providers to consider when determining whether there are adequate or inadequate levels of treatment progress. Additionally, we examined the incremental gains derived from incorporating additional data points into these predictions given that the use of more data may not necessarily improve predictions. It may be important to consider the balance between increasing levels of accuracy and timeliness in the provision of feedback. Results provide us with a better grasp on how one can interpret a client's observed status. Findings are not intended to provide specific guidelines regarding the exact number of data points to consider when determining a client's clinical status—rather, the goal was to provide a preliminary understanding regarding potentially reliable and reasonable approaches to the use of data from MFSs to assist in clinical decision-making.

Across all models, use of any number of data points appeared to have a strong association between predicted and client-reported observed scores. Consistent with our hypothesis, in most models, use of more data appeared to demonstrate greater ability to predict clients' current clinical status. However, the models focused on the 21st target assessment demonstrated a different trend. Although incorporation of additional data points did initially improve the strength of the relationship between predicted and observed scores for the models focused on the 21st target assessment, there was a point at which inclusion of additional data points resulted in an increasingly weaker prediction of the client's clinical status. This suggests that with increasingly large observation periods, there could be a tendency for the accuracy of older data to “expire,” such that a focus on only more temporally proximal observations could lead to better predictions.

In this study, the largest incremental gain across all models was observed between the use of three and four data points. The incremental ability to improve the accuracy of one's prediction of a client's current clinical status appeared to diminish after including the fourth data point. This suggests a need to weigh the balance between data accuracy and data efficiency. Reliance on 1 or 2 weeks of data may not provide sufficient information to formulate a reliable prediction of a client's current clinical status. However, although a provider can in theory continue to improve predictive accuracy with additional data, it may not be feasible or advisable for a provider to gather several months of data before assessing their client's clinical progress, especially given the noted diminishing returns after a few observations.

Though beyond the scope of this study, when interpreting data from MFSs, it is important to understand the expected window of change and to set the data collection schedule according to that timeframe. What is construed as a "recent" change may potentially vary depending on the phenomenon of interest. For example, we might speculate that clinical symptomatology may shift relatively quickly during treatment and thus warrant frequent measurement, whereas attitudes or values toward mental health concerns may take longer to shift. This suggests that the frequency in which data should be collected should depend on the phenomenon of interest given that the rate of change differs across phenomena. Guidelines developed for MFSs should not only be driven by statistical results stemming from client-specific data but should also incorporate alternative data sources, such as information from the psychological literature on clinical decision-making and human-computer design and interaction.

Another consideration for the design of MFSs is the manner in which the data from MFSs are interpreted. For instance, the current analyses assume that providers will draw conclusions regarding their client's current clinical status by visually examining the raw data that is displayed within the graphs generated by the MFS. However, with little empirical knowledge to inform providers as to how they should consume this data, some providers may overinterpret the most recent data point or be overly inclusive of dated information in their clinical decision-making. An alternative approach is to have an external source conduct analyses and then supply interpretive indicators to the provider (e.g., Lambert et al. 2001); however, such analyses must still contend with the rational construction of "best practice" rules for data interpretation and analyses and merely shifts that burden from the provider to the MFS application or its developer. If guidelines and training programs cannot be designed to effectively coach providers to reliably and accurately interpret data from MFSs, it may potentially be more beneficial to have other mental health professionals who specialize in interpretive analyses examine the data and provide subsequent recommendations to the treatment team. Technological solutions, such as advanced algorithms and refined dashboard displays, could also be explored as means to aid this process further. The use of complex algorithms to generate patient-specific recommendations has already been linked to improved provider performance within the medical health care field (Garg et al. 2005). Regardless of the approach taken, it is necessary to bolster research efforts to empirically link resulting interpretations of "poor" and "good" progress to "poor" or "good" status at post-treatment.

In addition to gathering additional evidence on the reliable and valid use of MFSs, enhancements to current MFSs can also be derived by exploring how different design features can alter a therapist's interpretation of the information that is displayed and further support therapists in their clinical decision-making. One possible venue for improving MFSs could be to investigate whether the interpretability of observed data from MFSs can be improved by incorporating information regarding corresponding expected values for the outcome measures in question. For example, a steady linear decline in symptoms may lead a therapist to conclude the effectiveness of their treatment plan, which would lead them to simply continue their course of action. Alternatively, if an MFS was also able to plot an expected treatment trajectory that indicates a much steeper decline in symptoms by the average client who is undergoing the same treatment, a therapist may instead conclude that their client is making insufficient levels of progress and thus attempt to alter their treatment plan. Although this approach follows previous models of feedback interpretation which have focused on the discrepancy between observed and expected values, it is unknown as to whether such an approach may actually improve data interpretability by providers as other concerns, such as potential data overload, can arise.

Limitations

Replication in larger studies on different populations using other MFSs, data collection schedules, and outcome measures can help determine whether these findings are generalizable beyond the confines of this current study. Furthermore, the majority of MFSs are currently used by providers to track the progress of individual clients and do not examine aggregated client data. However, the relationships found in this study may not hold for any particular individual. It may be that for a particular individual, examination of more or less data would provide a better prediction of a client's current clinical status. Findings in this study only provide a gross estimate of the amount of data that may be most reasonable for the majority of individuals. Additionally, less reliable measures may demonstrate poorer estimations with use of a three to four data point range. Our study only examined a single measure of symptomatology and so does not systematically consider measure reliability as a factor in determining ideal data volume for estimating a client's clinical status. It is likely that with less reliable measures, the data requirements to adequately characterize true status changes will increase. Lastly, these results are limited in that they do not include any covariates. Specifically, trajectories may have differed based on additional factors that were not accounted for within the study's prediction equations (e.g., client characteristics), and depending on the covariate's nature, it may impact the amount of data necessary for reasonable prediction. Despite these limitations, these results add to the limited knowledge regarding the interpretation and application of data from MFSs to clinical practice.

Conclusions

In the context of measuring psychopathology symptoms using a brief weekly measure in children and adolescents, we found that use of three to four data points were sufficient to provide a reasonable and reliable estimate of a client's current clinical status. Further research can help the field formulate more concrete guidelines regarding the use of data from MFSs. Currently, the status of clinical processes is frequently determined based on an

examination of the discrepancy between what is observed (e.g., client's reported level of symptomatology) and what is expected (e.g., clinical level of symptomatology as indicated by the literature). The field has been accumulating data on various forms of expected scores, such as clinically significant cutoffs, benchmarks, and group averages, which can be integrated into MFS interfaces, yet less focus has been placed on how one estimates and subsequently interprets the observed trend or status from real-time data. Once reliable and valid methods for interpreting observed data from MFSs can also be established, endeavors can move toward identifying the amount of discrepancy that is likely to be associated with poor outcome at treatment termination.

Given the ubiquitous escalation of health care costs (Bodenheimer 2005) and increasing emphasis on cost containment strategies (Cunningham et al. 2006), there is a need to develop decision support systems that can enhance the efficiency and effectiveness of mental health services. Accumulating evidence indicates the effectiveness of MFSs for improving mental health outcomes (Bickman et al. 2011; Reese et al. 2009); however, implementation of MFSs beyond the context of research has been limited and met with some resistance (Bickman 2008a). By improving our understanding of how, why, and under what conditions MFSs work best, it may be possible to enhance our use of such systems. By examining interpretive strategies for formulating inferences regarding client status when utilizing MFSs, it may be possible to build an evidence-base for standard guidelines, which could then support providers in the use and interpretation of MFSs and ultimately improve their effectiveness.

Acknowledgments

This work was supported by Grant 12-103104-000-USP from the John D. and Catherine T. MacArthur Foundation.

References

- Achenbach, TM., Rescorla, L. Manual for the ASEBA school-age forms & profiles. Burlington: Department of Psychiatry, University of Vermont; 2001.
- Bickman L. A measurement feedback system is necessary to improve mental health outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2008a; 47(10):1114–1119. DOI: 10.1097/CHI.0b013e3181825af8 [PubMed: 20566188]
- Bickman L. Why don't we have effective mental health services? *Administration and Policy in Mental Health and Mental Health Services Research*. 2008b; 35(6):437–439. DOI: 10.1007/s10488-008-0192-9 [PubMed: 18931904]
- 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. *Psychiatric Services*. 2011; 62(12):1423–1429. DOI: 10.1176/appi.ps.002052011 [PubMed: 22193788]
- Bodenheimer T. High and rising health care costs. Part 1: Seeking an explanation. *Annals of Internal Medicine*. 2005; 142(10):847–854. [PubMed: 15897535]
- Callaly T, Hyland M, Coombs T, Trauer T. Routine outcome measurement in public mental health: Results of a clinician survey. *Australian Health Review*. 2006; 30(2):164–173. DOI: 10.1071/AH060164 [PubMed: 16646765]
- Chorpita BF, Barlow DH, Albano AM, Daleiden EL. Methodological strategies in child clinical trials: Advancing the efficacy and effectiveness of psychosocial treatments. *Journal of Abnormal Child Psychology*. 1998; 26(1):7–16. DOI: 10.1023/A:1022626505280 [PubMed: 9566542]

- Chorpita BF, Bernstein A, Daleiden EL. Empirically guided coordination of multiple evidence-based treatments: An illustration of relevance mapping in children's mental health services. *Journal of Consulting and Clinical Psychology*. 2011; 79(4):470.doi: 10.1037/a0023982 [PubMed: 21787050]
- Chorpita BF, Daleiden EL, Weisz JR. Modularity in the design and application of therapeutic interventions. *Applied and Preventive Psychology*. 2005; 11(3):141–156. DOI: 10.1016/j.appsy.2005.05.002
- Chorpita BF, Reise S, Weisz JR, Grubbs K, Becker KD, Krull JL. Evaluation of the Brief Problem Checklist: Child and caregiver interviews to measure clinical progress. *Journal of Consulting and Clinical Psychology*. 2010; 78(4):526.doi: 10.1037/a0019602 [PubMed: 20658809]
- Cohen J. A power primer. *Psychological Bulletin*. 1992; 112(1):155–159. DOI: 10.1037/0033-2909.112.1.155 [PubMed: 19565683]
- Cunningham P, McKenzie K, Taylor EF. The struggle to provide community-based care to low-income people with serious mental illnesses. *Health Affairs*. 2006; 25(3):694–705. DOI: 10.1377/hlthaff.25.3.694 [PubMed: 16684733]
- Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux P, Beyene J, Haynes RB. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. *Journal of American Medical Association*. 2005; 293(10):1223–1238. DOI: 10.1001/jama.293.10.1223
- Garland AF, Kruse M, Aarons GA. Clinicians and outcome measurement: What's the use? *The Journal of Behavioral Health Services & Research*. 2003; 30(4):393–405. DOI: 10.1007/BF02287427 [PubMed: 14593663]
- Gilboa-Schechtman E, Foa EB. Patterns of recovery from trauma: The use of intraindividual analysis. *Journal of Abnormal Psychology*. 2001; 110(3):392–400. DOI: 10.1037//0021-843x.110.3.392 [PubMed: 11502082]
- Gold PB, Glynn SM, Mueser KT. Challenges to implementing and sustaining comprehensive mental health service programs. *Evaluation and the Health Professions*. 2006; 29(2):195–218. DOI: 10.1177/0163278706287345 [PubMed: 16645184]
- Hayes, SC., Barlow, DH., Nelson-Gray, RO. *The scientist practitioner: Research and accountability in the age of managed care*. Needham Heights, MA: Allyn & Bacon; 1999.
- Hayes AM, Laurenceau JP, Feldman G, Strauss JL, Cardaciotto L. Change is not always linear: The study of nonlinear and discontinuous patterns of change in psychotherapy. *Clinical Psychology Review*. 2007; 27(6):715–723. DOI: 10.1016/j.cpr.2007.01.008 [PubMed: 17316941]
- Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*. 1991; 59(1):12.doi: 10.1037/0022-006X.59.1.12 [PubMed: 2002127]
- Kazdin, AE. *Research design in clinical psychology*. Boston, MA: Allyn and Bacon; 2003.
- Kelley SD, Bickman L. Beyond outcomes monitoring: Measurement feedback systems in child and adolescent clinical practice. *Current Opinion in Psychiatry*. 2009; 22(4):363–368. DOI: 10.1097/YCO.0b013e32832c9162 [PubMed: 19417669]
- Knap C, Koesters M, Schoefer D, Becker T, Puschner B. Effect of feedback of treatment outcome in specialist mental healthcare: Meta-analysis. *The British Journal of Psychiatry*. 2009; 195(1):15–22. DOI: 10.1192/bjp.bp.108.053967 [PubMed: 19567889]
- Lambert MJ. Emerging methods for providing clinicians with timely feedback on treatment effectiveness: An introduction. *Journal of Clinical Psychology*. 2005; 61(2):141–144. DOI: 10.1002/jclp.20106
- Lambert MJ, Whipple JL, Smart DW, Vermeersch DA, Nielsen SL, Hawkins EJ. The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research*. 2001; 11(1):49–68. DOI: 10.1080/713663852 [PubMed: 25849877]
- Landes SJ, Carlson EB, Ruzek JI, Wang D, Hugo E, DeGaetano N, et al. Provider-driven development of a measurement feedback system to enhance measurement-based care in va mental health. *Cognitive and Behavioral Practice*. 2015; 22(1):87–100. DOI: 10.1016/j.cbpra.2014.06.004

- Lyon AR, Lau AS, McCauley E, Vander Stoep A, Chorpita BF. A case for modular design: Implications for implementing evidence-based interventions with culturally diverse youth. *Professional Psychology: Research and Practice*. 2014; 45(1):57. doi: 10.1037/a0035301
- Nelson TD, Steele RG. Beyond efficacy and effectiveness: A multifaceted approach to treatment evaluation. *Professional Psychology: Research and Practice*. 2006; 37(4):389–397. DOI: 10.1037/0735-7028.37.4.389
- Reese RJ, Norsworthy LA, Rowlands SR. Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy: Theory, Research, Practice, Training*. 2009; 46(4):418–431. DOI: 10.1037/a0017901
- Regan J, Daleiden EL, Chorpita BF. Integrity in mental health systems: An expanded framework for managing uncertainty in clinical care. *Clinical Psychology: Science and Practice*. 2013; 20(1):78–98. DOI: 10.1111/cpsp.12024
- Southam-Gerow MA, Weisz JR, Kendall PC. Youth with anxiety disorders in research and service clinics: Examining client differences and similarities. *Journal of Clinical Child and Adolescent Psychology*. 2003; 32(3):375–385. DOI: 10.1207/S15374424JCCP3203_06 [PubMed: 12881026]
- U.S. Department of Health and Human Services. *Mental health: A report of the surgeon general*. Rockville: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health; 1999.
- Weisz JR, Chorpita BF, Palinkas LA, Schoenwald SK, Miranda J, Bearman SK, et al. Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth: A randomized effectiveness trial. *Archives of General Psychiatry*. 2012; 69(3):274–282. DOI: 10.1001/archgenpsychiatry.2011.147 [PubMed: 22065252]
- Weisz JR, Jensen-Doss A, Hawley KM. Evidence-based youth psychotherapies versus usual clinical care: A metaanalysis of direct comparisons. *American Psychologist*. 2006; 61(7):671. doi: 10.1037/0003-066X.61.7.671 [PubMed: 17032068]
- Weisz, JR., Kazdin, AE. *Evidence-based psychotherapies for children and adolescents*. New York: Guilford Press; 2010.

Table 1

Correlation values (*r*) between predicted and observed BPC scores

Number of time points used in prediction equation	Target assessment point (years)					
	21	20	19	18	17	17
Correlation (<i>x</i>)						
3	0.7333	0.6484	0.6728	0.8524	0.7567	
Incremental gain						
3-4	0.1210	0.0785	0.1061	0.0382	0.0632	
4-5	0.0284	0.0348	0.0458	-0.0208	0.0214	
5-6	-0.0002	0.0258	0.0109	0.0019	0.0023	
6-7	0.0107	-0.0023	0.0046	-0.0035	0.0248	
7-8	0.0016	0.0117	0.0139	0.0204	0.0135	
8-9	-0.0031	-0.0155	-0.0028	0.0097	0.0092	
9-10	-0.0064	0.0007	0.0011	0.0004	-0.0059	
10-11	0.0099	0.0018	0.0048	0.0037	0.0038	
11-12	-0.0037	0.0145	0.0062	-0.0027	-0.0028	
12-13	-0.0140	0.0149	0.0006	-0.0028	-0.0038	
13-14	0.0049	0.0120	0.0041	-0.0039	-0.0009	
14-15	-0.0150	0.0024	0.0056	0.0009	-0.0045	
15-16	-0.0069	0.0062	-0.0035	-0.0003	0.0524	
16-17	-0.0004	0.0059	-0.0046	0.0268		
17-18	-0.0025	0.0024	0.0389			
18-19	-0.0061	0.0568				
19-20	0.0034					

Bolded numbers indicate the largest incremental gain for each assessment point. Incremental gain refers to the change scores between subsequent correlations of predicted and observed values (e.g., 3-4 = difference between correlation coefficients when incorporating 3 vs. 4 data points)