Assessing practical implementation of modular psychotherapy for youth in community-based settings using benchmarking

Daniel M. Cheron¹, Emily M. Becker-Haimes², H. Gemma Stern³, Aberdine R. Dwight⁴, Cameo F. Stanick⁵, Angela W. Chiu⁶, Eric L. Daleiden⁷ and Bruce F. Chorpita⁸

Abstract

Background
Achieving high quality outcomes in a community context requires the strategic coordination of many activities in a service system, involving families, clinicians, supervisors, and administrators. In modern implementation trials, the therapy itself is guided by a treatment manual; however, structured supports for other parts of the service system may remain less well-articulated (e.g., supervision, administrative policies for planning and review, information/feedback flow, resource availability). This implementation trial investigated how a psychosocial intervention performed when those non-therapy supports were not structured by a research team, but were instead provided as part of a scalable industrial implementation, testing whether outcomes achieved would meet benchmarks from published research trials.

Method
In this single-arm observational benchmarking study, a total of 59 community clinicians were trained in the Modular Approach to Therapy for Children (MATCH) treatment program. These clinicians delivered MATCH treatment to 166 youth ages 6 to 17 naturally presenting for psychotherapy services. Clinicians received substantially fewer supports from the treatment developers or research team than in the original MATCH trials and instead relied on explicit process management tools to facilitate implementation. Prior RCTs of MATCH were used to benchmark the results of the current initiative. Client improvement was assessed using the Top Problems Assessment and Brief Problem Monitor.

Results
Analysis of client symptom change indicated that youth experienced improvement equal to or better than the experimental condition in published research trials. Similarly, caregiver-reported outcomes were generally comparable to those in published trials.

¹Vice President of Clinical Programs, Implementation, & Training, Judge Baker Children’s Center, Boston, Massachusetts, USA
²University of Pennsylvania, Philadelphia, Pennsylvania, USA
³Rutgers University, New Brunswick, New Jersey, USA
⁴Northeastern University, Boston, Massachusetts, USA
⁵Vice President of Clinical Practice, Training, and Research and Evaluation, Hathaway-Sycamores Child and Family Services, Pasadena, California, USA
⁶Assistant Professor of Psychology in Clinical Psychiatry, Weill Cornell Medicine, New York, New York, USA
⁷Partner, PracticeWise, LLC, Satellite Beach, Florida, USA
⁸Professor of Psychology, The University of California, Los Angeles, Los Angeles, California, USA

Corresponding author:
Daniel M. Cheron, Judge Baker Children’s Center, 53 Parker Hill Avenue, Boston, MA 02120, USA.
Email: dcheron@jbcc.harvard.edu

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Conclusions
Although results must be interpreted cautiously, they support the feasibility of using process management tools to facilitate the successful implementation of MATCH outside the context of a formal research or funded implementation trial. Further, these results illustrate the value of benchmarking as a method to evaluation industrial implementation efforts.

Plain Language Summary: Randomized effectiveness trials are inclusive of clinicians and cases that are routinely encountered in community-based settings, while continuing to rely on the research team for both clinical and administrative guidance. As a result, the field still struggles to understand what might be needed to support sustainable implementation and how interventions will perform when brought to scale in community settings without those clinical trial supports. Alternative approaches are needed to delineate and provide the clinical and operational support needed for implementation and to efficiently evaluate how evidence-based treatments perform. Benchmarking findings in the community against findings of more rigorous clinical trials is one such approach. This paper offers two main contributions to the literature. First, it provides an example of how benchmarking is used to evaluate how the Modular Approach to Therapy for Children (MATCH) treatment program performed outside the context of a research trial. Second, this study demonstrates that MATCH produced comparable symptom improvements to those seen in the original research trials and describes the implementation strategies associated with this success. In particular, although clinicians in this study had less rigorous expert clinical supervision as compared with the original trials, clinicians were provided with process management tools to support implementation. This study highlights the importance of evaluating the performance of intervention programs when brought to scale in community-based settings. This study also provides support for the use of process management tools to assist providers in effective implementation.

Keywords
Modular treatment, community-based, benchmarking, process management, youth

Efficacy and effectiveness clinical trials are conducted in conjunction with substantial implementation support (Rudd et al., 2020). Studies continue to suggest mixed results for treatments’ success when these empirical approaches are tested in settings with more limited clinical and operational supports (e.g., Weisz et al., 2012; Southam-Gerow et al., 2010, Weisz et al., 2020). There are characteristics of community settings at the client, clinician, and agency-level that must also be thoughtfully addressed in the same way a treatment manual addresses psychopathology (see Weisz et al., 2013; Danschroder et al., 2009). For example, community settings often include far greater demographic diversity than their randomized, controlled trial (RCT) counterparts (U.S. Department of Health and Human Services, 2001), and greater amounts of diagnostic comorbidity (Southam-Gerow et al., 2003). Clinicians who work in community settings face greater productivity expectations than those funded by a research trial. Agencies providing community services also face greater challenges deriving income from fee-for-service reimbursements than those funded through a research trial (Stewart et al., 2016), and there can be considerable variability among organizations regarding the business practices that support successful implementation (e.g., supervision practices, use of outcome monitoring, structured documentation procedures, evidence-based planning and case review procedures; Daleiden & Chorpita, 2005).

When working in the context of an effectiveness research trial, many of these challenges and management strategies can be addressed by researchers regarding the implementation of the specific study interventions (e.g., postdoctoral researchers providing clinical consultation to community clinicians). Thus, it is less clear how to support successful industrial implementation, i.e., implementation conducted explicitly for increasing clinical service delivery of a chosen program, led by someone other than a research team, and not having an explicit research agenda. Further work needs to be done to develop implementation strategies that specifically target the characteristics of community settings at the client, clinician, and agency-level likely to support implementation and to understand the impact these strategies may have on intervention outcomes.

A related question involves how best to investigate the success of an implementation when these formal research supports are explicitly absent (e.g., Weersing & Hamilton, 2005). The desire to determine the effectiveness of a treatment (e.g., through use of randomization; clinical staff selection and/or management) can be at odds with the necessity of implementing research-supported treatment in community settings within the confines of their fiscal and administrative realities. Benchmarking has been proposed as a useful solution to address this problem. Benchmarking compares outcome data obtained in a clinical service setting (i.e., community practice) with data obtained from existing research from rigorous trials (i.e., RCTs) known to demonstrate significant results to evaluate
performance (McFall, 1996; Wade et al., 1998). Benchmarking, like other observational research, is not without its limitations. Threats to internal validity and alternate explanations for the outcomes observed may exist, as described by Weersing & Hamilton (2005). However, benchmarking can advance our understanding of how treatments perform within the natural characteristics of community settings and can help determine whether desired outcomes can be achieved without formal supports from the research team or developer.

This practical implementation report compared the outcomes of an industrial implementation of modular treatment for youth emotional and behavioral problems in a multi-site community mental health center (CMHC) with outcomes from previous RCTs that had demonstrated treatment effectiveness. The industrial implementation was a single-arm observational investigation of the effectiveness of the Modular Approach to Therapy for Children with Anxiety, Depression, Trauma and Conduct Problems (MATCH; Chorpita & Weisz, 2009) delivered in a multi-site community-based program. The current implementation of MATCH described here also represents the first use of explicitly developed clinical and operational process management support tools to facilitate adaptation and transfer of knowledge from treatment developers to local agency providers. These tools included a professional learning portfolio, a process management roadmap, and specific additions to the training curriculum dedicated to implementation management techniques. The purpose of these tools was to support robust scale-up as the implementation becomes industrialized and moves away from reliance on an academic implementation team. Implementation strategies used are described in Table 1 using the AACTT Framework for Implementation Interventions (Presseau et al., 2019).

MATCH is a flexible program comprised of 33 modules, each representing a therapeutic skill derived from previously developed evidence-based practices (EBP). The MATCH program organizes these modules into flexible approaches to treating anxiety, depression, traumatic stress, or conduct problems, so that when youth face multiple problems, treatment techniques from any area may be applied as needed during psychotherapy. MATCH is an established evidence-based program with multiple trials indicating that youth and families who receive MATCH see significant improvements in internalizing and externalizing symptoms at a quicker rate than those receiving usual care and standard manual-based EBPs (Chorpita et al., 2017; Weisz et al., 2012), and demonstrate a greater reduction in the number of problem areas faced by youth following treatment compared with those receiving usual care services (Weisz et al., 2012). Families who receive MATCH also demonstrate sustained improvements relative to usual care two years after receiving treatment (Chorpita et al., 2013), and utilize fewer additional behavioral health services and types of psychotropic medications (Park et al., 2016).

Importantly, a more recent trial of MATCH that used a different randomization strategy (individual instead of cluster randomization) and a different consultation format (group instead of individual) did not support the comparative effectiveness of MATCH relative to usual care (Weisz et al., 2020). This study illuminated how the potency of an intervention could be dependent on implementation factors, suggesting that shifting to less individualized (and less labor-intensive) implementation support may

---

Table 1. Implementation interventions utilized by project.

| Implementation Actions          | Implementation Actor, Context, Target, and Time |
|--------------------------------|-----------------------------------------------|
| MATCH Clinical Training        | MATCH Purveyor, in person, agency providers, 4 ½ days. | MATCH Developer, in person, agency providers, 6 days. | MATCH Developer, in person, agency providers, 6 days. |
| MATCH Consultation             | MATCH Purveyor, 100% virtual, groups of agency providers, bi-weekly for 12 months. | MATCH Developer, 100% in-person, individual agency providers, weekly for up to 2 years. | MATCH Developer, 80% virtual/20% in-person, groups of agency providers, weekly for up to 2 years. |
| Decision Support Data System   | Automated digital software surveys, electronically, youth and families, weekly. | Research assistants, individual telephone calls, youth and families, weekly. | Research assistants, individual telephone calls, youth and families, weekly. |
| Specific Training on           | MATCH Purveyor, in person, agency leaders and providers, ½ day | N/A | N/A |
| Implementation Management      | Agency providers and supervisors, Self-guided portfolio, Agency Providers, as needed | MATCH Developers, audiorecorded sessions, individual agency providers, weekly | MATCH Developers, audiorecorded sessions, individual agency providers, as needed |
| Process Management             | Agency providers and supervisors, Self-guided process management roadmap, MATCH providers, as needed | N/A | N/A |
have attenuated treatment outcomes. Specifically, the authors proposed that individual consultation, although not feasible at scale, allows for treatment developers to more closely monitor and guide clinician learning of the intervention. Weisz et al. (2020) therefore, concluded that moving to a group consultation format may have diluted the potency of consultation.

All three MATCH RCTs referenced above were conducted prior to the creation of process management tools used in the implementation effort described here. We hypothesized that when the clinical and operational supports typically provided in an RCT context were replaced by process management tools from an industrial implementation team, we could achieve similar outcomes to those obtained in research trials. To determine whether youth receiving MATCH in this study demonstrated similar improvement and to understand the potential impact that process management may have on outcomes, we benchmarked outcomes of these youth to those of two prior randomized effectiveness trials (Weisz et al., 2012; 2020). The 2012 and 2020 trials were selected for comparison because they represent the strongest demographic and geographic concordance with the present study and also used the same measurement model as that of the current implementation initiative. Furthermore, the 2012 and 2020 trials differed in the type of implementation supports (e.g., individual vs. group consultation) and did not utilize any of the clinical and operational support tools used in this implementation. Instead, clinical and operational support was provided directly by the research team, typically in weekly consultation conducted in a hybrid virtual/in-person format for up to 2 years. Clinicians in the prior MATCH trials received consultation in the model that was appropriate for a clinical trial, where demonstrating high-quality outcomes on a relatively small number of participants was a primary aim. This approach, however, would be considered expensive and difficult-to-scale outside of a research trial context. Specifically, consultants received extensive weekly guidance directly from the treatment developers along with support managing an information system dashboard. Clients received weekly telephone calls from research assistants to collect dashboard data and were compensated for their time completing these surveys.

In contrast, in this MATCH initiative, there was no team of investigators, postdoctoral fellows, or research assistants available to call clients weekly or support the information system dashboard. Clients received no compensation for completing surveys, and clinicians received no compensation for attending consultation. The present investigation substituted this extensive expert consultation with more abbreviated consultation (virtual group biweekly calls for 12 months), as well as a) a self-guided professional development portfolio to facilitate the development of competencies in the program; b) a self-guided process management roadmap to support process management of individual cases; and c) an explicit half-day training in adapting implementation to the inner context of each provider’s agency. This approach to implementation allowed the benchmarking analysis to evaluate whether industrial implementation resources were not only feasible but also able to support clinical outcomes comparable to those found in research trials.

**Methods**

**Study Design**

This study leveraged evaluation data obtained through a private philanthropic foundation-funded initiative to increase the availability of EBPs within a specific geographic area by implementing MATCH. This initiative provided funding to train clinicians in MATCH using less-extensive clinical and operational supports than those in the effectiveness RCTs and without a research agenda. During the initiative’s planning, additional funding was secured to support the formal assessment of whether such an EBP training program would increase access to EBPs, improve youth outcomes, and reduce treatment costs. Of note, this funding was primarily provided to determine the impact of improving EBP availability on youth in the area, and not to test the effectiveness of the treatment. Data on clinical trajectories of youth participating in this implementation initiative were analyzed and benchmarked against clinical trajectories from Weisz et al. (2012 and 2020).

The presence of philanthropic funding to support this training should not be confused with the scale of funding, scope, and resources available to RCTs. Indeed, implementing new practices takes resources; we hypothesize that the inclusion of process management tools may help supplant the research-intensive resources, direct more dollars towards increasing the scope of impact, and arrive at similar clinical outcomes.

**Setting**

Prior to selecting a provider organization, the authors conducted interviews with several agencies in the geographic region of interest. The selection criteria for choosing an agency included serving youth and families in the geographic region, serving a sufficient number of youth to support the initiative, employing enough clinical providers to support the initiative, and expressing a willingness to commit to the initiative. Only one of four agencies approached met all inclusion criteria. This single, private, multi-site, for-profit human service agency provided behavioral health services to more than 30,000 children, adolescents, adults, and families annually across approximately 16 sites, representing the largest single provider in the geographic region. The agency provided services through both private and public insurance reimbursement. Seven of the 16 service sites were selected for MATCH
implementation because they were, based on the agency’s perspective, a) within the geographic area of interest for the foundation and b) had enough clinicians seeing youth clients to make training clinicians in MATCH feasible.

Implementation Model
A complete description of the implementation process, training curriculum, and implementation successes and challenges has been previously reported (citation blinded for review). Briefly, clinicians were trained in two separate cohorts separated by approximately one year. Training in the MATCH model consisted of five consecutive days of face-to-face training including didactic presentations and experiential activities (e.g., modeling and role-playing techniques). Following the clinical training, clinicians attended 25 telephone-based consultation sessions in groups of eight clinicians provided every other week. A select subgroup of clinicians was further trained as in-house supervisors/trainers of the model to maintain sustainability after the end of the funded initiative. Following their training, clinicians began enrolling clients in MATCH. Since the initiative was designed to increase the amount of EBPs being provided in the area, clinicians were free to begin using MATCH with existing clients or new clients. Additionally, they were encouraged to start MATCH with only a few clients while they were learning the techniques, and then gradually increase the percentage of their caseload with whom they utilized MATCH. Clinicians were provided with guidance during the training about clients that would likely be a good fit for the program, but clinicians were ultimately allowed to independently determine if a client was provided with MATCH.

Throughout the initiative, clinicians and supervisors were provided with process management tools designed to facilitate successful implementation of the MATCH program by enhancing the Competency drivers of implementation (i.e., training, coaching, and fidelity; NIRN, 2017). The professional development portfolio is akin to a structured curriculum vitae that helps organize the developmental process, monitor and track trainee experiences and expertise, and facilitate formative and summative evaluation of direct provider and supervisor achievements. The process management roadmap is a clinical decision-making framework that provides reference to case-specific information needed and actions to be taken at key decision points in treatment. The instructional changes include specific didactic and experiential training designed to help clinicians better adapt the MATCH curriculum to their unique practice setting and client population. All three of these tools were not present in any of the prior MATCH trials and were added in this initiative to facilitate implementation in the absence of intensive support (i.e., weekly consultant with program developers for the entire course of the project; research assistants calling families weekly to collect progress data) typically provided in a research setting.

Participants
Clinicians
A total of 59 community-based masters-level clinicians were trained in MATCH over the course of two years at the seven sites within the agency. Clinicians provided MATCH services in outpatient-, schools-, or home-based modalities, with many clinicians operating in more than one service-delivery setting. Participating clinicians averaged 29.7 (SD = 5.34) years old with three years of clinical experience. Clinicians were 87% female, predominantly white (92%), and most clinicians reported a professional background in counseling (53%). These demographic data parallel those of the larger mental health workforce (Hewitt et al., 2008). All clinicians completed the initial MATCH training, enrolled at least one client, and were included in the intent-to-treat analyses, regardless of whether they completed consultation. Twelve clinicians (20%) either left the agency or transferred to non-clinical positions during the course of the consultation.

Clients
A total of 166 youth ages 6 to 17 (M = 12.46, SD = 2.64) naturally presenting for outpatient, school, or home-based psychotherapy services were provided MATCH. Clients were predominantly white (59%) and 50% were male. All clients were receiving publicly supported reimbursement for services through Medicaid. The average number of sessions attended per client was 18.92 (SD = 14.46) delivered over a mean treatment length of 193.66 (SD = 134.61) days. Youth were included in analyses regardless of whether they completed treatment. Written informed consent was obtained from all participants or guardians and all study procedures were overseen and conducted in compliance with the Institutional Review Board at the first author’s organization (H100913).

Variables and Data Sources
Data were collected through a progress monitoring and feedback system, also funded through the foundation support for this initiative, and were intentionally designed to support future benchmarking against original trials. The cloud-based application surveyed clients and caregivers weekly after the start of psychotherapy. Surveys included the Top Problems Assessment (TPA; Weisz et al., 2011) as well as the Brief Problem Monitor (BPM; Achenbach et al., 2011). The TPA is a brief, idiographic assessment of chief complaints identified by the child and the caregiver independently. The test-retest reliability, convergent and discriminant validity, sensitivity to change, slope reliability, and the association of Top Problems slopes with standardized measure slopes all support the psychometric
strength of the TPA (Weisz et al., 2011). At the start of services, the child and the caregiver were separately prompted to identify their top three chief concerns and rate the current severity from 0 (not a problem) to 10 (severe problem). Clinicians were trained to elicit top problems that were measurable, observable, and changeable in psychotherapy and use developmentally appropriate language familiar to the child and caregiver. Children younger than eight years old typically were unable to provide top problems independently, and only caregiver problems were collected in those instances. The BPM is a 19-item version of the Child Behavior Checklist and Youth Self Report. Respondents answer on a scale from 0–2 the degree to which a statement has been true in the past week. Responses include “Not at All” (0), “Sometimes” (1), and “Often” (2). The BPM Total Scale is the combined score of all items. The Internalizing, Externalizing, and Attention Subscales are calculated from this total scale and represent a subset of relevant items measuring anxiety, depression, and traumatic stress (Internalizing Subscale), disruptive behavior (Externalizing Subscale), and lack of concentration and inattention (Attention Subscale). Internal consistency of the BPM is satisfactory (Cronbach’s alpha = 0.78 to 0.91), and high correlations were demonstrated between the CBCL and BPM (r = 0.86 to 0.95). The BPM and Subscales were sensitive and identified significantly higher behavioral and emotional problems among children whose caregiver reported a psychiatric diagnosis (Piper et al., 2014).

The TPA and BPM were administered through the digital progress monitoring and feedback system. Surveys were electronically sent to youth and families every seven days. However, on average, youth completed a total of 13.5 (SD = 10.1) surveys during a course of treatment, and a survey was completed every 14.1 (SD = 17.0) days on average. Caregivers completed a total of 15.9 (SD = 13.5) surveys during a course of treatment and a survey was completed approximately every 13.4 days (SD = 16.4). Surveys were completed independently at home via email by the youth and caregiver approximately 25% and 57% of the time, respectively, and were surveyed upon arrival at their session the remaining 75% and 43% of the time, respectively.

**Analytic Plan**

Initial analyses compared youth demographics from the current trial to the first MATCH effectiveness trial using t-tests and Pearson chi-square statistics. To benchmark outcomes in the current trial relative to the 2012 and 2020 MATCH effectiveness trials, data analysis followed the same methods used in those trials (Weisz et al., 2012, 2020). Hierarchical linear modeling using SAS Version 9.2 examined trajectories of change in client symptoms over time on measured outcome variables (Average Top Problems, BPM Internalizing Subscale, the BPM externalizing Subscale, and a BPM Total score comprising the Internalizing and Externalizing Subscales). To determine whether the current sample’s effect sizes were comparable to the 2012 and 2020 trials, independent sample t-tests were conducted on slope estimates from this sample and the 2012 and 2020 trials, as slope estimates served as a common metric across all studies. Projected 1-year change values were also calculated in the absence of follow-up data; this statistic was selected as it is an easily interpreted effect size metric that closely coordinates with the RCTs upon which this study benchmarks. It is important to note that these are projected change values based on estimated slope, as the average length of treatment was only approximately 6 months. Given tendencies for caregivers and youth to provide unique perspectives on youth symptoms (De Los Reyes & Ohannessian, 2013), and consistent with the 2012 and 2020 trials, models were run separately for youth and caregiver informant reports of symptoms. Time was indexed in log days, consistent with both previous trials.

The 2012 and 2020 trials used 2-level hierarchical linear models to index symptom change over the treatment period. To determine whether a 2-level analysis was also suitable in the current sample, intra-cluster correlation coefficients were calculated from the unconditional model to determine the proportion of variance attributable to each potential level of data analysis in the current sample. Based on the data structure, there was a maximum of 4 levels of analysis: Level 1 = Time, Level 2 = Client, Level 3 = Therapist, Level 4 = Clinical Site. ICCs suggested most of the variance in symptoms over time was a function of client level change (ICCs range from 0.09 to 0.14) with negligible variance attributable to any other level of analysis for most outcome variables (ICCs < 0.03). For some outcomes (TPA, BPM Externalizing), ICCs suggested the potential added benefit of a three-level model accounting for therapist effects (ICCs range = .12 to .13). While there is no formal guideline for an ICC “cut-off,” ICCs of 10% or greater generally suggest the need for multilevel analysis. To avoid oversimplification of the data structure, all analyses were run both as 2-level and 3-level models. Examination of fit indices and parameter estimates suggested that the 3-level models produced comparable findings to that of the 2-level models. To ensure parsimony and facilitate comparison to the prior trials, we present findings of the 2-level models.

**Results**

**Participants and Descriptive Data**

Comparison of clinician and client demographics between the Weisz et al. (2012, 2020) trials and this sample showed several differences. Clinicians in this sample were on average younger, more likely to be from the counseling discipline, and had fewer years of experience. Clinicians...
in this sample were also less ethnically diverse than the Weisz et al., 2012 sample (all ps <.01; see Table 2).

Clients in this sample were on average older and more likely to be female than in both previous trials, and less ethnically diverse than the Weisz et al., 2012 trial (all ps <.05; see Table 3). Clients in this sample also showed higher child-reported baseline severity scores on the BPM and lower caregiver-rated baseline Top Problems Assessment scores than both previous trials. Caregiver reported baseline externalizing problems were higher in the Weisz et al., 2020 trial than this sample. Clients were also more likely to be receiving services for traumatic stress and less likely to be receiving services for conduct problems in this trial (see Table 3). There were no differences in child-reported Top Problems Assessment scores, mean length of treatment, or frequency of treatment.

**Youth Reported Outcomes**

Table 4 and Figure 1 show trajectories and projected 1-year change values for the current and benchmarked comparison samples for youth reported symptoms. Trajectory estimates in this sample indicated that significant reductions in youth-reported symptoms over time occurred for all outcome variables (slopes range = −.55 to −1.77, all ps <.001). Overall, youth reported outcome variables showed comparable to larger effect sizes than those reported in the 2012 and 2020 trials (i.e., steeper trajectories). As seen in Table 4, reductions in youth reported BPM were significantly greater in this sample (ts range = 2.12 to 4.26, ps <.05) relative to those reported in Weisz et al., (2012 and 2020). Reductions in youth reported TPA were comparable to both benchmark samples (all ps >.05). Figure 1 illustrates the 1-year change projections based on estimated slope in the current sample next to those reported in Weisz et al. (2012 and 2020).

**Caregiver Reported Outcomes**

Table 5 and Figure 2 show slope estimates and projected 1-year change values for the current and benchmarked comparison samples for caregiver-reported symptoms. Similar to youth-reported outcomes, trajectories of all caregiver-reported symptom measures showed significant symptom reduction over time (bs range = −.36 to −1.28, all ps <.001). Caregiver reported outcomes were generally comparable to both the 2012 and 2020 effectiveness trials. Table 5 shows the estimated slope values in the current sample and 2012 and 2020 trials. Comparing change in slopes from this sample and the prior trials indicate that caregiver reported TPA and BPM reductions in this sample were not significantly different (ts range = .34 to 1.79, all ps >.05) except for the BPM Internalizing Subscale, which showed a slightly steeper trajectory in this sample compared to the Weisz et al., 2020 trial (t(241) = 2.43, p <.05). Figure 2 illustrates that 1-year change projections based on estimated slopes in the current sample were similar to those reported in Weisz et al. (2012).

**Discussion**

As researchers continue to provide evidence for the use of EBPs in controlled contexts, direct service organizations
have gained significant ground in getting those EBPs into varied service settings. Nevertheless, although there is ample evidence that these treatments should work in these settings, there is limited evidence that they do work, especially in the absence of the quality assurance and administrative procedures that occur naturally in the research context. Do they outperform prior treatment practices when they are implemented with more limited resources? Do they explicitly address characteristics and implementation needs of the community? The fiscal and logistical burden of testing whether interventions outperform usual care through traditional RCTs can be a significant barrier to industrial implementation in varied settings. Yet, given equivocal evidence for the successful adaptation of EBPs from laboratories to clinics (e.g., Southam-Gerow et al., 2010), some method of evaluation is clearly warranted or at least justified, especially when new programs are utilized in an industrial implementation initiative with a more limited level of supports than an effectiveness trial. This practical implementation report demonstrates how by using measurements that align with RCTs, and careful selection from multiple RCTs to maximize sample comparability, the use of benchmarking analyses provides an opportunity to understand how treatments are working in the context for which they were oftentimes designed. This report also provides evidence that successful implementation is feasible in the absence of the externally-supported and labor-intensive strategies characteristic of research trials, when using explicit implementation tools to facilitate proper utilization of the program in context.

The results of this initiative indicated that, when used in a community environment with explicit process management tools supplanting some of the resources and controls of a research trial, the MATCH program may be as effective as demonstrated in published clinical trials. This is important because the 2012 trial demonstrated MATCH’s superiority to both usual care and standard manualized treatment (Weisz et al., 2012). Extrapolating those results to this population, the reader, and even more importantly the service organizations, can somewhat reasonably assume that individuals in this trial experienced a greater reduction in symptoms than those who might have received such alternate treatment. The benchmarking analysis here also provides an enhanced context for interpreting the findings of the Weisz et al., 2020 trial that failed to

### Table 3. Client demographic and baseline clinical characteristics.

|                                | Present Study (n = 166) M (SD) | Weisz et al., 2012 (n = 174) M (SD) | Benchmark analysis | Weisz et al., 2020 (n = 156) M (SD) | Benchmark analysis |
|--------------------------------|--------------------------------|-----------------------------------|--------------------|-----------------------------------|--------------------|
| **Demographics**               |                                |                                   |                    |                                   |                    |
| Age, M (SD)                    | 12.46 (2.64)                   | 10.59 (1.76)                      | t(338) = 7.72**    | 10.50 (2.5)                       | t(320) = 6.83**    |
| Male, n (%)                    | 96, 50%                        | 121, 70%                          | χ²(1, N = 340) = 5.04† | 75, 48%                           | χ²(1, N = 322) = 3.07 |
| Race/Ethnicity                 |                                |                                   |                    |                                   |                    |
| Asian or Pacific Islander      | 2, 1%                          | 7, 4%                             | χ²(5, N = 340) = 19.17** | 0, 0%                             | χ²(4, N = 322) = 22.10** |
| Black / African American       | 16, 10%                        | 16, 9%                            |                    | 8, 5%                             |                    |
| White / Caucasian             | 98, 59%                        | 78, 45%                           |                    | 125, 80%                          |                    |
| Latino/Latina                  | 18, 11%                        | 10, 6%                            |                    | 3, 2%                             |                    |
| Multiethnic/Other              | 32, 19%                        | 59, 34%                           |                    | 20, 13%                           |                    |
| Prefer not to disclose         | 0, 0%                          | 4, 2%                             |                    | 0, 0%                             |                    |
| **Brief Problem Monitor**      |                                |                                   |                    |                                   |                    |
| Child Internalizing            | 4.34 (2.97)                    | 3.44 (2.82)                       | t(338) = 2.87**    | 2.65 (2.36)                       | t(320) = 5.63**    |
| Child Externalizing            | 4.13 (2.60)                    | 3.08 (2.65)                       | t(338) = 3.69**    | 3.56 (2.25)                       | t(320) = 2.10†     |
| Child Total                    | 8.50 (4.55)                    | 6.48 (4.44)                       | t(338) = 4.14**    | 6.21 (3.83)                       | t(320) = 4.87**    |
| Caregiver Internalizing        | 5.14 (3.12)                    | 4.90 (3.23)                       | t(338) = 0.70      | 4.95 (2.48)                       | t(320) = 0.60      |
| Caregiver Externalizing        | 5.26 (3.02)                    | 5.31 (3.16)                       | t(338) = -0.15     | 6.17 (2.97)                       | t(320) = -2.72**   |
| Caregiver Total                | 10.44 (4.73)                   | 10.21 (5.08)                      | t(338) = 0.43      | 11.11 (4.15)                      | t(320) = -1.35     |
| **Top Problem Assessment**     |                                |                                   |                    |                                   |                    |
| Child Top Problem              | 4.99 (2.67)                    | 5.52 (2.81)                       | t(338) = -1.78     | 5.54 (3.13)                       | t(320) = -1.70     |
| Caregiver Top Problem          | 6.31 (2.22)                    | 7.04 (1.86)                       | t(338) = -3.29**   | 7.82 (1.20)                       | t(320) = -7.52**   |
| Mean Treatment Length          | 193.66 (134.61)                | 210.15 (115.75)                   | t(226) = -0.85     | 191.44 (103.41)                   | t(241) = 0.13      |
| Mean Days Between Session      | 10.91 (9.69)                   | 11.96 (4.65)                      | t(285) = -1.10     | 12.92 (6.74)                      | t(241) = -1.64     |
| **Primary Protocol**           |                                |                                   |                    |                                   |                    |
| Anxiety                        | 51, 31%                        | 59, 34%                           |                    | 27, 17%                           | χ²(3, N = 322) = 15.02** |
| Depression                     | 42, 25%                        | 37, 21%                           |                    | 42, 27%                           |                    |
| Trauma                         | 19, 11%                        | 0, 0%                             |                    | 9, 6%                             |                    |
| Conduct                        | 54, 33%                        | 78, 45%                           |                    | 78, 50%                           |                    |

†p < .05, *p < .01, **p < .001.
demonstrate clinically superior results to its usual care control group. Notably, Figure 1 illustrates that the usual care group in that trial had consistently higher caregiver report effects than the usual care group in the 2012 trial and the 2020 usual care group had effects that appear comparable to the MATCH group in the 2012 trial. Thus, interpretation of the observed failure to replicate superior results should address the unusual success of usual care with caregivers in the 2020 trial in addition to the characteristics of the treatment group (e.g., group supervision) potentially related to reduced effectiveness.

Although prior trials of MATCH and this observational study were all conducted in community mental health centers, there is a distinct difference between testing a treatment in a community mental health center and the day-to-day treatment that occurs in community mental health. Whereas both types of trials may occur in the same therapeutic environment, the support available for

### Table 4. Youth reported trajectories of change in client symptoms over time.

| Measure                  | Current Sample (N = 166) | Weisz et al 2012 (MMT sample N = 62) | Weisz et al., 2020 (STEPS sample n = 77) |
|--------------------------|--------------------------|--------------------------------------|------------------------------------------|
|                          | Slope (SE)               | Projected 1-y change Slope (SE)      | Projected 1-y change Benchmark Slope (SE) | Projected 1-y change Benchmark |
| BPM Total Scale          | -1.17 (.10)              | -6.89                                | -4.04                                    | -48 (.10)                      |
| BPM Internalizing Scale  | -61 (.06)                | -3.59                                | -2.28                                    | -24 (.06)                      |
| BPM Externalizing Scale  | -55 (.05)                | -3.23                                | -1.74                                    | -24 (.06)                      |
| Average Top Problems     | -80 (.06)                | -4.75                                | -3.57                                    | -64 (.09)                      |

Note. The slope is the estimate of the change in scale score per log day, and the 1-year change is the change in scale score that would be projected 1 year after the initial assessment.

BPM; Brief Problem Monitor.

†p < .05, *p < .01, **p < .001.

**Figure 1.** Youth reported 1-year change Projections.
Implementation is often different than the unique support provided to a research trial. The impact of differing levels of management support on implementation seems to be present even across research trials. Again, Weisz et al. (2020) noted that when consultation is provided in a group instead of an individual format, it may be difficult to maintain the superiority of MATCH over usual care. In other words, the potency of MATCH consultation may be diminished in a group format. Given that the more general strategies for managing cases are not an explicit part of the MATCH protocol (see Regan et al., 2019), implementation efforts may need to instill and support these formerly implicit and externally supported process management strategies within organizations pursuing long-term implementation.

In contrast to Weisz et al. (2020), the present study incorporated those explicit resources to support learning in the group-based MATCH consultation calls. The process management tools summarized earlier were used to structure provider development and expertise in

Table 5. Caregiver reported trajectories of change in client symptoms over time.

| Measure                  | Current Sample (N = 166) | Weisz et al 2012 (MMT sample N = 62) | Weisz et al., 2020 (STEPs sample n = 77) |
|--------------------------|--------------------------|--------------------------------------|-----------------------------------------|
|                          | Projected Slope (SE)     | Projected Slope (SE)                 | Benchmark Slope (SE)                    |
|                          | Projected 1-y change     | Projected 1-y change                 | Benchmark 1-y change                    |
| BPM Total Scale          | −.89 (.09)               | −.94 (.12)                           | −.75 (.10)                              |
| BPM Internalizing Scale  | −.53 (.06)               | −.50 (.08)                           | −.30 (.05)                              |
| BPM Externalizing Scale  | −.36 (.05)               | −.45 (.07)                           | −.45 (.07)                              |
| Average Top Problems     | −.50 (.05)               | −.65 (.07)                           | −.65 (.07)                              |

Note. The slope is the estimate of the change in scale score per log day, and the 1-year change is the change in scale score that would be projected 1 year after the initial assessment. BPM; Brief Problem Monitor
†p < .05, *p < .01, **p < .001

Figure 2. Caregiver reported 1-year change Projections³.
MATCH (i.e., process guides and managing integrity to the information-rich decisional processes as opposed to adherence to specific procedures in the manual). As noted by Chorppita and Daleiden (2018), the decisional process, or “knowing what to do when”, requires the complex assembling of information and techniques from a variety of sources (e.g., research evidence, individual client evidence, and clinical theory). Expert consultants may be able to support this complex assembling. In fact, Regan et al. (2019) found that providers delivering MATCH ended up following the consultant’s recommendation more than twice as often as following the manual when MATCH consultants recommended a strategic deviation from the standard MATCH protocol. The results of Weisz et al. (2020) suggest that this nuance–i.e., that following the manual is not necessarily following the program as designed–could easily get lost in group-based consultation. Although empirical evaluation is needed to confirm this hypothesis, results from this trial suggest that this coordination process of “knowing what to do when” can be retained in a group format through the utilization of explicit process management tools in consultation. Thus, the decisional process may require either individual consultation by an expert or the additional implementation of explicit, decisional support tools that help inform the actions of a therapist in the absence of a consultant, the latter of which is clearly more cost effective.

Limitations

Limitations to this work parallel those associated with benchmarking methodology. As with any observational study, there may be alternate explanations for our results. Notably, the sample of clients and clinicians had significantly different demographic characteristics than the 2012 and 2020 trials. It may be the case that the client or clinician demographics had more to do with the improvement in our sample than the MATCH program did. Specifically, the youth in this sample were older than the previous trials because older adolescents were permitted to enroll in MATCH services in this study. This has two potential implications. First, the higher age range for inclusion in this study sample resulted in a greater number of youth respondents who were able to complete surveys. Perhaps, as a result, youth in this sample reported higher baseline symptoms than their RCT counterparts, especially regarding externalizing problems. Thus, it is possible the greater youth-reported improvements in this sample could be explained by the fact that youth had more room for change over time, rather than MATCH being more effective in this sample compared with the two RCTs. Second, aspects of effective treatment for adolescent disruptive behavior are somewhat different than those for children (i.e., commands and time out are used with less frequency whereas communication skills, problem-solving, and family therapy practices are used more frequently in the EBP literature). Thus, given the modular nature of the intervention, youth in this study might not have received the same intervention, strictly speaking, as the participants in the two RCTs we used for benchmarking.

Additionally, both the comparison RCT (Weisz et al. 2012 and 2020) samples were more heavily represented by youth with externalizing problems and younger youth compared to the current sample. Given that there is a historic reliance on caregiver report for assessing externalizing problems, this may mean that having an older sample of youth in this study resulted in a slower trajectory of symptoms improvement on the Caregiver TPA. Thus, in the analysis of caregiver-reported measures, the difference between the prior trials’ results and those in this study may need to be interpreted in the context of the several factors (e.g., youth age, primary problem area, and informant); benchmarking analysis alone cannot address these nuances. Furthermore, clinicians were free to choose which youth they enrolled in MATCH. Clinicians may have chosen cases that were more impaired than the prior trial and thus in need of specialized EBP services; conversely, clinicians could have elected less complicated and impaired individuals to receive MATCH.

Furthermore, to compare across studies, we used projected change estimates from this implementation study sample given that 1-year follow-up data were not collected. While it is possible that this methodological approach may have inflated results, the approach is consistent with that of Weisz et al. (2012). Furthermore, Chorpita et al. (2013) found that the difference between MATCH and usual care conditions did not converge in their long-term follow-up study. MATCH’s advantage over usual care was most pronounced in the first six months of treatment, and again at the 1-year observation period, after which few additional gains were observed in any condition. Rates of change were near zero for all conditions for the second year of observation, indicating that gains were maintained.

Conclusions and Future Directions

Although the results obtained in this initiative were obtained without those rigorous research controls in place, they suggest that there is still further work to be done to better understand the performance of MATCH in diverse community settings. Benchmarking in this manner allows these results to be considered as part of a larger literature informing about what works for whom and when. For example, it may be advisable for EBP registries to provide guidance on potential “gold standard RCT comparators” and their measurement models so that future agencies can conduct comparable benchmarking analyses such as those reported here. Results from trials such as this one can be especially helpful to those focused on implementation because the success or failure of a particular program in a particular setting can inform how researchers, policymakers, and service providers approach EBP selection.
As noted earlier, the characteristics of this industrial implementation allowed clinicians to select who received the treatment. This self-selection behavior associated with implementing a new program within an organization is an area for future attention. It may be particularly helpful for future work to incorporate qualitative research to elucidate the clinician intent and experience with the protocol, potentially addressing these issues of sample composition. Such investigations might also continue utilizing qualitative methods to understand and adapt implementation techniques to the characteristics of community settings at the agency level as well.

Structured implementation supports other than the clinical intervention are clearly important for services systems and need to be considered. Through this benchmarking comparison, it seems as though the structured implementation supports developed for this investigation were feasible and appear sufficient to support comparable outcomes in a service setting to those in a clinical trial. Furthermore, benchmarking should continue to be considered as a valuable tool in understanding how interventions fare when transported into similar contexts with different types of implementation. As implementation specialists consider the characteristics of the community settings they serve – working with real clients and real therapists in real organizations with real budgets – developing support tools and benchmarking their results would likely be welcome by agency directors and senior leaders in their efforts to maintain positive outcomes with fewer resources.

Declaration of conflicting interests
The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Drs. Chorpita and Daleiden are partners in PracticeWise, LLC, which publishes the MATCH-ADTC program referred to in this study. Drs. Stanick and Chiu were consultants for PracticeWise, LLC at the time of the study.

Funding
This work was supported by grants from the Peter and Elizabeth C. Tower Foundation, The Blue Cross and Blue Shield of Massachusetts Foundation, and The John D. and Katherine T. MacArthur Foundation.

ORCID iD
Cameo F. Stanick https://orcid.org/0000-0002-6076-3726

Notes
1. When considering the length of training, readers are encouraged to recognize that the 5 days of training (40 hours) encompassed approximately 4 h on implementation training and 36 h on a combined 4 treatment protocols. Thus, approximately 9 h each were spent on training providers to treat anxiety, depression, disruptive behavior, and traumatic stress.

2. It is important to note that the Weiss et al. (2012, 2020) RCTs upon which these data are benchmarked used an earlier, proprietary version of this measure called the Brief Problem Checklist (BPC). The BPC includes the same items from the Internalizing and Externalizing Subscales but omits the Attention Subscale. For simplicity, we reference the BPM throughout even though the Weiss et al. (2012, 2020) manuscripts utilized the shorter BPM.

3. Findings for the 3-level models are available upon request.
4. One-year change projections were also available from another MATCH trial (Chorpita et al., 2017) that used a different analytical model and are included for reference.

References
Achenbach, T. M., McConaughy, S. H., Ivanova, M. Y., & Rescorla, L. A. (2011). Manual for the ASEBA brief problem monitor (BPM). ASEBA, 1–33.
Chorpita, B. F., & Daleiden, E. L. (2018). Coordinated strategic action: Aspiring to wisdom in mental health service systems. Clinical Psychology: Science and Practice, 25(4), e12264.
Chorpita, B. F., Daleiden, E. L., Park, A. L., Ward, A. M., Levy, M. C., Cromley, T., Chiu, A. W., Letamendi, A. M., Tsai, H. H., & & Krull, J. L. (2017). Child STEPs in California: A randomized effectiveness trial comparing modular treatment with community implemented treatment for youth with anxiety, depression, conduct problems, or traumatic stress. Journal of Consulting and Clinical Psychology, 85(1), 13–25. https://doi.org/10.1037/ccp0000133
Chorpita, B. F., & Weisz, J. R. (2009). Modular approach to therapy for children with anxiety, depression, trauma, or conduct problems (MATCH-ADTC). PracticeWise, LLC.
Chorpita, B. F., Weisz, J. R., Daleiden, E. L., Schoenwald, S. K., Palinkas, L. A., & Miranda, J., Austin, A. A., Borntrager, C. F., Ward, A., Wells, K. C., Gibbons, R. D., & Research Network on Youth Mental Health. (2013). Long-term outcomes for the child STEPs randomized effectiveness trial: A comparison of modular and standard treatment designs with usual care. Journal of Consulting and Clinical Psychology, 81(6), 999–1009. https://doi.org/10.1037/a0034200
Daleiden, E., & Chorpita, B. F. (2005). From data to wisdom: quality improvement strategies supporting large-scale implementation of evidence based services. Child and Adolescent Psychiatric Clinics of North America, 14(2), 329–349. https://doi.org/10.1016/j.chc.2004.11.002
Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. Implementation Science, 4(1), 1–15. https://doi.org/10.1186/1748-5908-4-50
De Los Reyes, A., & Ohannessian, C. M. (2016). Introduction to the special issue: Discrepancies in adolescent–parent perceptions of the family and adolescent adjustment. Journal of Youth and Adolescence, 45(10), 1957–1972.
Hewitt, A., Larson, S., Edelstein, S., Seavey, D., Hoge, M., & Morris, J. (2008). A synthesis of direct service workforce demographics and challenges across intellectual/developmental disabilities, aging, physical disabilities, and behavioral health. University of Minnesota, Institute on Community Integration, Research and Training Center on Community Living.
McFall, R. M. (1996, July 1). Consumer satisfaction as a way of evaluating psychotherapy: ecological validity and all that
versus the good old randomized trial (panel discussion). 6th Annual Convention of the American Association of Applied and Preventative Psychology, San Francisco.

Modular Approach to Therapy for Children with Anxiety, Depression, Trauma, or Conduct Problems (MATCH-ADTC). (n.d.). In SAMHSA’s National Registry of Evidence-based Programs and Practices online.

National Implementation Research Network. (2017). Module 1: an overview of active implementation frameworks. Retrieved from http://implementation.fpg.unc.edu/module-1/implementation-drivers

Piper, B. J., Gray, H. M., Raber, J., & Birkett, M. A. (2014). Reliability and validity of brief problem monitor, an abbreviated form of the child behavior checklist. *Psychiatry and Clinical Neurosciences, 68*(10), 759–767. https://doi.org/10.1111/pcn.12188

Presseau, J., Mc Cleary, N., Lorencatto, F., Patey, A. M., Grimshaw, J. M., & Francis, J. J. (2019). Action, actor, context, target, time (AACTT): A framework for specifying behaviour. *Implementation Science, 14*(1), 1–13. https://doi.org/10.1186/s13012-019-0951-x

Regan, J., Park, A. L., & Chorpita, B. F. (2019). Choices in treatment integrity: considering the protocol and consultant recommendations in child and adolescent therapy. *Journal of Clinical Child & Adolescent Psychology, 48*(sup1), S79–S89. https://doi.org/10.1080/15374416.2016.1212360

Rudd, B. N., Davis, M., & Beidas, R. S. (2020). Integrating implementation science in clinical research to maximize public health impact: A call for the reporting and alignment of implementation strategy use with implementation outcomes in clinical research. *Implementation Science, 15*(1), 1–11. https://doi.org/10.1186/s13012-019-0962-7

Southam-Gerow, M. A., Weisz, J. R., Chu, B. C., McLeod, B. D., Gordis, E. B., & Connor-Smith, J. K. (2010). Does cognitive behavioral therapy for youth anxiety outperform usual care in community clinics? An initial effectiveness test. *Journal of the American Academy of Child and Adolescent Psychiatry, 49*(10), 1043–1052. https://doi.org/10.1016/j.jaac.2010.06.009

Southam-Gerow, M. A., Weisz, J. R., & Kendall, P. C. (2003). Youth with anxiety disorders in research and service clinics: examining client differences and similarities. *Journal Of Clinical Child And Adolescent Psychology, 32*(3), 375–385. https://doi.org/10.1207/S15374442JCCP3203_06

Stewart, R. E., Adams, D. R., Mandell, D. S., Hadley, T. R., Evans, A. C., & Rubin, R., Erney, J., Neimark, G., Hurford, M. O., & Beidas, R. S. (2016). The perfect storm: Collision of the business of mental health and the implementation of evidence-based practices. *Psychiatric Services, 67*(2), 159–161. https://doi.org/10.1176/appi.ps.201500392

Wade, W. A., Treat, T. A., & Stuart, G. L. (1998). Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. *Journal of Consulting and Clinical Psychology, 66*(2), 231. https://doi.org/10.1037/0022-006X.66.2.231

Weersing, V. R., & Hamilton, J. (2005). Benchmarking the effectiveness of psychotherapy: Program evaluation as a component of evidence-based practice. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*(10), 1058–1067. https://doi.org/10.1097/01.chi.0000172682.71384.80

Weisz, J. R., Bearman, S. K., Ugueso, A. M., Herren, J. A., Evans, S. C., & Cheron, D. M., Alleyne, A. R., Weissman, A. S., Tweed, J. L., Pollack, A. A., Langer, D. A., Southam-Gerow, M. A., Wells, K. C., & Jensen-Doss, A. (2020). Testing robustness of child STEPs effects with children and adolescents: A randomized controlled effectiveness trial. *Journal of Clinical Child & Adolescent Psychology, 49*(6), 883–896.

Weisz, J. R., Chorpita, B. F., Frye, A., Ng, M. Y., Lau, N., & Bearman, S. K., Ugueso, A. M., Langer, D. A., Hoagwood, K. E., & The Research Network on Youth Mental Health. (2011). Youth top problems: using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy. *Journal of Consulting and Clinical Psychology, 79*(3), 369. https://doi.org/10.1037/a0023307

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., & Bearman, S. K., Daleiden, E. L., Ugueso, A. M., Ho, A., Martin, J., Gray, J., Alleyne, A., Langer, D. A., Southam-Gerow, M. A., Gibbons, R. D., & the Research Network on Youth Mental Health. (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth: A randomized effectiveness trial. *Archives of General Psychiatry, 69*(3), 274–282. https://doi.org/10.1001/archgenpsychiatry.2011.147

Weisz, J. R., Kuppens, S., Eckstein, D., Ugueso, A. M., Hawley, K. M., & Jensen-Doss, A. (2013). Performance of evidence-based youth psychotherapies compared with usual clinical care: A multilevel meta-analysis. *JAMA Psychiatry, 70*(7), 750–761.