AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Consulting and Clinical Psychology

ISSN: 0022-006X

2020, Vol. 88, No. 12, 1065–1078
http://dx.doi.org/10.1037/ccp0000536

Clinician Training, Then What? Randomized Clinical Trial of Child STEPs Psychotherapy Using Lower-Cost Implementation Supports With Versus Without Expert Consultation

John R. Weisz
Harvard University

Kristel Thomassin
University of Guelph

Jacqueline Hersh
Appalachian State University

Lauren C. Santucci
Harvard University

Heather A. MacPherson
Brown University

Gabriela M. Rodriguez
Indiana University School of Medicine

Sarah Kate Bearman
University of Texas at Austin

Jason M. Lang and Jeffrey J. Vanderploeg
Child Health and Human Development Institute, Farmington, Connecticut

Timothy M. Marshall
Connecticut State Department of Children and Families, Hartford, Connecticut

Jack J. Lu
Child Health and Human Development Institute, Farmington, Connecticut

Amanda Jensen-Doss
University of Miami

Spencer C. Evans
Harvard University

Objective: Implementation of evidence-based treatments in funded trials is often supported by expert case consultation for clinicians; this may be financially and logistically difficult in clinical practice. Might less costly implementation support produce acceptable treatment fidelity and clinical outcomes? ***Method:*** To find out, we trained 42 community clinicians from four community clinics in Modular Approach to Therapy for Children (MATCH), then randomly assigned them to receive multiple lower-cost implementation supports (LC) or expert MATCH consultation plus lower-cost supports (CLC). Clinically referred youths ($N = 200$; ages 7–15 years, $M = 10.73$; 53.5% male; 32.5% White, 27.5% Black, 24.0% Latinx, 1.0% Asian, 13.5% multiracial, 1.5% other) were randomly assigned to LC ($n = 101$) or CLC ($n = 99$) clinicians, and groups were compared on MATCH adherence and competence, as well as on

John R. Weisz, Department of Psychology, Harvard University; Kristel Thomassin, Department of Psychology, University of Guelph; Jacqueline Hersh, Department of Psychology, Appalachian State University; Lauren C. Santucci, Department of Psychology, Harvard University; Heather A. MacPherson, Department of Psychiatry and Human Behavior, Warren Alpert Medical School, Brown University; Gabriela M. Rodriguez, Department of Psychiatry, Indiana University School of Medicine; Sarah Kate Bearman, Department of Educational Psychology, University of Texas at Austin; Jason M. Lang and Jeffrey J. Vanderploeg, Child Health and Human Development Institute, Farmington, Connecticut; Timothy M. Marshall, Connecticut State Department of Children and Families, Hartford, Connecticut; Jack J. Lu, Child Health and Human Development Institute; Amanda Jensen-Doss, Department of Psychology, University of Miami; Spencer C. Evans, Department of Psychology, Harvard University.

The study was supported by funding to John R. Weisz from the Child Health and Human Development Institute and the Norlien Foundation. We are grateful to the youths, caregivers, and clinicians who participated in this study, as well as to the clinic administrators and state officials whose leadership facilitated and supported the project. Our special thanks to Kyle Barrette, Chris Bory, Ana Carvalho, Grace Cho, Caroline Cooke, Katherine Corteselli, Tiffany Franceschetti, Linda Gallion, Patricia Garibaldi, Megan Howard, Natalie Klapholz, Nicole Long, Sarah Oravecz, Lea Petrovic, Dana Tagavi, Ana Ugueto, and Kelsey Yonce for their many contributions to the study.

John R. Weisz is a coauthor of the treatment manual used in this study (i.e., *Modular Approach to Therapy for Children With Anxiety, Depression, Trauma, or Conduct Problems*) and receives royalties for its sales. The other authors report no conflicts of interest related to this study.

Correspondence concerning this article should be addressed to John R. Weisz, Department of Psychology, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, MA 02183. E-mail: john_weisz@harvard.edu

multiple clinical outcomes using standardized measures (e.g., Child Behavior Checklist, Youth Self-Report) and idiographic problem ratings (Top Problems Assessment). ***Results:*** Coding of therapy sessions revealed substantial therapist adherence to MATCH in both conditions, with significantly stronger adherence in CLC; however, LC and CLC did not differ significantly in MATCH competence. Trajectories of change on all outcome measures were steep, positive, and highly similar for LC and CLC youths, with no significant differences; a supplemental analysis of posttreatment outcomes also showed similar LC and CLC posttreatment scores, with most LC–CLC differences nonsignificant. ***Conclusions:*** The findings suggest that effective implementation of a complex intervention in clinical practice may be supported by procedures that are less costly and logistically challenging than expert consultation.

What is the public health significance of this article?
Expert consultation, sometimes used to support implementation of evidence-based therapies, may be difficult to arrange and fund in some practice settings. Our findings suggest that effective implementation in clinical practice may be supported by a combination of lower-cost procedures in the absence of expert consultation.

Keywords: children, adolescents, youths, implementation, psychotherapy

A recurring challenge for clinical and implementation science is determining what therapist support is needed after an initial training in an evidence-based treatment (EBT) to produce effective implementation by clinicians in practice settings. The question is important for dissemination efforts because most forms of posttraining implementation support come with a price tag, and the ability of provider organizations to bring EBTs into everyday clinical care will depend, inevitably, on their ability to manage the concomitant costs (Lang & Connell, 2017; Roundfield & Lang, 2017; Schoenwald, Kelleher, Weisz, & the Research Network on Youth Mental Health, 2008). In fact, implementation costs have been identified in some research as the most significant factor in implementing and sustaining EBTs (Aarons, Wells, Zagursky, Fettes, & Palinkas, 2009; Pegg, Walsh, Becker-Haimes, Ramirez, & Jensen-Doss, 2019). Our capacity to bring EBTs into everyday mental health care may depend in part on our ability to limit potentially costly implementation support to the forms and levels that are necessary and sufficient.

Posttraining implementation support may take a variety of forms, some more costly than others (Powell et al., 2015; Proctor et al., 2009). At the lower end of the cost spectrum, videorecordings of the formal training may be saved by the service organization and made available for review. Handouts and other learning aids used during the training may also be provided for clinicians. For some treatments, video tutorials are available, with treatment sessions illustrated by actors, to convey the ideal content, structure, and flow of each session. In addition, some treatments are supported by websites that provide free access to therapist tips, printable worksheets, and flyers for parents that describe treatment components being implemented with their children. For treatments that are guided by measurement-based care (MBC), clinicians may also be given access to routine feedback on each client's response to treatment during episodes of care (Fortney et al., 2017; Scott & Lewis, 2015).

At the costlier end of the spectrum is case consultation by experts in the treatment being implemented. For many randomized trials, this form of support is standard practice—for example, at 1 hr per week for each of the study clinicians delivering the treatment (e.g., Weisz et al., 2012). Such expert consultation, common in funded trials, may be more difficult to arrange in everyday clinical practice. One challenge is the often limited pool of expertise. For some EBTs, particularly newer ones, there may be relatively few experts, with limited flexible time outside of their current employment. A second challenge is cost, in several forms. One of these is opportunity cost: The time staff clinicians devote to receiving consultation is time away from their primary role, patient care, and if many clinicians are involved, service delays may result and waitlists may grow. Other costs associated with expert consultation are financial, and some research ranks consultation as the costliest form of posttraining support (see, e.g., Lang & Connell, 2017; Roundfield & Lang, 2017). The direct cost of hiring expert consultants can be substantial, and that cost grows as the number of clinicians increases. In addition, consultation time generates indirect costs; every clinician hour devoted to consultation is an hour of lost reimbursement income for services—income upon which most service organizations depend for survival (Schoenwald et al., 2008). Lost reimbursement income has been ranked second only to direct consultation cost among posttraining agency implementation expenses (Lang & Connell, 2017). Individual clinicians receiving consultation also lose time that might otherwise be devoted to client care and thus lose income if they are paid on a fee-for-service basis by their employer. Similarly, private practitioners must weigh the cost of expert consultation in both forms—that is, the consulting payment to the expert and the client fees they would forego by taking time off work for consultation. For all these reasons, expert consultation may be a difficult form of implementation support to arrange in everyday clinical practice.

Research on the contributions of posttraining consultation has generally shown beneficial effects on clinician adherence and competence in delivering treatment protocols (e.g., Beidas, Edmunds, Marcus, & Kendall, 2012; Schwalbe, Oh, & Zweben, 2014; Sholomskas, Syracuse, Rounsaville, Carroll, 2005; Webster-Stratton, Reid, & Marsenich, 2014) with evidence more limited and mixed regarding improved clinical outcomes (e.g., Nadeem, Gleacher, & Beidas, 2013; Schoenwald, Sheidow, & Chapman, 2009; Smith-Boydston, Holtzman, & Roberts, 2014). The contribution of expert consultation may depend to some extent on the type of intervention being implemented. At one extreme, interven-

tions that are brief and highly scripted (see, e.g., single-session interventions reviewed in Schleider & Weisz, 2017) may require relatively little support. At the other extreme, multisession interventions that are highly flexible, modular in design, and transdiagnostic in their focus may require substantial guidance and support for therapists; this, at least, has been the assumption guiding implementation of some such treatments.

One example of such a treatment approach is Child STEPs (sometimes referred to as "STEPs"), which consists of a transdiagnostic manual called Modular Approach to Treatment of Children (MATCH; Chorpita & Weisz, 2009) and a system for MBC—that is, for monitoring youth treatment response and providing weekly feedback to the treating clinicians (e.g., Chorpita et al., 2017; Weisz & Chorpita, 2011; Weisz et al., 2012). MATCH includes 33 modules (brief summaries of empirically supported treatment procedures) organized within four protocols (modules for treatment of anxiety, depression, trauma, and conduct problems), with flowcharts to guide selection of initial treatment focus and other flowcharts guiding navigation within each of the protocols. A therapist using Child STEPs needs to make judgments throughout an episode of care regarding which youth problems should be addressed, with which protocols and modules, and in which order, and these judgments need to be informed—and modified throughout treatment—by the clinician's monitoring of weekly feedback from the youth and caregiver regarding the youth's treatment response. In addition to this decision making, the therapist needs to implement the treatment procedures in each module competently, adjusting procedures in relation to the youth's characteristics, emotions, and behaviors. As this description implies, fidelity of MATCH implementation is closely linked conceptually to clinical outcomes in the STEPs model (Chorpita & Weisz, 2009; Weisz & Chorpita, 2011; Weisz, Krumholz, Santucci, Thomassin, & Ng, 2015). The MATCH modules were derived from components of evidence-based treatments and are thus expected to be effective only to the extent that they are implemented faithfully and competently, consistent with procedures in prior studies that demonstrated effectiveness.

Faithfulness of implementation is gauged by measures of adherence, the extent to which intervention follows the prescribed content and method. Competence of implementation is gauged by measures of the extent to which an intervention is carried out skillfully in relation to characteristics of the individual youth (e.g., capacity to understand). Expert consultation has been provided to help clinicians maintain adherence and competence in delivering MATCH. For clinicians who are learning to use MATCH, success in implementing the procedures and requirements has been assumed to require consultation from a MATCH expert. Thus, all published MATCH randomized controlled trials (RCTs) to date have included a provision that therapists implementing MATCH will have—in addition to an initial training program—weekly case consultation throughout the treatment phase of the study, to help them apply the procedures appropriately (Chorpita et al., 2017; Merry et al., in press; Weisz et al., 2012, 2019).

What cannot be determined from the prior studies—of MATCH and of other modular, flexible treatments that use MBC—is the counterfactual. That is, previous studies of MATCH, all using case consultation, cannot tell us whether outcomes would have been weakened if less costly forms of implementation support had been used. What is needed to address this question is an RCT in which newly trained clinicians are randomly assigned to receive available implementation supports with versus without expert consultation, as they implement the treatment. In this study, we carried out such a trial. We enrolled clinicians from four large community mental health clinics that served children and adolescents (herein "youths"). These clinicians received MATCH training, followed by random assignment to (a) a lower-cost (LC) implementation support group, which was given access to videorecordings of the training sessions, handouts and learning aids from the training, MATCH video tutorials, and free access to online MATCH support and to an online MBC system, or (b) a consultation + lower-cost supports (CLC) group, which received the same resources as the LC group plus weekly case consultation from MATCH experts. Youths referred for treatment in the four clinics were randomly assigned to receive treatment from clinicians in one of the two study conditions. Data analyses examined whether the two study conditions differed in (a) the degree of MATCH adherence and competence shown by clinicians and (b) youth clinical outcomes. The study was thus designed to test whether adding MATCH expert consultation to a combination of lower-cost implementation supports would result in more adherent and competent MATCH delivery by clinicians and/or better youth outcomes than the use of lower-cost supports alone. We reasoned that addressing this question might be helpful to providers and provider organizations in their efforts to balance effective implementation with effective fiscal management.

Method

Informed consent was obtained from clinicians and caregivers, as well as assent from youths, prior to study enrollment. All procedures were approved by two institutional review boards—one affiliated with the researchers' university and the other affiliated with the state's Department of Children and Families.

Youth Participants and Procedures

The study sample included 200 youths referred to one of four partner community mental health clinics. Upon first contact with the clinics, families were provided with basic information about the study, and if they expressed interest, they were referred to the research team for a follow-up phone call. The research team called the family to provide further information about the study and screen them for eligibility. Consent and assent were acquired over the phone followed by the initial baseline assessment with youth and caregiver. After confirming study eligibility, youths were then randomly assigned to one of the two treatment conditions. Written consent and assent were then obtained by mail. As shown in the main CONSORT flow diagram (see Figure 1), a total of 584 families (111–176 per clinic) were referred by the four partner clinics. Of these, 444 completed phone screens and 264 completed full baseline assessments, yielding 210 cases (51–55 at each clinic) who met all eligibility requirements and were randomly allocated across the two conditions within clinic. Five cases in each condition failed to attend the first treatment session, leaving a final sample of 200 youths in the LC (n = 101) and CLC (n = 99) conditions who received the intervention, completed assessments, and were included in outcomes analyses (by clinic, ns = 25–28 and 26–27, respectively).

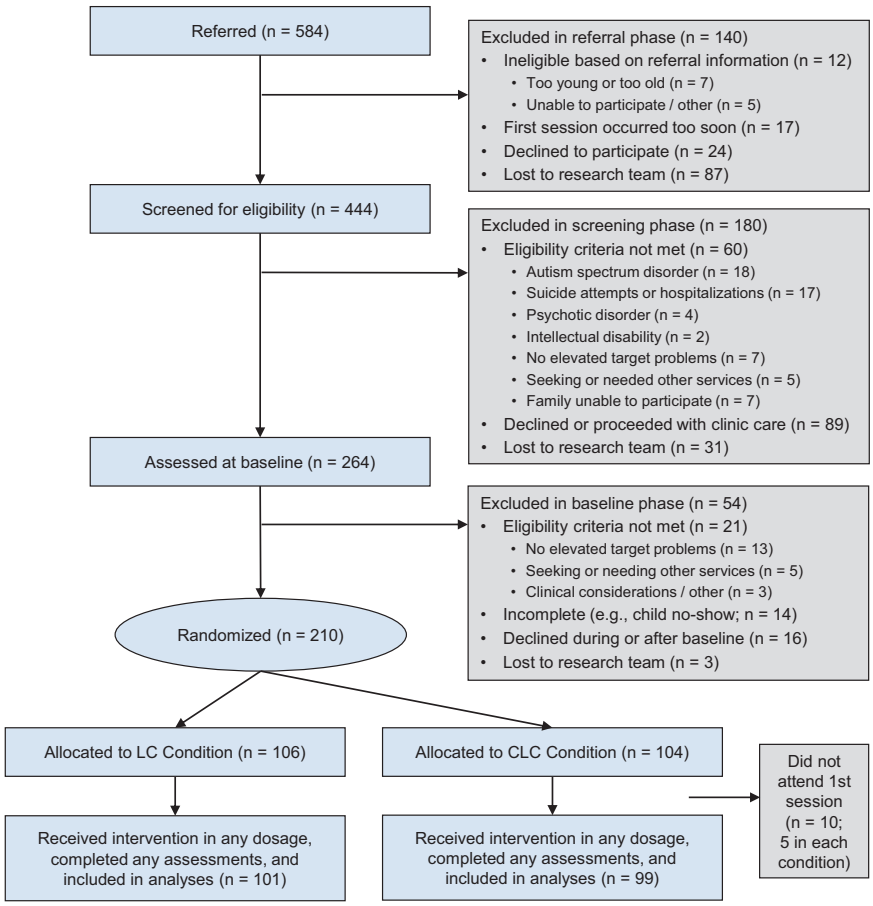

Figure 1. CONSORT flow diagram for participating families. LC = lower-cost supports only; CLC = consultation plus lower-cost supports. See the online article for the color version of this figure.

Inclusion criteria were youths (a) ages 6–15 and their caregivers, (b) seeking services at community mental health clinics, and (c) with primary problem or disorder related to anxiety, depression, traumatic stress, or conduct problems or any combination of the four. Exclusion criteria included (a) referral issues or primary presenting problems falling outside the scope of MATCH—that is, attention problems/hyperactivity (but youths diagnosed with attention-deficit/hyperactivity disorder were included if referred for a MATCH-relevant problem), (b) primary caregiver not willing to be involved in treatment and complete research assessments, and (c) presence of psychotic spectrum disorders, autism spectrum disorders, eating disorders, intellectual disability, or past-year hospitalization for suicidal thoughts or behaviors.

The final sample included 200 youths aged 7–15 years (M_{age} = 10.73, SD = 2.42; 53.5% male, 46.5% female) at baseline, within the developmental range appropriate for MATCH and closely resembling the age range of previous STEPs trials. Families were ethnically and racially diverse, with 32.5% identifying as White/Caucasian, 27.5% as Black/African American, 24.0% as Hispanic/Latinx, 1.0% as Asian, 13.5% as multiracial, and 1.5% as other. About one third (34.5%) reported an annual family income of less than $20,000, while 29.5% reported $20,000–$39,000, 13.0% reported $40,000–$59,000, 6.5% reported $60,000–$79,000, 7.0% reported $80,000–$99,000, and 5.0% reported $100,000 or more; 4.5% did not respond. Some 32.5% of the sample received some form of psychiatric medications during their study treatment episode. The sample included only youths referred through normal community pathways (For another application of these data, see the Appendix).

To ensure that random assignment was indeed random, we assessed for differences between conditions at baseline. There were no significant differences at baseline between the two treatment conditions on any of the aforementioned demographic or treatment variables (all ps $> .09$). Regarding clinical outcomes measured at baseline, there were no differences on any of the caregiver-reported measures (all ps $> .08$) or the youth-reported measures (ps $> .08$), with one exception: internalizing scores on the youth-reported Brief Problem Monitor (BPM) were slightly higher ($p = .017$) for the LC condition ($M = 3.86$, $SD = 3.37$) than the CLC condition ($M = 2.78$, $SD = 2.92$). This one significant difference out of 34 baseline variables fell within chance expectancy, and the measure is a small subset of a more comprehensive and reliable measure (Youth Self-Report [YSR]—below) that showed no significant condition differences at baseline. Thus, the pattern suggests that the groups

were indeed randomly assigned and generally equivalent at baseline.

The mean number of MATCH sessions was 10.78 (SD = 9.28), nonsignificantly fewer in LC (M = 9.63, SD = 7.65) than CLC (M = 11.95; SD = 10.61). Average duration from baseline to last session was 166.8 days (SD = 135.2; Mdn = 132), with no significant difference (p = .215) between the two conditions (M_{LC} = 154.9, SD = 122.4; M_{CLC} = 178.8, SD = 146.7).

Clinics and Clinicians

Four large community mental health clinics in the U.S. Northeast participated. The clinics were freestanding organizations serving urban, suburban, and rural populations, with services paid by third-party reimbursement, primarily Medicaid. Agency heads at these clinics invited clinicians to participate in the study. In total, 49 clinicians began training, 47 were randomized, and 42 treated at least one study case (see Figure 2, clinician CONSORT, and see CONSORT footnote regarding reallocation of one therapist). The clinics provided an array of mental health services to youths and families referred from a variety of sources, including schools, Department of Children and Families, and self-referrals. All therapists were asked to use MATCH in the treatment they provided but not discouraged from employing other treatment procedures they believed would be helpful. The chief difference between conditions was the consultation component as described below.

Therapists averaged 35.39 (SD = 10.34) years of age, 5.54 (SD = 5.60) years of postundergraduate training, and 5.86 (SD = 6.67) years of professional experience. Of the 42, 37 were female; 29 were White/Caucasian, 2 Black/African American, 7 Hispanic/Latinx, 0 Asian, 1 multiracial, 2 "other," and 1 did not respond. Some 21 of the clinicians were social workers, 8 were marriage and family therapists, 6 were counselors, 5 were psychologists, 1 was a behavioral health clinician, and 1 was a school psychologist; 5 had doctorates, 35 had master's degrees or equivalent (e.g., MA, MSW, MS, LCSW, MFCC/MFT), 1 was currently in a doctoral program, and 1 did not respond. When asked whether they were state licensed, 15 responded yes, 20 responded no, and 2 did not respond.

To assess therapists' prior experience with EBTs, relevant to interpreting our findings, we used the Experience in Evidence-Based Practice Survey (EEBPS—see Measures and Figure 3), on which clinicians rate their prior experience with specific EBTs and with EBTs overall. The last EEBPS item asked. "Overall, how experienced would you say you are in delivering evidence-based therapies?" Ratings, on a scale from 0 (*not at all experienced*) to 5 (*very experienced*), averaged 2.93 (SD = 1.21, Mdn = 3), with 30.9% rating 4 or 5 and another 31.0% selecting 3. No clinician endorsed "not at all experienced." On all clinician background variables and EEBPS items, t tests and χ^2 tests revealed no significant LC versus CLS differences (all ps $>$ .19).

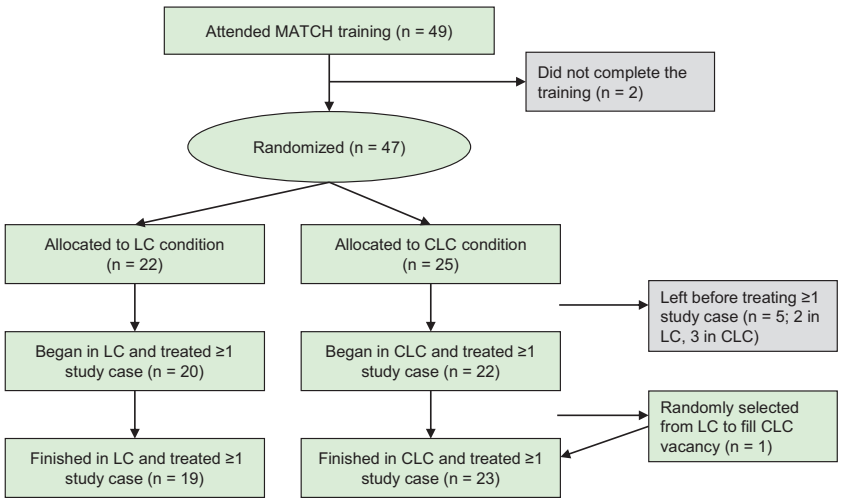

Figure 2. CONSORT flow diagram for participating therapists. Data regarding reasons for leaving were not always available, and data that are available can only be reported in aggregate. Leaving one's agency was the most common reason, occurring in a majority of cases, followed by various personal and administrative reasons (e.g., maternity leave, career changes, retirement, budgetary issues). To our knowledge, no therapist left the study for reasons that were not primarily due to external factors such as these. At one point, this therapist turnover caused a problematic shortage of CLC therapists in one agency. It became necessary to randomly select one of the agency's LC therapists for reallocation to the CLC condition. Thus, as shown above, the original allocation of therapists was 20 in LC and 22 in CLC, but the allocation in the end and for most of the study was 19 in LC and 23 in CLC. For patient-level analyses, this therapist's study case data (one in LC; seven in CLC) were modeled separately in the two conditions. For clinician-level analyses, this therapist's data were included in the CLC condition given that the large majority of this person's time and study caseload were in the CLC condition. LC = lower-cost supports only; CLC = consultation plus lower-cost supports. See the online article for the color version of this figure.

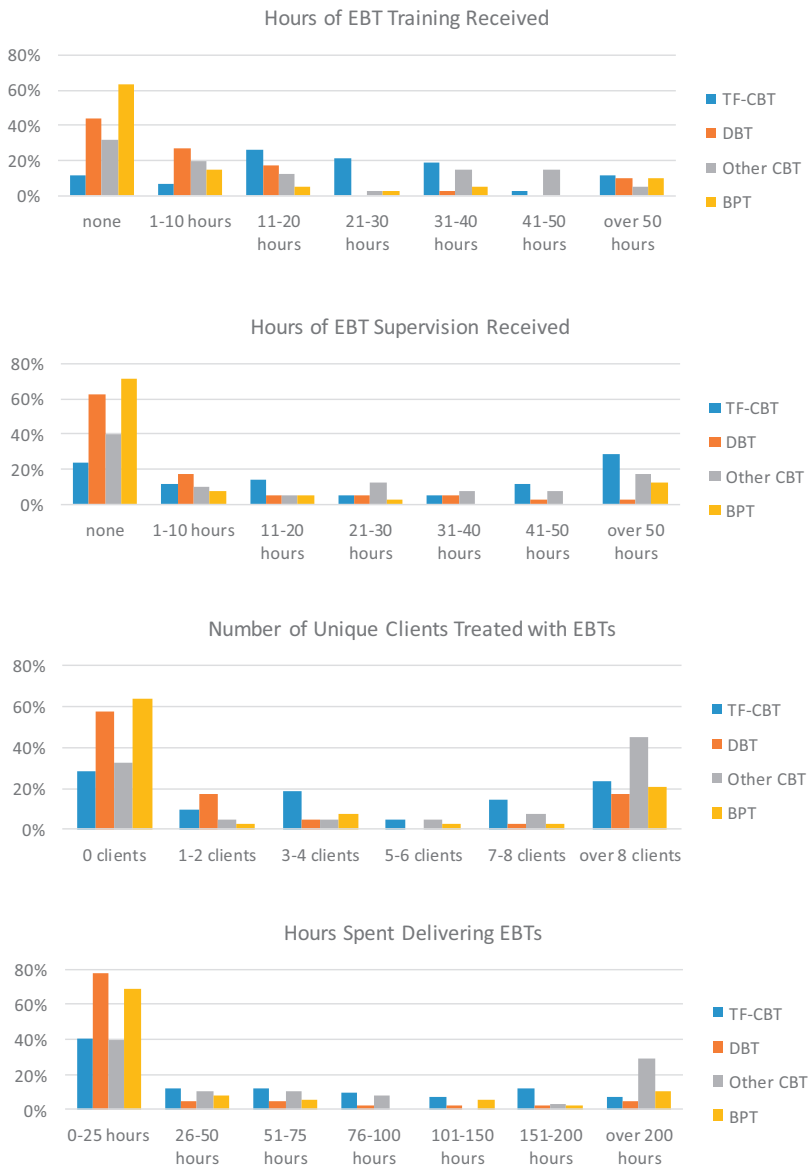

Figure 3. Therapists' self-reported previous training, supervision, and delivery of EBTs. EBT = evidence-based treatment; TF = trauma focused; CBT = cognitive–behavior therapy; DBT = dialectical behavior therapy; BPT = behavioral parent training.

Experimental Design and Study Conditions

Clinicians and youths were randomly assigned within the clinic to one of the two conditions: LC or CLC. Elements common and unique to each condition are described below.

Included in both conditions—training plus lower-cost supports. Clinicians in both conditions (LC and CLC) attended the same 6-day training. As in prior STEPs trials, the training was led by doctoral-level clinical psychologists with expertise in MATCH and evidence-based youth psychotherapy. Trainings included didactic instruction, video illustrations, and interactive exercises (e.g., role-playing). Clinicians in both groups had access to lower-cost resources that included (a) MATCH handouts, (b) videos of the full MATCH training, (c) video tutorials in which MATCH experts and young actors illustrated treatment sessions, (d) free access to the PracticeWise website that provides additional MATCH information and downloadable materials (e.g., youth worksheets, caregiver handouts), and (e) free access to weekly feedback on each youth's response to treatment via our web-based MBC system called Progress Assessment in Therapy (PATH). The PATH system displayed a dashboard for each youth, updated weekly, with youth- and caregiver-report scores on the BPM and the TPA (see both below), plus as a record of which modules were covered, week by week.

Included in CLC only—expert consultation. Therapists in the CLC condition received weekly group consultation led by MATCH consultants. As in prior MATCH trials, these were doctoral-level psychologists who had expertise in MATCH and had coled MATCH trainings. Consultation groups were kept small

(average of two clinicians per call; range = 1–3) to provide sufficient time for discussion of all study cases on each clinician's caseload. Consultation included review of each youth's treatment response, using the PATH dashboard to facilitate MBC, discussion of modules used and the youth's response, identification of treatment challenges for each case and possible solutions, planning for subsequent sessions, and an overall focus on personalizing treatment to make each MATCH case as successful as possible.

Although LC clinicians did not receive MATCH consultation via the study, some of them organized peer consultation with their LC colleagues to discuss using MATCH with their cases. We obtained posttherapy reports from 41 of the 42 therapists on the amount of MATCH consultation they had received on their study cases. LC therapists reported a mean of 5.75 hr per case (SD = 7.60); CLC therapists reported 13.82 hr (SD = 10.07), $t(39) = -2.857$, $p = .007$, $d = 0.90$. Viewed another way, 100% of CLC therapists received MATCH consultation via the study; LC therapists received no MATCH consultation via the study, but 78.9% participated in peer consultation, albeit for significantly less consultation time than CLC therapists had.

Measures

We measured clinical progress, outcomes, and clinician fidelity to MATCH. Clinical progress and outcomes were assessed following two longitudinal measurement schedules: (a) comprehensive standardized measures administered at baseline and 3, 6, 9, 12, and 18 months thereafter, as well as at posttreatment, and (b) brief rating scales administered weekly during treatment. All trained research assistant assessors were kept naive to participant study condition.

TPA. To provide an idiographic measure of progress on the problems identified as most important by each youth and caregiver, we administered the TPA (Weisz et al., 2011) weekly throughout treatment. The TPA consists of youth and caregiver severity ratings for the three problems identified by the youth and caregiver as most important to them, in separate structured baseline interviews. Psychometric analyses in Weisz et al. (2011) with clinic-referred youths and caregivers supported the test–retest reliability (.69–.91, across 5- to 21-day intervals), convergent and discriminant validity (relative to standardized measures), and sensitivity to change during treatment. The TPA also showed superiority of MATCH over usual care in previous trials (Chorpita et al., 2017; Weisz et al., 2012).

BPM. To provide a standardized measure of progress on a common set of clinically significant problems, we administered the BPM (Achenbach, McConaughy, & Rescorla, 2011; Piper, Gray, Raber, & Birkett, 2014) weekly throughout treatment. The BPM caregiver form includes 19 items derived from the Child Behavior Checklist (CBCL; see below); the youth-report BPM includes 19 similar items from the Youth Self-Report (YSR; see below). Caregivers and youths rate the items on a 3-point scale (0 = *not true,* 1 = *somewhat true,* 2 = *very true*), yielding scores for internalizing problems (6 items; range), externalizing problems (7 items), attention problems (6 items), and total problems (19 items). Evidence supports the reliability and validity of the BPM (Piper et al., 2014). Given MATCH's focus on affective and disruptive behavior problems, our primary focus was on the BPM internalizing and externalizing scales, but we included the standard BPM total scale and the BPM attention problems scale to offer a more complete picture of the results.

CBCL and YSR. The CBCL and YSR (Achenbach & Rescorla, 2001) were used to provide a more comprehensive assessment. The CBCL is a caregiver-report measure with 113 youth problem items, each rated on a 0–2 scale (2 = *very/often true*). The YSR is a 112-item youth self-report measure with most items corresponding to those of the CBCL. Both measures generate T scores, adjusted for age and gender, for internalizing, externalizing, and total problems. Evidence for CBCL and YSR validity and reliability is extensive (Achenbach & Rescorla, 2001). Both measures showed significant superiority of MATCH over usual care in the first STEPs RCT (Weisz et al., 2012).

EEBPS. The therapist-report EEBPS measure, developed specifically for this study, is a multiple-choice questionnaire with item sets referencing treatments relevant to MATCH: (a) trauma-focused cognitive–behavior therapy (CBT), (b) dialectical behavior therapy (DBT), (c) other kinds of CBT, and (d) behavioral parent training (BPT) for disruptive behavior. For each of these, therapists are asked to report "About how many . . ." hours of training, hours of supervision, unique clients you have treated, and hours spent delivering the treatment. The final item assesses overall experience in EBTs on a scale from 1 (*not at all experienced*) to 5 (*very experienced*). As shown in Figure 2, the majority of the 42 participating therapists had received training and supervision in, and had treated young clients with, both trauma-focused CBT and other forms of CBT; they had less experience with BPT and very little experience using DBT (but a majority had had some form of DBT training).

Therapist Integrity in Evidence-Based Interventions. As in previous STEPs trials (Chorpita et al., 2017; Weisz et al., 2012, 2019), session recordings were coded for adherence to the evidence-based treatment procedures of MATCH. Sessions were coded in 5-min segments for presence/absence of 27 items reflecting therapist adherence and competence in the use of MATCH. To generate a percent adherence score for each coded treatment session for each MATCH content item, the number of 5-min segments in which any MATCH item was coded present was summed, multiplied by 5 (number of minutes in each segment), and divided by the total time of the session in minutes. The resulting percentages were averaged across all the coded sessions for each youth's treatment episode. Coding of therapist competence is based on coders' global ratings of skillfulness of delivery of each item of MATCH content, ranging from 0 (*not at all*) to 4 (*expert*). To generate a competence score for each youth's treatment episode, we calculated the averages of all nonzero global competence scores for all EBT items on the Therapist Integrity in Evidence-Based Interventions (TIEBI; Bearman, Herren, & Weisz, 2012; Bearman, Schneiderman, & Zoloth, 2017).

Coders were 12 bachelor's- and master's-level research assistants supervised by the primary TIEBI developer, including training to pass a reliability standard on practice sessions before study coding. On study sessions, mean intercoder agreement (on 53 randomly selected sessions [i.e., 10%]) was M ICC (intraclass correlation coefficient) for adherence [1,1] = 0.79 and M ICC for competence [1,1] = 0.70. For this study, 529 sessions (272 from LC cases, 256 from CLC) were selected with the following constraints: no more than three session recordings per case; randomly selected from the first, middle, and last thirds of the full treatment

episode; and excluding initial sessions (these often had administrative content) and sessions of unrepresentative length (< 15 min or > 75 min). All coders were kept naive to participant identity, characteristics, and study condition.

Data and Analyses

To investigate therapist adherence and competence, we used TIEBI data. Group differences in TIEBI adherence and competence were evaluated as in previous trials using *t* tests, with Cohen's *d* effect sizes. Following Cohen (1988) and Lipsey and Aiken (1990), we interpreted *d as* large if > 0.80–0.90, medium if > 0.45–0.50, and small if > 0.15–0.20.

Following prior STEPs trials, our primary analyses of youth clinical outcomes focused on trajectories of change over time, both during treatment (via the weekly BPM and TPA) and long term up to 18 months postbaseline (via the quarterly CBCL/YSR). These analyses used multilevel models with repeated observations (Level 1) nested within each individual (Level 2) and each individual nested within therapist (Level 3). Clinic effects were controlled for by using fixed-effects dummy codes at Level 3. As in prior STEPs trials, time was modeled as number of days since baseline, using the natural logarithm of days since baseline + 1 for analyses. Questions of interest were as follows: (a) Do youths in each condition improve over time (i.e., is each condition's slope significantly negative)? and (b) Do youths in one condition improve faster than those in the other condition (i.e., are the conditions' slopes significantly different)?

The magnitude and clinical significance of these results were explored in several ways. As in prior STEPS trials, effect sizes (ESs) were computed as the difference between two groups' slopes divided by the square root of the overall slope variance. This yields an ES interpretable similar to Cohen's *d*. Clinical significance was investigated by using the multilevel models to generate two types of model-implied estimates. First, we examined average improvement in clinical outcomes at 1 and 2 years postbaseline. Second, in a supplemental analysis, we examined youths' posttreatment scores on outcome measures, estimated by centering the longitudinal intercept term such that "day 0" represented the day their posttreatment assessment occurred (in 78.5% of cases) or should have occurred (21.5%). Both approaches offer additional insights, beyond the primary slope contrasts, into the extent to which youths in each condition improved following treatment, in both relative and absolute terms. Results are interpreted according to the clinically meaningful properties of the measures, including published clinical cutoffs, normed *T* score points, and descriptively anchored scales.

Regarding data availability, repeated assessments for both the weekly measurement schedule (with an undefined number of occasions because treatment durations varied) and the quarterly schedule (up to six occasions: 0, 3, 6, 9, 12, and 18 months) were designed to collect ample data to accommodate missing observations and still facilitate analyses of trajectories based on the available data for each participant. Data availability considerations thus focus on number of total observations collected for each informant, measurement schedule, and condition. Families completed a mean of 23.69 (SD = 16.55) weekly youth-report observations and 25.05 (SD = 17.36) weekly caregiver-report observations. The CLC and LC groups did not differ significantly on number of weekly youth reports (M_{CLC} = 25.42, SD = 17.46; M_{LC} = 22.00, SD = 15.52; p = .147) or caregiver reports (M_{CLC} = 27.17, SD = 15.93; M_{LC} = 22.97, SD = 18.55; p = .087). All (100%) caregivers and 99% of youths completed at least one weekly assessment (one youth in each condition elected not to participate in weekly assessments).

Regarding the long-term assessment schedule, there was a mean of 4.47 (SD = 1.86) quarterly observations by youth report, and 4.63 (SD = 1.78) quarterlies by caregiver report. The CLC and LC groups did not differ significantly in number of completed quarterly assessments by youth report (M_{CLC} = 4.74, SD = 1.80; M_{LC} = 4.21, SD = 1.90, $t(198)$ = 2.023, p = .044, d = 0.29) or caregiver report (M_{CLC} = 4.94, SD = 1.69; M_{LC} = 4.33, SD = 1.82; $t(198)$ = 2.468, p = .014, d = 0.35). Longitudinal attrition for quarterly measures (0, 3, 6, 9, 12, 18 months) followed a similar pattern by caregiver report (0%, 15%, 23%, 30%, 32%, 39%) and youth report (1%, 17%, 26%, 34%, 37%, 32%). During the latter part of study, participation rates appeared to diverge by condition. At 6 months, caregiver (and youth in parentheses) missing data rates for CLC versus LC were 17% (20%) versus 29% (32%), respectively; at 9 months, these missing data rates were 21% (27%) versus 38% (40%); at 12 months, 23% (28%) versus 40% (43%); and at 18 months, 28% (31%) versus 49% (51%).

Power was simulated a priori to guide sample size determination and estimated again post hoc to identify minimum detectable effect sizes (Bloom, 1995) given the data we obtained. We used the Optimal Design (Raudenbush et al., 2011) protocol for a cluster-randomized trial, with therapists as clusters, and repeated measures of patient-level outcomes at Level 1. Assuming standard thresholds for significance (α = .05) and power ($1 - \beta$ = 0.8), with 42 therapists averaging 4.76 study cases and an average intraclass correlation coefficient estimate of 0.02, we obtained adequate power to detect a minimum effect size of approximately 0.47 for the primary outcomes of youth CBCL/YSR symptom trajectories (averaging 4–5 quarterly observations) and 0.45 for the BPM/TPA problem trajectories (averaging 24–25 weekly observations). Applying the same parameters to the TIEBI analyses, we were powered to detect a smaller effect size of 0.25 on adherence and competence outcomes. Trajectory models were estimated using HLM Version 7 (Raudenbush et al., 2011) with full-information maximum likelihood to accommodate missing observations in the data. Other analyses were conducted in SPSS Version 25 (IBM, 2017).

Results

Therapist Adherence and Competence

TIEBI session coding showed therapist adherence to be relatively strong in both the LC (M = 66.55%, SD = 37.99) and CLC (M = 79.41%, SD = 31.24) conditions, but CLC therapists showed stronger adherence than LC therapists, $t(526)$ = 4.23, p < .001, d = 0.37. On TIEBI competence ratings, the two therapist groups did not differ significantly (CLC M = 1.93, SD = 0.80, LC M = 1.80, SD = 0.92, $t(526)$ = 1.74, p = .082, d = 0.15).

Trajectories of Change in Clinical Outcomes Over Time

Results of the primary outcome analyses are shown in Table 1, with the concomitant change estimates presented in Table 2. As shown, the two groups did not differ from one another in slopes of improvement on any outcome measure (all *p*s > .10; see Table 1). Further, the effect sizes for these contrasts clustered around zero, with specific effects falling in the negligible to small range in both directions (ES range: −0.26 [favoring CLC] to 0.18 [favoring LC]).

However, results did show within-group improvements. Within each condition, the log-linear slope estimates for all measures (see Table 2) were consistently significantly negative (*p*s < .02). As shown, these changes amounted to meaningful reductions in problem severity on all measures as reflected in 1- and 2-year estimates. For example, 2 years after starting treatment, both groups' YSR score estimates had fallen by about 8–13 *T* score points, and CBCL score estimates had fallen by about 7–11 *T* score points. This is a meaningful change, considering that ~10 *T* score points represent one standard deviation in the population. Results of similar magnitude were found for the BPM and TPA outcomes, although these were analyzed on raw scale score metrics to retain their statistical sensitivity to change (Achenbach et al., 2011). Indeed, 1 year after beginning treatment, each group's change reflected an improvement of more than one half standard deviation relative to the sample's baseline scores. Overall, these indicators suggest that changes within each group reflect clinically meaningful improvements.

So, the trajectory analyses indicate that youths receiving MATCH in the LC and CLC conditions showed similar statistically and clinically significant improvements over time. This same pattern of improvement was documented on multiple outcome measures, using both youth and caregiver report, and based on short-term and long-term assessment schedules.

Supplemental Analyses of Posttreatment Outcomes

The analyses of improvement over time were primary, just as in prior STEPs trials, in part because duration varies widely, producing marked individual and group differences in dose of treatment received and number of assessments completed. However, posttreatment outcomes are of interest and were included in secondary analyses. Figure 4 presents the model-implied posttreatment intercept estimates for youths on all outcome measures by condition. Like the slope coefficients interpreted above, posttreatment intercept coefficients showed little to no LC versus CLC difference on most outcome measures, but there were a few statistically significant exceptions. The CLC condition showed somewhat more favorable posttreatment outcomes than the LC condition on youth-reported BPM internalizing problems (p = .016, ES = −0.33), BPM total problems (which overlaps BPM internalizing; p = .041, ES = −0.30), and top problem severity (p = .027, ES = −0.49), as well as on caregiver-reported BPM internalizing (p = .045, ES = −0.28). Because these findings fall within an overall pattern of largely nonsignificant differences, plus the finding of no slope difference on any measure, they may not be clinically meaningful.

What does appear meaningful in Figure 4 is the absolute level of outcomes for both groups. At posttreatment, youths in both conditions had average estimated CBCL *T* scores between 55 and 59 and YSR *T* scores around 45–48, showing that both conditions had fallen to within the normal range. Findings were similar when we used BPM norms to extrapolate from our results and convert BPM raw score estimates into approximated *T* scores, aggregated across the gender and age norms for our sample. At posttreatment, the caregiver-rated BPM estimates translate to approximate *T* scores of 54–64 and youth-rated BPM estimates in the range of 51–61, below the BPM cutoff of 65. Thus, average scores in LC and CLC fell within the nonclinical range at posttreatment. Finally, mean TP severity scores fell about 2 points from baseline, to posttreatment model-estimated means of 1.42–2.09, closer to the scale value 0 (*not a problem*) than 4 (*a very big problem*). Thus, supplemental posttreatment analyses showed a pattern similar to that found in the primary outcome analyses: significant, meaningful improvement on all outcome measures, with most outcomes similar for LC and CLC.

Table 1
Estimates Comparing Groups' (LC Versus CLC) Trajectories of Change on Youth-Reported and Caregiver-Reported Scores

Variable	Youth report			Caregiver report		
	Estimate	*p* value	Effect size	Estimate	*p* value	Effect size
Weekly monitoring measures						
BPM internalizing	−0.005	.96	−0.01	−0.085	.40	−0.15
BPM externalizing	−0.083	.39	−0.14	0.061	.52	0.11
BPM attention	−0.083	.37	−0.15	0.024	.80	0.04
BPM total	−0.170	.48	−0.11	−0.018	.94	−0.01
Top problem severity	−0.062	.13	−0.25	−0.057	.12	−0.26
Quarterly outcome measures						
CBCL/YSR internalizing	0.159	.58	0.12	−0.205	.37	−0.21
CBCL/YSR externalizing	0.200	.43	0.18	−0.039	.86	−0.04
CBCL/YSR total	0.053	.84	0.04	−0.193	.39	−0.19

Note. Negative values in the "Estimate" columns indicate a faster reduction in problems/symptoms for the CLC group than the LC group; positive values reflect the opposite. LC = lower-cost supports only; CLC = consultation plus lower-cost supports; BPM = Brief Problem Monitor; CBCL = Child Behavior Checklist; YSR = Youth Self-Report.

Table 2
Log-Linear Slope and Change Estimates by Treatment Condition

Variable	LC condition			CLC condition		
	Slope	1-year change	2-year change	Slope	1-year change	2-year change
Youth-report weekly measures						
BPM-Y internalizing	−0.22	−1.33	—	−0.23	−1.35	—
BPM-Y externalizing	−0.20	−1.19	—	−0.20	−1.18	—
BPM-Y attention	−0.26	−1.54	—	−0.34	−2.03	—
BPM-Y total	−0.67	−3.97	—	−0.84	−4.97	—
TPA-Y	−0.27	−1.60	—	−0.33	−1.97	—
Youth-report quarterly measures						
YSR internalizing	−2.03	−11.97	−13.38	−1.87	−11.04	−12.33
YSR externalizing	−1.39	−8.18	−9.14	−1.19	−7.00	−7.82
YSR total	−1.92	−11.32	−12.65	−1.87	−11.00	−12.30
Caregiver-report weekly measures						
BPM-P internalizing	−0.20	−1.16	—	−0.28	−1.66	—
BPM-P externalizing	−0.29	−1.69	—	−0.23	−1.33	—
BPM-P attention	−0.23	−1.34	—	−0.20	−1.20	—
BPM-P total	−0.68	−3.99	—	−0.69	−4.10	—
TPA-P	−0.26	−1.51	—	−0.31	−1.85	—
Caregiver-report quarterly measures						
CBCL internalizing	−1.50	−8.87	−9.91	−1.71	−10.08	−11.27
CBCL externalizing	−1.11	−6.57	−7.34	−1.15	−6.80	−7.59
CBCL total	−1.35	−7.94	−8.88	−1.54	−9.08	−10.15

Note. Negative signs for slopes, 1-year change, and 2-year change indicate reduced problem/symptom levels over time. For example, consider YSR total problems: The 1-year change results indicate that 12 months after starting treatment in the LC condition, the average youth's YSR total problems *T* score dropped by about 11.32 points. Estimates for 2-year change are only reported for the long-term outcome measures; 2-year change was not estimated for the BPM and TPA scales because these were both administered for a shorter period of time, during active treatment only. BPM = Brief Problem Monitor; CBCL = Child Behavior Checklist; YSR = Youth Self-Report; TPA = Top Problems Assessment; Y = youth; P = parent; LC = lower-cost supports only; CLC = consultation plus lower-cost supports.

Discussion

We tested whether a diverse combination of lower-cost implementation supports might provide sufficient scaffolding for skill building by community clinicians learning a complex transdiagnostic treatment or whether fidelity and clinical outcomes would be boosted markedly by adding weekly expert consultation. Our findings suggested that adding expert consultation did significantly improve adherence to the evidence-based MATCH procedures but not MATCH competence. Clinical outcomes on standardized and idiographic measures were quite similar in the two study conditions, with just a few differences in supplemental analyses favoring the consultation group. The findings suggest that giving community clinicians a combination of multiple lower-cost implementation supports may provide the skill building they need to produce clinical outcomes quite comparable to those achieved when expert consultation is added to the mix. Offering such lower-cost supports could be viewed as part of a stepped process of skill building, beginning with intensive training and then stepping down to inexpensive, sustainable ways of enhancing and applying the trained skills. This could be helpful to providers and organizations who seek training in EBTs but whose resources cannot accommodate the cost and complexity of arranging expert consultation after the training is completed.

Youths in both groups responded well to treatment. Across multiple methods, measures, and informants, analyses revealed that youths in both conditions showed statistically and clinically significant improvement in internalizing, externalizing, attention, total problems, and youth- and caregiver-identified top problems—with problem levels dropping to nonclinical levels. Our primary clinical outcome analysis, focused on trajectories of change, showed very similar slopes for the LC and CLC groups, with no significant difference on any measure. Our supplemental analysis of posttreatment scores showed similar outcomes for LC and CLC, with 4 of 16 measures showing a significant advantage for CLC. Thus, it appeared that adding consultation to the lower-cost supports may have produced a slight advantage in MATCH adherence and a small number of secondary outcomes, but for MATCH competence and the great majority of clinical outcomes, there was no evidence that adding MATCH consultation enhanced clinical outcomes over and above the lower-cost MATCH resources and supports that were freely available to both groups.

This RCT adds to a growing body of evidence on strategies for transporting Child STEPs to community mental health settings. A previous study (Weisz et al., 2018) indicated that training staff clinicians to supervise MATCH implementation in their clinics can provide ample levels of clinician fidelity and youth outcomes. Another study (Weisz et al., 2019) suggested that thinning out consultation by multiplying the number of clinicians per consultation hour might undermine fidelity and outcomes. The present findings suggest that a relatively lower-cost combination of multiple implementation supports may actually provide a reasonably effective substitute for expert consultation.

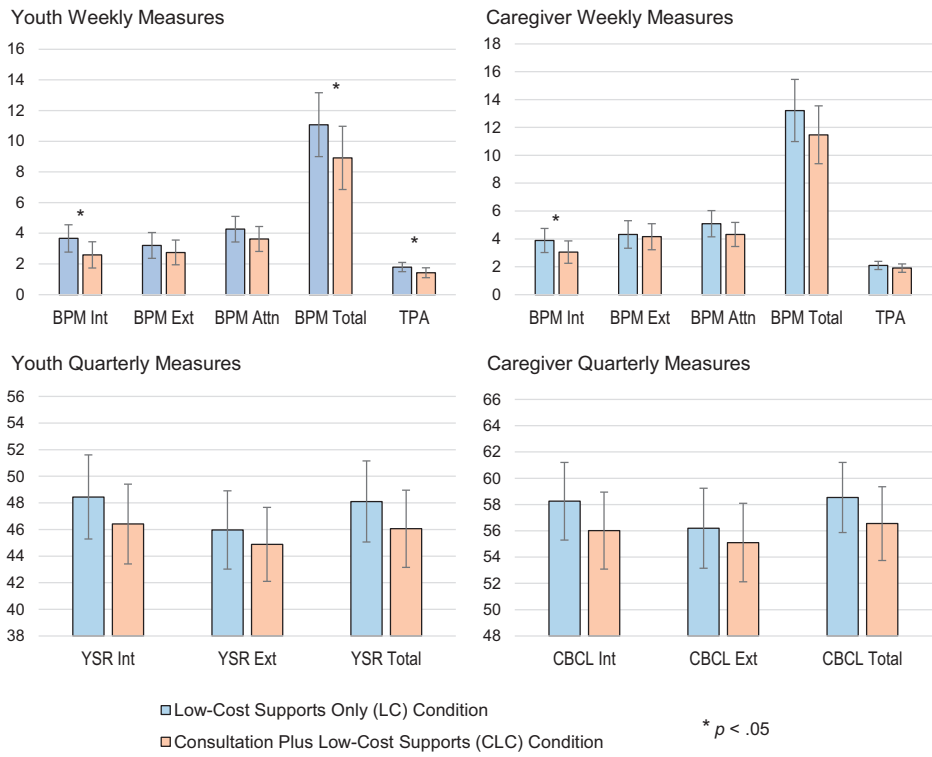

Figure 4. Model-implied posttreatment outcomes on all measures by condition. Estimates are scaled with the *y*-axis representing raw scores for BPM and TPA and *T* scores for CBCL and YSR. Error bars show 95% confidence interval bands (computed as ± 1.96**SE*). BPM = Brief Problem Monitor; TPA = Top Problems Assessment; CBCL = Child Behavior Checklist; YSR = Youth Self-Report; Int = internalizing; Ext = externalizing; Attn = attention; Total = total problems.

In its emphasis on both implementation and clinical outcomes, our study approximates what Landes, McBain, and Curran (2019) have identified as a Type 3 hybrid implementation-effectiveness trial (see also Curran, Bauer, Mittman, Pyne, & Stetler, 2012). In fact, it is useful to consider this study in the broader context of implementation science and efforts to address "the implementation cliff" (Weisz, Ng, & Bearman, 2014), the drop-off in benefit often seen when EBTs move from efficacy studies to implementation in real-world clinical care contexts. Williams and Beidas (2019) have proposed that "optimizing the implementation of effective treatments in community care for youth with psychiatric disorders is a defining challenge of our time" (p. 430) and have suggested that training approaches tested thus far have generally not been very effective in producing behavior change. Other authors have raised similar concerns. For example, in a recent systematic review of research on EBT training, Valenstein-Mah et al. (2020) found a few encouraging studies but generally mixed evidence on the effectiveness of various kinds of posttraining consultation. Nadeem and colleagues (2013), in a rich analysis of the consultation research literature, noted that "we have very limited understanding of how expert consultation fits into the larger implementation support system, or the most effective consultation strategies" (p. 439). Offering a useful perspective, Wandersman, Chien, and Katz (2012) have argued that successful implementation of EBTs requires tools, training, technical assistance, and quality assurance and that technical assistance may take multiple forms, only one of which is expert consultation. Consistent with that view, our findings suggest that pooling multiple, lower-cost, alternative forms of technical assistance may approximate the benefits of the expert consultation that many of us have thought to be a necessity but that some providers and organizations may find difficult to arrange or afford. In this respect, our findings may suggest one pragmatic step toward addressing the implementation cliff.

That conclusion must be qualified, however, by certain limitations of the study. First, although it seems clear that our LC condition was lower in cost than our CLC condition, we did not conduct a formal cost analysis. It is true that exactly the same lower-cost supports were available to clinicians in LC and CLC, with only the CLC condition involving the additional expense of external consultants. However, our inability to monitor all the activities of all clinicians throughout their workdays left us unable to calculate the amount of time LCs and CLCs devoted to accessing the lower-cost supports, and that rules out a precise cost assessment. Experts have noted the striking absence of cost assessments in implementation research (e.g., Eisman, Kilbourne, Dopp, Saldana, & Eisenberg, 2020; Valenstein-Mah et al., 2020) and the potency such assessments would have in making the business case with providers and administrators who must watch their bottom line; we agree, and we hope to see this gap addressed in future research. The fact that we could not monitor all clinician

activity at all times also meant that we could not ensure that no CLC clinicians communicated consultation content with any LC clinician; all CLCs had committed not to do so as part of their agreement to participate in the study, but the fact that we could not monitor all their behavior makes this another study limitation. Our inability to monitor closely also meant that we lack information on the extent to which clinicians in the two conditions accessed and used each of the various lower-cost implementation supports; that would be useful information to collect in future research. A fourth limitation is that the study did not provide data on feasibility or the perspectives of agency leaders or clinicians. Regardless of the relative cost of LC and CLC, providers and service administrators might find one approach to be more feasible than the other, and we might have detected this if we had included such methods as qualitative assessments and focus groups or quantitative measures of acceptability, appropriateness, and feasibility (e.g., Weiner et al., 2017); this might also have provided a lens into other elements of the taxonomy of implementation outcomes proposed by Proctor et al. (2011). A fifth aspect of our study could also be considered a limitation, at least in regard to generalizability: The participating clinicians already had some degree of experience with EBTs prior to the study. The prior experience included multiple trainings in EBTs and some collection and reporting of treatment outcome data. That base of experience may have produced a readiness to absorb Child STEPs without external consultation, as long as support was provided in such forms as youth and caregiver handouts, access to a website with MATCH information and materials, access to the MATCH training videos and tutorials, and measurement-based care feedback. Whether such lower-cost supports would be sufficient without external consultation for groups of clinicians who have less prior experience with EBTs is an interesting empirical question for future research. Whatever the findings, such research could contribute usefully to our efforts to build cost-effective implementation strategies.

The present study joins with other efforts in our field to make effective delivery of EBTs a part of the landscape of everyday clinical practice. To attain this goal may require work that brings together the perspectives of clinical science and clinical practice. Such a collaboration may inform the development of clinician skill-building strategies that are sufficient to produce measurable clinical benefit and feasible for providers who operate within limited financial resources. Our findings suggest one strategy that may hold promise for some mental health care settings. We hope this will contribute to an ever-expanding collaboration between science and practice, to the benefit of those who seek effective evidence-based mental health care in their communities.

References

Aarons, G. A., Wells, R. S., Zagursky, K., Fettes, D. L., & Palinkas, L. A. (2009). Implementing evidence-based practice in community mental health agencies: A multiple stakeholder analysis. *American Journal of Public Health, 99,* 2087–2095. http://dx.doi.org/10.2105/AJPH.2009.161711

Achenbach, T. M., McConaughy, S. H., & Rescorla, L. A. (2011). *Manual for the ASEBA Brief Problem Monitor (BPM)*. Burlington: University of Vermont, Research Center for Youths, Youth and Families.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington: University of Vermont, Research Center for Youths, Youth and Families.

Bearman, S. K., Herren, J., & Weisz, J. R. (2012). *Therapy integrity in evidence based interventions: Observational coding system, coding manual*. Austin: University of Texas at Austin.

Bearman, S. K., Schneiderman, R. L, & Zoloth, E. (2017). Building an evidence base for effective supervision practices: An analogue experiment of supervision to increase EBT fidelity. *Administration and Policy in Mental Health and Mental Health Services Research, 44,* 293–307.

Beidas, R. S., Edmunds, J. M., Marcus, S. C., & Kendall, P. C. (2012). Training and consultation to promote implementation of an empirically supported treatment: A randomized trial. *Psychiatric Services, 63,* 660–665. http://dx.doi.org/10.1176/appi.ps.201100401

Bloom, H. S. (1995). Minimum detectable effects: A simple way to report the statistical power of experimental designs. *Evaluation Review, 19,* 547–556. http://dx.doi.org/10.1177/0193841X9501900504

Chorpita, B. F., Daleiden, E. L., Park, A. L., Ward, A. M., Levy, M. C., Cromley, T., . . . Krull, J. L. (2017). Child STEPs in California: A cluster 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,* 13–25. http://dx.doi.org/10.1037/ccp0000133

Chorpita, B. F., & Weisz, J. R. (2009). *MATCH-ADTC: Modular approach to therapy for children with anxiety, depression, trauma, or conduct problems*. Satellite Beach, FL: PracticeWise.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.

Curran, G. M., Bauer, M., Mittman, B., Pyne, J. M., & Stetler, C. (2012). Effectiveness-implementation hybrid designs: Combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical Care, 50,* 217–226. http://dx.doi.org/10.1097/MLR.0b013e3182408812

Eisman, A. B., Kilbourne, A. M., Dopp, A. R., Saldana, L., & Eisenberg, D. (2020). Economic evaluation in implementation science: Making the business case for implementation strategies. *Psychiatry Research, 283,* 112433. http://dx.doi.org/10.1016/j.psychres.2019.06.008

Fortney, J. C., Unutzer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H. T. (2017). A tipping point for measurement-based care. *Psychiatric Services, 68,* 179–188. http://dx.doi.org/10.1176/appi.ps.201500439

IBM. (2017). *IBM SPSS Statistics for Windows, Version 25.0*. Armonk, NY: Author.

Landes, S. J., McBain, S. A., & Curran, G. M. (2019). An introduction to effectiveness-implementation hybrid designs. *Psychiatry Research, 280,* 112513. http://dx.doi.org/10.1016/j.psychres.2019.112513

Lang, J. M., & Connell, C. M. (2017). Measuring costs to community-based agencies for implementation of an evidence-based practice. *Journal of Behavioral Health Services & Research, 44,* 122–134. http://dx.doi.org/10.1007/s11414-016-9541-8

Lipsey, M. W., & Aiken, L. S. (1990). *Design sensitivity: Statistical power for experimental research* (Vol. 19). Thousand Oaks, CA: Sage.

Merry, S. N., Hopkins, S., Lucassen, M. F. G., Stasiak, K., Weisz, J. R., Frampton, C. M. A., . . . Crengle, S. (in press). Clinician training in the modular approach to therapy for children to enhance the use of empirically supported treatments and clinical outcomes: A randomized clinical trial. *Journal of the American Medical Association Psychiatry Open*.

Nadeem, E., Gleacher, A., & Beidas, R. S. (2013). Consultation as an implementation strategy for evidence-based practices across multiple contexts: Unpacking the black box. *Administration and Policy in Mental Health and Mental Health Services Research, 40,* 439–450. http://dx.doi.org/10.1007/s10488-013-0502-8

Pegg, S. L., Walsh, L. M., Becker-Haimes, E. M., Ramirez, V., & Jensen-Doss, A. (2019). Money makes the world go 'round: A qualitative examination of the role funding plays in large-scale implementation and sustainment of youth evidence-based practice. *Psychological Services*. Advance online publication. http://dx.doi.org/10.1037/ser0000399

Piper, B. J., Gray, H. M., Raber, J., & Birkett, M. A. (2014). Reliability and validity of the Brief Problem Monitor, an abbreviated form of the Child Behavior Checklist. *Psychiatry and Clinical Neurosciences, 68,* 759–767. http://dx.doi.org/10.1111/pcn.12188

Powell, B. J., Waltz, T. J., Chinman, M. J., Damschroder, L. J., Smith, J. L., Matthieu, M. M., . . . Kirchner, J. E. (2015). A refined compilation of implementation strategies: Results from the Expert Recommendations for Implementing Change (ERIC) project. *Implementation Science, 10,* 21. http://dx.doi.org/10.1186/s13012-015-0209-1

Proctor, E. K., Landsverk, J., Aarons, G., Chambers, D., Glisson, C., & Mittman, B. (2009). Implementation research in mental health services: An emerging science with conceptual, methodological, and training challenges. *Administration and Policy in Mental Health and Mental Health Services Research, 36,* 24–34. http://dx.doi.org/10.1007/s10488-008-0197-4

Proctor, E. K., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., . . . Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38,* 65–76. http://dx.doi.org/10.1007/s10488-010-0319-7

Raudenbush, S. W., Spybrook, J., Congdon, R., Liu, X. F., Martinez, A., Bloom, S. W., & Hill, C. (2011). Optimal design software for multi-level and longitudinal research (Version 3.01) [Computer software]. Retrieved from https://sites.google.com/site/optimaldesignsoftware/home

Roundfield, K. D., & Lang, J. M. (2017). Costs to community mental health agencies to sustain an evidence-based practice. *Psychiatric Services, 68,* 876–882. http://dx.doi.org/10.1176/appi.ps.201600193

Schleider, J. L., & Weisz, J. R. (2017). Little treatments, promising effects: Meta-analysis of single-session interventions for youth psychiatric problems. *Journal of the American Academy of Child & Adolescent Psychiatry, 56,* 107–115. http://dx.doi.org/10.1016/j.jaac.2016.11.007

Schoenwald, S. K., Kelleher, K., Weisz, J. R., & the Research Network on Youth Mental Health. (2008). Building bridges to evidence-based practice: The MacArthur Foundation Child System and Treatment Enhancement Projects (Child STEPs). *Administration and Policy in Mental Health and Mental Health Services Research, 35,* 66–72. http://dx.doi.org/10.1007/s10488-007-0160-9

Schoenwald, S. K., Sheidow, A. J., & Chapman, J. E. (2009). Clinical supervision in treatment transport: Effects on adherence and outcomes. *Journal of Consulting and Clinical Psychology, 77,* 410–421. http://dx.doi.org/10.1037/a0013788

Schwalbe, C. S., Oh, H. Y., & Zweben, A. (2014). Sustaining motivational interviewing: A meta-analysis of training studies. *Addiction, 109,* 1287–1294. http://dx.doi.org/10.1111/add.12558

Scott, K., & Lewis, C. (2015). Using measurement-based care to enhance any treatment. *Cognitive and Behavioral Practice, 22,* 49–59. http://dx.doi.org/10.1016/j.cbpra.2014.01.010

Sholomskas, D., Syracuse, G., Rounsaville, B. J., & Carroll, K. M. (2005). We don't train in vain: A randomized trial of three strategies for training clinicians in CBT. *Journal of Consulting and Clinical Psychology, 73,* 106–115.

Smith-Boydston, J. M., Holtzman, R. J., & Roberts, M. C. (2014). Transportability of Multisystemic Therapy to community settings: Can a program sustain outcomes without MST Services oversight? *Child & Youth Care Forum, 43,* 593–605. http://dx.doi.org/10.1007/s10566-014-9255-0

Valenstein-Mah, H., Greer, N., McKenzie, L., Hansen, L., Strom, T. Q., Wiltsey-Stirman, S., . . . Kehle-Forbes, S. M. (2020). Effectiveness of training methods for delivery of evidence-based psychotherapies: A systematic review. *Implementation Science, 15,* 1–17.

Wandersman, A., Chien, V. H., & Katz, J. (2012). Toward an evidence-based system for innovation support for implementing innovations with quality: Tools, training, technical assistance, and quality assurance/quality improvement. *American Journal of Community Psychology, 41,* 171–181. http://dx.doi.org/10.1007/s10464-008-9174-z

Webster-Stratton, C. H., Reid, J., & Marsenich, L. (2014). Improving therapist fidelity during implementation of evidence-based practices: Incredible Years Program. *Psychiatric Services, 65,* 789–795. http://dx.doi.org/10.1176/appi.ps.201200177

Weiner, B. J., Lewis, C. C., Stanick, C., Powell, B. J., Dorsey, C. N., Clary, A. S., . . . Halko, H. (2017). Psychometric assessment of three newly developed implementation outcome measures. *Implementation Science, 12,* 1–12.

Weisz, J. R., Bearman, S. K., Ugueto, A. M., Herren, J. A., Evans, S. C., Cheron, D. M., . . . Jensen-Doss, A. (2019). Testing the robustness of Child STEPs effects: A randomized controlled effectiveness trial. *Journal of Clinical Child and Adolescent Psychology*. Advance online publication. http://dx.doi.org/10.1080/15374416.2019.1655757

Weisz, J. R., & Chorpita, B. F. (2011). Mod squad for youth psychotherapy: Restructuring evidence-based treatment for clinical practice. In P. C. Kendall (Ed.), *Child and adolescent therapy: Cognitive-behavioral procedures* (4th ed., pp. 379–397). New York, NY: Guilford.

Weisz, J. R., Chorpita, B. F., Frye, A., Ng, M. Y., Lau, N., Bearman, S. K., . . . Hoagwood, K. E. (2011). Youth top problems: Using idiographic, consumer-guided assessment to identify treatment needs and track change during psychotherapy. *Journal of Consulting and Clinical Psychology, 79,* 369–380. http://dx.doi.org/10.1037/a0023307

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., & Bearman, S. K. (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,* 274–282. http://dx.doi.org/10.1001/archgenpsychiatry.2011.147

Weisz, J. R., Krumholz, L. S., Santucci, L., Thomassin, K., & Ng, M. (2015). Shrinking the gap between research and practice: Tailoring and testing youth psychotherapies in clinical care contexts. *Annual Review of Clinical Psychology, 11,* 139–163. http://dx.doi.org/10.1146/annurev-clinpsy-032814-112820

Weisz, J. R., Ng, M. Y., & Bearman, S. K. (2014). Odd couple? Re-envisioning the relation between science and practice in the dissemination-implementation era. *Clinical Psychological Science, 2,* 58–74. http://dx.doi.org/10.1177/2167702613501307

Weisz, J. R., Ugueto, A. M., Herren, J., Marchette, L. K., Bearman, S. K., Lee, E. H., . . . Jensen-Doss, A. (2018). When the torch is passed, does the flame still burn? Testing a "train the supervisor" model for the Child STEPs treatment program. *Journal of Consulting and Clinical Psychology, 86,* 726–737. http://dx.doi.org/10.1037/ccp0000331

Williams, N. J., & Beidas, R. S. (2019). Annual research review: The state of implementation science in child psychology and psychiatry: A review and suggestions to advance the field. *Journal of Child Psychology and Psychiatry, 60,* 430–450. http://dx.doi.org/10.1111/jcpp.12960

(*Appendix follows*)

Appendix

Data Transparency

The present article is the only report of the clinicaltrials.gov randomized clinical trial (NCT number: NCT03153904) as preregistered that has been submitted to any journal. There is a separate, student-led paper under review, based on secondary analyses of some data from this parent study, unrelated to the randomized nature of the trial. The student project used baseline reports by youths and caregivers on the problems identified as most important, in order to code the extent of youth–caregiver agreement. Baseline agreement/disagreement, thus assessed, was then used to predict multiple measures of treatment process (e.g., youths' completion of therapy homework) and two measures of clinical outcome (Child Behavior Checklist and Youth Self-Report). The only overlap in measures between the present parent study and the student project is that both included the CBCL and YSR, but the student project analyzed those measures for the entire sample as a whole without regard to the treatment condition to which participants had been randomized (i.e., the two treatment conditions were collapsed for the student project). In contrast, the present parent study analyzed those two measures, plus four other measures of clinical outcome (Brief Problem Monitor youth report, Brief Problem Monitor caregiver report, Youth Top Problems youth report, and Youth Top Problems caregiver report), two measures of treatment fidelity (MATCH adherence and competence), and the Experience in Evidence Based Practice Survey, all as part of the RCT, comparing outcomes for the two study conditions to which participants had been randomly assigned, as per the registered trial. Thus, the parent RCT (the present article) and the student-led study overlap only in their use of two measures; otherwise, the two studies differ in the measures used, their goals and study questions, their attention to randomization versus not, their methods of data analysis, and of course their findings and all implications of those findings presented in the Discussion section.

Received April 26, 2020
Revision received August 16, 2020
Accepted August 31, 2020 ■