Making Fidelity an Intramural Game: Localizing Quality Assurance Procedures to Promote Sustainability of Evidence-Based Practices in Usual Care

Aaron Hogue, The National Center on Addiction and Substance Abuse at Columbia University
Timothy J. Ozechowski, Michael S. Robbins, and Holly Barrett Waldron, Oregon Research Institute

Although the science of disseminating empirically supported behavioral treatments has made remarkable advances, the ultimate goal of dissemination—sustaining the implementation of evidence-based practices (EBPs) in usual care with a high degree of fidelity—remains challenging. This article presents a rationale and guidelines for transitioning from conventional purveyor-driven dissemination methods to intramural quality assurance procedures that can be maintained with routine agency resources. Three innovations for localizing EBP quality assurance are described: adaptation of observational fidelity methods for therapist self-report and supervisor observation of EBPs, process control benchmarking methods for continuous tracking of EBP fidelity strength and consistency, and development of intramural clinical expertise grounded in local management of EBP implementation and outcome data. These innovations exemplify a fundamentally empirical approach to sustaining quality EBP implementation in frontline settings.

Key words: dissemination and implementation science, evidence-based practices, quality assurance, sustainability, treatment fidelity, usual care. [Clin Psychol Sci Prac 20: 60–77, 2013]

The emerging discipline of dissemination and implementation science is focused on discovering efficient methods for transporting empirically supported treatments (ESTs) into everyday practice, where they can be delivered as sustainable evidence-based practices (EBPs).1 Several explanatory theories of diffusion of health interventions have been described (Gallo & Barlow, 2012; Stirman, Cits-Christoph, & DeRubeis, 2004), and an even greater number of phase-based models for behavioral health interventions (BHI) have been presented to map out the key sequences and elements of successful dissemination and highlight common roadblocks and difficult passages in moving from laboratory-developed ESTs to routinely implemented EBPs (Aarons, Hurlburt, & Horowitz, 2011; Glasgow, Vogt, & Boles, 1999; Glisson & Schoenwald, 2005; McHugh & Barlow, 2010; Simpson & Flynn, 2007). These models each have unique merits and points of emphasis, but broadly speaking, they concur in describing four basic phases of BHI dissemination: (a) Preparation: recruiting stakeholders and funding, assessing fit between intervention and host agency/system, conducting needs and barriers assessment regarding regulatory, provider, and client issues, and so forth; (b) Training: enhancing or modifying administrative and supervisory infrastructures, training supervisors and clinicians, and piloting and adapting the intervention to match site needs; (c) Implementation and Assessment:
rolling out the intervention to the target population, accommodating expert oversight on intervention delivery, and evaluating provider success in model delivery and client outcomes in targeted domains of functioning; and (d) **Sustainability**: solidifying the funding and regulatory resources, agency and supervision infrastructures, model modifications, and clinical oversight procedures required for intervention maintenance and (gradual) withdrawal of external consultation.

Although remarkable advances in the science of EST transportation have occurred in every phase of BHI dissemination, research studies (and success stories) in the sustainability phase are hardest to come by (Stroul & Manteuffel, 2007). The primary reason for this lag is self-evident: Advances in sustainability are predicated on at least moderate success in all three preceding phases. That is a lot of ground to cover, most of it new territory for host agencies and systems. Invariably, hosts initiate the transportation process by contracting with recognized experts in the given model to govern activity in all four phases; in many cases such contracts are required for proper credentialing to implement the EST. Contracted experts are typically model purveyors who either developed the model in research contexts or are licensed by developers to stimulate and manage model dissemination. This arrangement creates a confound that is unique to the sustainability phase and a second major obstacle to completing the transportation process: Hosts and purveyors are functionally reliant on one another and financially co-invested in each phase of the work, yet the end goal of dissemination is host self-sufficiency as defined by divestment or minimization of extramural involvement.

The end goal of this article is to promote the transition from expert-dependent to self-sufficient delivery of EBPs during the sustainability phase by describing methods for enhancing local quality control of model implementation. The proposed methods are designed to ensure EBPs are implemented in accordance with the main principles and procedures of the given model (Aarons et al., 2011; Damschroder & Hagedorn, 2011). In laboratory settings, such methods are usually called integrity or fidelity procedures, owing to their tight focus on the delivery of the intervention package itself in the context of controlled research in which therapist, client, and organizational variables are favorable by design. In field settings, such methods are called quality assurance (QA) procedures (Bond, Becker, & Drake, 2011; Schoenwald, 2011). As detailed in the next section, many EST developers have designed comprehensive provider training and QA procedures to facilitate model dissemination. These QA “superstructures” contain several basic components (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005; McHugh & Barlow, 2010; Schoenwald, Sheidow, & Chapman, 2009): (a) standardized training toolkits that include a treatment manual, protocol for training workshops, demonstration videos and clinician workbooks, on-site supervision procedures, and fidelity checklists; (b) procedures for ongoing training and consultation from model experts, often including observational coaching of sessions via audio- or videotape; (c) quality improvement (QI) procedures to evaluate implementation data collected on-site, feed selected data back to therapists, and buttress organizational support; and (d) certifications granted to providers who successfully complete training and maintain quality standards.

This article presents innovative methods to support the transition from extramural QA structures to intramural QA procedures that are capable of sustaining EBP fidelity—competent delivery of prescribed interventions and avoidance of proscribed interventions for target populations (Perepletchikova, 2011)—in frontline settings. The first section reviews purveyor-driven QA structures for EBPs in usual care, outlining a continuum of intensity of involvement with model developers. The second section outlines the rationale for developing locally driven QA methods to complement (and eventually minimize) extramural stewardship in the sustainability phase. The third section details three specific methods for localizing QA: adaptation of observational fidelity technology, process control benchmarking (PCB), and development of intramural clinical expertise.

**THE INTENSITY CONTINUUM OF EXTRAMURAL QA PROCEDURES FOR SUSTAINING EBP FIDELITY**

Quality assurance procedures for supporting the transportation and maintenance of ESTs in usual care are crafted to fit the particular therapeutic content, administrative and infrastructure requirements, and extramural...
consultation needs of the given EST—as such, there are as many different QA structures as there are transport-ready ESTs. To exemplify a few basic components and configurations of existing QA structures, we describe QA procedures for two treatment models that each boast a wealth of efficacy and effectiveness data and are widely implemented in frontline settings: functional family therapy (FFT) and motivational interviewing (MI). These two models also represent opposite ends of a continuum of intensity with regard to involvement and interaction between purveyor organizations and local providers: FFT procedures include structured multilevel involvement with a centralized network of local providers governed by a single purveyor organization, whereas MI procedures are flexibly managed by a diffuse network of trained individuals.

Functional family therapy (Alexander, Pugh, & Parsons, 1998) is an integrative family-based EST for adolescent conduct problems and substance use (Alexander, Robbins, & Sexton, 2000; Waldron, Slesnick, Brody, Turner, & Peterson, 2001; Waldron & Turner, 2008). FFT is widely disseminated by FFT, Inc. (www.fftinc.com), which oversees FFT implementation in over 250 local, state, national, and international organizations in 44 U.S. states and five other countries, collectively providing services to over 12,000 families annually. FFT, Inc. has developed a comprehensive training and dissemination protocol that is implemented within provider organizations contracting with FFT, Inc. (or an affiliated “network partner”) to become licensed FFT implementation sites. The transportation of FFT to an individual site begins with establishing buy-in among agency administration and key stakeholders in the community, starting with a formal application process that highlights the expectations of implementation sites (such as staffing, team size, case-loads, etc.). This application information provides a foundation for initiating an ongoing dialogue with prospective sites via in-person visits and phone conferences with FFT, Inc. that is critical for developing or maintaining funding streams, establishing appropriate referral mechanisms, facilitating communication plans, identifying site administrative and clinical staff to be involved in transportation, and implementing the training program. The goal of this preparation work is to confirm site readiness, identify any critical factors impacting model implementation, and build necessary community/stakeholder understanding of implementation challenges.

FFT, Inc. training and implementation monitoring is organized in three structured phases. Phase I is dedicated to establishing a core FFT implementation team (designated supervisor(s) and clinicians) and systematically building therapist competence in delivering FFT to youth and families. Training consists of five workshops, delivered to FFT therapists over the course of the first year, that include a mix of didactic, experiential (role play), tape review, and case review, followed by weekly group (and individual, when possible) consultation sessions. The workshops and weekly consultation sessions are led by a certified expert trainer from FFT, Inc. Also during Phase I, FFT, Inc. initiates the process of supporting site independence by identifying (in collaboration with the site) an individual from the designated FFT team who will take over consultation as the on-site supervisor in Phase II and beyond. The on-site supervisor participates in an intensive externship experience at the end of Phase I at a national training hub that involves live case consultation during three 3-day visits over a three-month period. During Phase II, the on-site FFT supervisor attends two 2-day supervisor trainings and is then supported by FFT, Inc. via monthly phone consultation. During this phase, the on-site supervisor is also trained in a sophisticated Web-based QA application, the client services system (CSS), to monitor therapist fidelity and client outcomes. The CSS is described below in the “Intramural Expertise Development” section below. Finally, in Phase III, FFT, Inc. provides only minimal direct site consultation in the form of monthly calls focused on administrative and data review as well as relevant clinical challenges raised by the on-site supervisor.

The developers of the MI model have adopted a markedly different approach to model dissemination. MI is a client-centered therapeutic approach designed to strengthen clients’ motivation and commitment to change and to support client progress toward achieving desired goals (Miller & Rollnick, 2002). MI transportation into community sites is facilitated by the Motivational Interviewing Network of Trainers (MINT; http://www.motivationalinterviewing.org). MINT is a nonprofit organization whose membership is composed
of between 1,000 and 1,500 clinicians throughout the world who have been credentialed as trainers in MI by virtue of successfully completing a three-day “training of new trainers” (TNT) workshop sponsored by MINT. Once credentialed, MI trainers may be hired by individual clinicians or service organization seeking to learn and adopt MI. The amount and format of MI training are negotiated on a case-by-case basis according to the needs of contracting individuals or organizations. MI training typically includes one or more of the following components: (a) a one-day introductory workshop designed to familiarize clinicians with fundamental MI principles and practices; (b) a two- to three-day basic training workshop involving a mixture of didactic presentation, clinical demonstration, and practice exercises; and (c) a two- to three-day advanced training workshop involving feedback on recorded samples of trainees’ clinical work and more clinically intricate demonstrations and practice exercises. In addition to these workshops, MI trainers may be contracted to provide booster training, ongoing case consultation, and clinical supervision. MI trainers may also provide instruction for reviewing and coding recorded treatment sessions for adherence to the MI model.

Motivational interviewing trainers are not formally monitored or evaluated by MINT beyond completion of the TNT workshop. Consequently, MI trainers may exercise considerable leeway and flexibility regarding the amount, content, and format of the training they provide to those seeking to learn and adopt MI. Moreover, in keeping with the MI philosophy, the MI model and training process may be locally customized for various clinical and cultural contexts (Miller & Rollnick, 2002). The principal advantage of this relatively flexible and loosely structured approach to model transportation is that adopting MI is administratively, clinically, and financially feasible for most settings. However, a primary drawback of the MI approach is lack of quality control and evidence of sustained effectiveness. For example, a randomized trial of MI training methods by Miller, Yahne, Moyers, Martinez, and Pirritano (2004) indicated that licensed therapists in community-based settings receiving a two-day MI training workshop followed by four months of feedback and coaching achieved and maintained MI proficiency standards up to 12 months posttraining, whereas gains in MI proficiency tended to erode by four months posttraining among therapists receiving only workshop instruction. In contrast to these results with experienced therapists, studies of therapists with less experience and credentials indicate that MI training methods appear to have little effect on long-term skill development and proficiency (Mitcheson, Bhavsar, & McCambridge, 2009; Moyers et al., 2008), although supplemental coaching and supervision may be effective for those practitioners not achieving required skill levels after initial rounds of training (Beidas, Edmunds, Marcus, & Kendall, 2012; Martino et al., 2011).

Both ends of the QA continuum present benefits and drawbacks with regard to local capacity to sustain EBPs. At the higher-intensity end, for EBPs that require more centralized and regimented QA structures, sustainability is reliably grounded in adherence to standardized QA procedures and long-term dependence on purveyor involvement. Yet, this fundamentally extramural approach may be feasible only for high-capacity and well-funded provider organizations and networks (Hogue, 2010). In the long run intensive QA structures may thrive principally in government-operated sectors of care in which clients are especially high risk, services are high cost, and stakes for success are high profile: criminal justice, juvenile justice, welfare, child welfare, homeless services, and so forth (Chambers, Ringeisen, & Hickman, 2005; Zazzali et al., 2008). At the lower-intensity end, for EBPs with more decentralized and flexible QA structures, sustainability is nurtured via development of an organically tailored fit with the organizational and resource capacity of host agencies. Yet, this largely intramural approach is susceptible to erosion of clinical skills and lapses in model fidelity acquired during initial training in the absence of systematic and ongoing training, supervision, and quality control (Aarons, Sommerfeld, Hecht, Silovsky, & Chaffin, 2009; McHugh & Barlow, 2010).

As discussed in the following section, the best means to achieve EBP sustainability across the full range of existing (and needy) service settings may be a balanced approach that utilizes purveyor-facilitated QA processes and mechanisms imported during the implementation phase to systematically build local capacity for model-based supervision, adherence monitoring, and quality control during the sustainability phase. BHI dis-
semination models of this kind—in which building local capacity for sustaining purveyor-developed QA procedures is a primary dissemination objective—are beginning to emerge (Chamberlain et al., 2012). Yet, very few specific guidelines and mechanisms for building local capacity to sustain EBPs have been offered, and there are virtually no empirical data to lead the way.

RATIONALE FOR ENHANCING THE SUSTAINABILITY OF EBP FIDELITY VIA INTRAMURAL QA

All stakeholders in BHI dissemination—purveyors, government systems, providers, clients—stand to profit in multiple ways from increasing the sustainability of EBPs in routine care. Unfortunately, despite noteworthy achievements for several models during the first three phases of dissemination (Henggeler & Sheidow, 2012), success in the sustainability phase has been elusive. Three specific strategies for achieving sustainability are commonly endorsed: continued intensive involvement with purveyors, train-the-trainer procedures, and peer consultation among providers. Whereas each of these strategies holds promise as an end-phase solution, each also faces significant challenges in frontline settings, and none has yet generated convincing evidence of long-term effectiveness (Aarons et al., 2011; McHugh & Barlow, 2010).

As discussed above, the primary obstacle to ongoing involvement in purveyor-driven QA structures is the cost incurred by continuous investment in material resources and technical support for multiple QA components. High therapist turnover in agency settings also creates costly demand for reenactment of training phase elements for new line staff (Knudsen, Ducharme, & Roman, 2008). And there are ample obstacles to purveyor involvement beyond cost. Sustainability that relies chiefly on extramural methods is vulnerable to vicissitudes in funding and regulatory practices, decrease in the availability or commitment of purveyors, decrease in the provider stamina to honor contractual QA procedures for an extended period, and demoralization among frontline staff if external agents are responsible for continuous judgments about clinical competence and EBP adaptation (Aarons et al., 2011; Gallo & Barlow, 2012; Simpson & Flynn, 2007), although staff fatigue may be counterbalanced to a large degree by the additional EST training and professional enrichment offered by such arrangements (Aarons et al., 2009).

The other two sustainability strategies face equally stiff challenges. Train-the-trainer procedures aim to engender EBP sustainability by judiciously selecting and training in-house supervisors to become experts in implementing and monitoring the given treatment model, whereupon extramural support is reduced in favor of greater QA responsibility for the trainer. Evidence for this strategy has been mixed. For example, Forgatch and DeGarmo (2011) reported an uneven “training cascade” effect in transporting a parenting skills EST: Strong fidelity was achieved among cohort-1 providers who were trained by purveyors, followed by a significant fidelity decline among cohort-2 providers trained by cohort-1, followed by return to strong fidelity among cohort-3 providers trained by cohorts 1 and 2. Martino and colleagues (Martino et al., 2011) found that MI implementation and outcome effects degraded or disappeared once external support was fully withdrawn from the train-the-trainer study condition. Also, trainer turnover remains a prevalent and costly problem across the board. Finally, the peer consultation strategy, wherein a regional coalition of sites trained in a given model collaborates to share QA resources (Chamberlain et al., 2012), has not yet provided a reliable blueprint for creating the efficiencies of coordination, cost, and scale necessary to sustain EBPs (McHugh & Barlow, 2010). Foremost among the sustainability challenges for these two strategies is “fidelity drift”: the tendency for EBP fidelity levels achieved during the training and implementation phases to fade substantially during the sustainability phase as expert-led QA is reduced or withdrawn (McHugh & Barlow, 2010). Overall, because sustainability studies are few and success stories fewer, there are virtually no empirical guidelines that stipulate how much time and financial investment is required to prevent fidelity drift and conserve gains accrued during training and implementation phases, how many resources are needed to sponsor a “good enough” intramural QA system, or whether for more complex ESTs it is advisable or even possible to graduate from outside technical support to localized QA.

The section that follows introduces three methods for facilitating greater local control of QA procedures
during the sustainability phase. The methods are designed to aid hosts in attaining a more fluid balance between external and internal QA management, ideally one that is both structured (procedures-based) and flexible to meet the clinical and capacity needs of given providers. To achieve such balance, it seems fair to say that significantly more ballast is needed on the intramural side.

THREE PROPOSED INNOVATIONS IN INTRAMURAL QA PROCEDURES TO PROMOTE SUSTAINABILITY

In a forerunning article on sustaining EBPs in everyday settings, Altman (1995) observed that “sustaining interventions requires that communities assume ownership and control over programs. Assuming ownership and control requires the development of local expertise and only selected reliance on outside experts” (p. 529). Nearly two decades later, tools for cultivating local expertise in sustaining EBPs remain underdeveloped and understudied (Sorensen & Kosten, 2011). In this article, we describe three methodological “innovations” to serve this purpose. We view these methods as innovations in the sense that they entail novel and fundamental shifts in standard operating practices within most local treatment settings. Altman (1995) emphasized that sustaining interventions necessarily entails “overcoming the status quo” within clinical environments, that is, propelling through administrative, clinical, and cultural inertia to embrace new practices that foster the adoption and long-term maintenance of EBPs. The three empirically grounded methods can advance these efforts within almost any type of agency, including small or stand-alone clinics laboring to adopt ESTs independently of any model purveyor.

Observational Fidelity Methods

The logical first step in shifting the balance from outside expertise toward intramural QA is developing EBP fidelity assessment tools that are not only reliable and clinically valid but also locally manageable (Schoenwald, 2011). In research settings, the most rigorous method for assessing intervention fidelity is observational rating by nonparticipant judges (Hogue, Liddle, & Rowe, 1996; Mowbray, Holter, Teague, & Bybee, 2003). Although time and resource intensive (Hill, 1991), observational methods can produce objective and highly specific data on how clinicians perform in session (Hill, O’Grady, & Elkin, 1992; Hogue et al., 1998). Top-notch observational fidelity measures usually contain four design features (Carroll et al., 2000; Hogue, 2002; Waltz, Addis, Koerner, & Jacobson, 1993): (a) quantitative metrics that assess the intensity and frequency of interventions; (b) specification of both model-specific interventions essential to the underlying clinical theory and common therapeutic elements endorsed by most BHI models (e.g., working alliance); (c) attention to intervention quality (competence) as well as quantity (adherence); and (d) attention to both therapist and client contributions to the treatment process.

What if research-based observational fidelity methods could themselves be transported into agency settings? There is growing enthusiasm for using observational EST measures as a foundation for developing QA tools in routine care (Henggeler, Sheidow, Cunningham, Donohue, & Ford, 2008; Weisz, Ugueto, Herren, Afienko, & Rutt, 2011), on the premise that adapting these methods is likely to yield instruments with both strong psychometric properties and high specificity in measuring EBP techniques (Schoenwald, 2011). However, EBP assessment tools intended for frontline clinics may require some design compromises in the service of a broader clinical relevance (Kolko, 2006; Schoenwald et al., 2010): applicability to heterogeneous client populations with diverse profiles of co-occurring problems; ability to capture a wide variety of treatment practices rather than a narrow band of prespecified (manualized) techniques; and flexibility in marking duration, ancillary interventions, and organizational features of treatment. Overall, the most effective QA tools for assessing usual care interventions will balance measurement precision with breadth of interventions assessed, evaluate fidelity to both the content and processes of EBP implementation, and accommodate multiple uses and yield data that inform multiple levels and standards of accountability (Bearsley-Smith, Sellick, Chesters, & Francis, 2008; Schoenwald, 2011; Schoenwald et al., 2010). Below, we detail two promising methods for adapting observational fidelity methods to support quality EBP implementation across three levels of QA monitoring: (a) therapist self-report measures and (b) supervisor observation techniques.
Therapist Self-Report. Self-report measures of EBP fidelity offer several methodological strengths that may afford the desired balance between rigor and relevance in QA systems: They are quick, inexpensive, and non-intrusive; they can be completed throughout treatment, which facilitates measurement of infrequent but clinically meaningful interventions; and they can assess therapist intentions as well as observed behaviors (Carroll, Nich, & Rounsaville, 1998; Weersing, Weisz, & Donenberg, 2002). If proven reliable and valid, therapist-report measures could support EBP sustainability and inform QI via feedback loops of several kinds: as a self-check by therapists to mark their own progress in treating individual cases; as a supervision aid for on-site and external trainers to monitor treatment fidelity; and as administrative data for stakeholders and regulators to evaluate therapist- and agency-level clinical performance (Bearsley-Smith et al., 2008; Bond et al., 2011; Garland, Bickman, & Chorpita, 2010), among others. As an example, therapist reports of adherence to multisystemic therapy for antisocial youth have demonstrated strong construct and predictive validity and been linked to various client, therapist, and organizational factors (Schoenwald, Letourneau, & Halliday-Boykins, 2005), including ongoing expert consultation (Schoenwald, Sheidow, & Letourneau, 2004).

Unfortunately, the handful of studies testing the validity of therapist-reported EBP implementation via comparison with observational ratings has produced disappointing results (Hurlburt, Garland, Nguyen, & Brookman-Frazee, 2010; Miller et al., 2004; Schoenwald et al., 2009). Even studies that directly retrofitted observational EST fidelity measures for use as self-report instruments have logged modest to weak correspondence between the new measure and its original version (Carroll et al., 1998, 2009). Nevertheless, there remains hope for salvaging the validity, and thus the utility, of therapist-reported EBP implementation: Train and monitor frontline clinicians to be fluent in self-report fidelity ratings using procedures analogous to those successfully used with observational coders (Hurlburt et al., 2010). Succinctly put, these procedures consist of (a) adapting and piloting an existing observational measure and its companion coding manual to capture the desired EBPs as items on a self-report scale; (b) training clinicians to use the adapted self-report scale during weekly sessions over two to three months that include didactic instruction and group discussion of scale items, review of practice tapes carrying gold standard ratings for each item made by knowledgeable experts, and coding exercises designed to test and expand item knowledge; and (c) continuous monitoring of self-report data via peer review of coded sessions, with the possibility of tracking the self-report reliability of therapists against observations made by their own supervisors (see below section), and subsequently retraining for any items and/or clinicians who evidence weak or declining reliability (Hill, 1991).

Is such a rigorous approach feasible in everyday practice settings? There are reasons to believe so. First, EST purveyors with existing observational measures are primed to fashion appropriate self-report tools and training procedures, and the initial training of line clinicians in self-report reliability appears to fit snugly and perhaps cost-effectively within the broader goals of the training and implementation phases of dissemination (see below section on promoting the feasibility of observational fidelity methods in usual care). Second, it may not be prohibitively difficult to train community therapists to be reliable self-report fidelity coders, given that several observational studies have employed practicing therapists as coders (Hogue et al., 2008). Third, as described above, existing QA structures already require therapists to provide self-report data of various kinds, setting in place the procedures and media needed to collect and analyze self-report fidelity data. An important caveat is that therapist-report measures cannot legitimately capture the competence of model implementation, though treatment adherence data themselves play a pivotal role in broader judgments about service quality.

Finally, to strengthen the feasibility and utility of EBP self-reports in usual care (Fals-Stewart & Birchler, 2002), prospective fidelity tools should (a) be designed as postsession rather than posttermination measures (Weersing et al., 2002) to maximize reliability and acuity; (b) assess discrete intervention techniques (aka clinical strategies or practice elements), which is middle ground between molar versus molecular specification of treatment processes and hence well suited for describing eclectic clinical practices in routine care (Garland, Hurlburt, Brookman-Frazee, Taylor, &
Accurso, 2010); and (c) address the “contours” of EBP implementation (Schoenwald, 2011), including treatment parameters (service delivery aspects of implementation: to whom, where, and how often) and prescribed content (therapeutic focus or domains of functioning) within and across sessions (Hogue, Liddle, Dauber, & Samuolis, 2004; Hurlburt et al., 2010; Kelley, de Andrade, Sheffer, & Bickman, 2010).

Supervisor Observation. There is virtual consensus in the EST training and dissemination literature that ongoing expert clinical supervision of therapists is essential for effective and sustainable EST transportation in community settings (Beidas & Kendall, 2010; Schoenwald et al., 2009; Stirman et al., 2010). Unfortunately, in many community agencies, clinical supervision is either minimal or nonexistent due to a range of organizational and financial barriers (Amodeo, Storti, & Larson, 2010; McLellan & Meyers, 2004). Shortages of clinical practitioners and supervisors are particularly acute in rural areas (Jameson, Blank, & Chambless, 2009). The inability to recruit and retain highly qualified and experienced clinicians to fill supervisory roles is a prime factor in agencies’ inability to adopt ESTs (Massatti, Sweeney, Panzano, & Roth, 2008).

Fortunately, many historical barriers to the availability of on-site clinical supervision can be overcome presently given the penetration of Web-based video conferencing (WVC) within the delivery of psychotherapy services (Richardson, Christopher Frueh, Grubaugh, Egede, & Elhai, 2009). In addition to expanding the reach of clinical services to difficult-to-access populations, WVC is increasingly used for training, consultation, and supervision (Rees, Krabbe, & Monaghan, 2009; Reese et al., 2009; Weingardt, Cucciare, Bellotti, & Lai, 2009; Xavier, Shepherd, & Goldstein, 2007). With regard to sustaining EBPs, agencies lacking in-house supervisors experienced in a given treatment model can seek out such personnel externally and contract with them to provide clinical supervision remotely via WVC. The equipment required for this is minimal (computer with webcam, high-speed Internet connection, video conferencing software), widely available, and affordable in retail. Moreover, the cost of contracting for several hours per week from an off-site expert supervisor is far less than the expense of recruiting and hiring such personnel into part- or full-time employed positions.

Beyond mere availability, clinical supervision in sites endeavoring to sustain newly imported ESTs should entail a significant amount of direct observation of the clinical work. In many agency settings, clinical supervision is composed entirely of case consultation based on therapist descriptions of the given case. Unfortunately, case reports from therapists are prone to be filtered, selective, and may omit or misrepresent critical aspects of the therapist’s work (Stirman et al., 2010). For agencies in which supervisory case consultation is the status quo, clinical supervision based on direct observation will constitute a monumental shift in both practice and paradigm. Nevertheless, this shift is a virtual requirement for sustainability based on evidence from technology transfer studies of all kinds indicating that performance feedback and coaching are essential for adopting ESTs with high fidelity in usual care (Fixsen et al., 2005; Miller, Sorensen, Selzer, & Brigham, 2006). Performance feedback and coaching for novice and experienced clinicians alike can be made more precise, incisive, applicable, and effective if such guidance is informed by direct observation rather than post hoc anecdotal reports of clinical activities (Sheidow, Donohue, Hill, Henggeler, & Ford, 2008; Stirman et al., 2010). Encouragingly, recording of therapy sessions for supervisory purposes is dramatically more feasible than ever given advances in digital recording and computer software that can promptly download, store, and share audio or video files via secure encrypted transmission. Direct-observation supervision delivered in-person or via WVC is now becoming commonplace in dissemination work (Miller et al., 2004; Sholomskas et al., 2005; Weingardt et al., 2009).

Promoting the Feasibility of Observational Fidelity Methods in Usual Care: Finding Time. As described above, advances in several areas have primed the field of dissemination science to promote the use of observational fidelity methods in everyday settings: validated observational measures of EST fidelity, EST training protocols that emphasize treatment fidelity, QA requirements for therapist reports of EBP delivery, and fidelity supervision guidelines and session recording technology. But even accepting that the technology is
available and the costs reasonable, where to find (non-billable) time to engage in fidelity-promoting client observation? One possibility is to include specific procedures within initial EST training protocols that instruct therapists on how to assign self-ratings of their own model adherence in a manner that matches how EST trainers evaluate trainee performance via observational methods. Laying this groundwork in the training phase will provide the foundation for regular fidelity consistency checks between therapists and local supervisors in the sustainability phase, ideally carried out on a monthly basis during routine supervision hours. Fidelity consistency checks could take the form of (a) comparing archived therapist ratings to supervisor ratings of a common session and (b) therapist and supervisor together reviewing a taped session segment during supervision hours and conferring on fidelity ratings, or some combination. In this manner, therapist and supervisor ratings of EBP fidelity could be continuously recalibrated via case observation with no additional time required and minimal sacrifice of standard case supervision.

Process Control Benchmarking

Benchmarking is a method for assessing the fidelity and effectiveness of BHI in routine clinical settings. Typically, benchmark methods are employed to investigate whether community therapists delivering ESTs in everyday practice can approximate the performance standards set by their research-funded counterparts in controlled trials (Addis, 1997; Nathan, Stuart, & Dolan, 2000; Weisz, Han, Granger, Weiss, & Morton, 1995), focusing on critical areas such as client retention, treatment implementation, and clinical outcomes (Hunsley & Lee, 2007; Weersing, 2005). Benchmarks can also be derived from a variety of recognized sources such as local or national performance criteria (Weersing, 2005), national data warehouses (Mellor-Clark, Barkham, Mothersole, McInnes, & Evans, 2006; Mullin, Barkham, Mothersole, Bewick, & Kinder, 2006), or clinical research studies aggregated via quantitative review (Chorpita et al., 2002; Weersing & Weisz, 2002) or meta-analyses (Minami, Wampold, Serlin, Kircher, & Brown, 2007). By and large, benchmarking studies have reported that ESTs exported to community sites via expert-guided training achieved outcomes that were roughly comparable to those produced in research settings (Barkham et al., 1996; Hill, 1991; McEvoy & Nathan, 2007; Merrill, Tollbert, & Wade, 2003; Tuschen-Caffier, Pook, & Frank, 2001; Wade, Treat, & Stuart, 1998; Weersing & Weisz, 2002), although effects in community sites may be less durable over time (Barkham et al., 1996; Henggeler, Brondino, Melton, Scherer, & Hanley, 1997; Henggeler, Pickrel, & Brondino, 1999).

The inclusion of benchmarking analyses in QA systems can provide a valuable perspective on whether EBPs are potent and sustainable in the real world by depicting the strength and consistency of fidelity effects in host agencies (Gaston, Abbott, Rapee, & Neary, 2006; Minami et al., 2007). An efficient and flexible tool for conducting fidelity benchmarking analysis is statistical process control (SPC) analysis. SPC is a method developed in industrial psychology for monitoring the amount of variability in a continuous production process (Deming, 1986). SPC has been used in a few BHI studies to measure consistency across sites in delivering treatment services (Green, 1999), analyze variations over time in clinical staff performance (Dey, Sluyter, & Keating, 1994; Pfadt, Cohen, Sudhalter, Romanczyk, & Wheeler, 1992), and examine consistency across clients in implementing prescribed interventions in residential drug treatment (Faw, Hogue, & Liddle, 2005).

Although not widely used in BHI research, SPC appears well suited for benchmarking EBP implementation in usual care. Most benchmarking studies have relied on primarily descriptive analyses that juxtapose effect sizes or percentages of improved outcomes in community versus criterion samples (Franklin, Abramowitz, Kozak, Levitt, & Foa, 2000; Tuschen-Caffier et al., 2001; Wade et al., 1998). Other analytic techniques favored in benchmarking research include equivalence testing (Fals-Stewart & Birchler, 2002; Self-Brown et al., 2012), inference testing using reliable change indices (McEvoy & Nathan, 2007), and meta-analysis techniques that test effect size differences across samples using a priori “good enough” critical values (Hill, 1991; Knudsen et al., 2008). In contrast, SPC employs probability sampling procedures (Weersing & Weisz, 2002) in which continuous samples are taken from the process under investigation and plotted on a
control chart that contains upper and lower control limits based on either prespecified criterion values or the distributive properties of the given sample. Plotted data points are inspected to determine whether an “out of control” pattern emerges that indicates a systematic influence or change in the production process (e.g., nine consecutive data points on the same side of the mean); these control criteria have evolved in industrial settings to identify patterns unlikely to occur when a given process obeys a routine, expectable level of variance (Hoyer & Ellis, 1996).

Applications of SPC for monitoring and evaluating EBP implementation in real-world settings—called PCB—are straightforward and exciting. PCB methods allow tracking of both fidelity strength (via analysis of means) and fidelity consistency (via analysis of variances) within the target group. Because it is designed to distinguish between normal variation and systematic uncontrolled variation, PCB may be ideal for supporting line clinicians who are responsible for consistent delivery of prescribed interventions yet also expected to show natural variation in implementation across sessions and caseloads. PCB methods can be readily merged into clinical data management systems used to guide decision making about who gets what treatment, how well treatment is progressing, and what corrective action is needed to improve case progress (Chorpita, Bernstein, & Daleiden, 2008). For example, PCB data could be principal information transmitted in clinician feedback loops, which provide regular appraisals of individual client progress on standardized indicators to therapists and supervisors during treatment (Godley, Garner, Smith, Meyers, & Godley, 2011; Whipple et al., 2003). Clinician feedback loops have demonstrated excellent utility for preventing early treatment failure, reversing symptom deterioration, and enhancing general outcomes (Harmon et al., 2007; Shimokawa, Lambert, & Smart, 2010). And because SPC is a flexible and user-friendly technology, it can serve the nuts-and-bolts data management needs of PCB for EBP fidelity data collected during routine QA procedures in agencies with even rudimentary computing resources.

Process control benchmarking can also provide unique insight into links between EBP implementation and outcome. For example, if PCB analyses showed that community therapists placed well outside research-based control limits for model fidelity, this would suggest that the best means to improve outcomes is to enhance EBP fidelity levels by strengthening training and QA protocols. On the other hand, if therapists were basically adherent to fidelity control limits but still missed the mark in outcome success, this may direct efforts toward revising the treatment model itself to improve its potency in community settings. In these ways, PCB can produce a call to action for shaping QA and QI procedures based on site fidelity data (Bond et al., 2011), informed in equal measure by provider consistency in reaching fidelity benchmarks and observed correlations between fidelity and outcomes.

Finally, PCB methods are a perfect fit for designing “fidelity drift alarms” that can be activated during the sustainability phase as the balance shifts from external to internal QA monitoring. Because many ESTs have a sufficient research base to establish implementation benchmarks, drift alarms can be created to signal deviations from fidelity control limits, which would initiate intramural self-correction mechanisms and alert purveyors to a possible need for interstitial training and QA boosts. Alarms could be calibrated to model-specific adherence and competence control limits and then fixed at a single general setting or at different settings for different phases of treatment. Alarm calibrations would vary over time based on the expanding research portfolio of the original EST as well as grassroots experience in fidelity benchmarking with community providers. Also, whereas alarm thresholds and controls would usually be set by purveyors (Glisson & Schoenwald, 2005), systems for monitoring drift alarms and providing corrective feedback to line staff could be fully managed by host agencies using inexpensive and widely available technology (for similar ideas regarding local QA technology, see Chorpita et al., 2008; Wengardt, 2004).

Intramural Expertise Development
Success in the sustainability phase is predicated on success in all previous phases. During the preparation phase, establishing agency buy-in, funding streams, appropriate referral mechanisms, intra- and extramural communication plans, and staff selection processes are all necessary precursors for facilitating a context of
empowerment and ownership at the host site and surrounding community that is critically important to sustaining EBPs. Similarly, the clinical foundation for self-sufficiency is laid early in the training phase by identifying a staff member with the nascent skills and motivation to become an on-site supervisor of the EBP. The multistep process of supervisor development plays out across the training and implementation phases, typically entailing an intensive training and/or externship experience to enhance clinical skills in the model, weekly (titrated to monthly) phone consultation from model experts on model implementation and (gradually) model supervision, and increased supervisor responsibilities in proportion to decreased purveyor involvement. Finally, as the sustainability phase extends in time, the goal is for direct consultation from purveyors to become progressively focused on review of clinical and administrative data generated by the site, with modest clinical support and booster training for on-site supervisors as needed. To illustrate purveyor-designed data management and review procedures that support localization of EBP fidelity during the sustainability phase, we again draw on the example of FFT.

**FFT Quality Assurance/Quality Improvement.** As mentioned above, EST purveyors develop comprehensive procedures for tracking clinical processes and outcomes throughout the dissemination process with the intent of improving performance at every level of implementation: therapist adherence/competence, supervisor skills and training, organizational support, and relationships with referral sources and community stakeholders. Aspects of implementation are continuously monitored with respect to performance benchmarks defined by respective ESTs (QA), and performance data are used to formulate and implement plans to systematically improve or shape the implementation process (QI). As such, the collection and utilization of various types of performance data—including client, therapist, and supervisor reports—is a key aspect of the QA/QI process and thereby fundamental to sustaining the EST model with fidelity in community settings.

FFT, Inc. maintains a multifaceted Web-based application, the CSS, to monitor clinical implementation and client outcomes. The CSS is designed to monitor highly structured FFT therapist progress notes as well as supervisor and client ratings of therapist adherence and competence. The monitoring process helps maximize sustainability for community programs by limiting costs, at the same time building local clinical management resources by infusing the on-site supervision process with direct feedback to therapists and supervisors based on progress notes and fidelity ratings. The CSS also allows therapists to track the modalities essential for successful implementation: session process goals, comprehensive client assessments, and clinical outcomes. The computer-based format provides easy access to a wide variety of performance data to inform clinical decisions and evaluate case success. In addition, several CSS reports have been developed to provide host agencies with QA/QI reviews, therapist development planning and tracking, and supervision planning.

Initially, the expert consultant from FFT, Inc. assumes responsibility for generating and reviewing reports, and developing and implementing QI plans. However, in the sustainability phase, this responsibility shifts to the on-site supervisor. In these latter stages, the role of FFT, Inc. is to provide guidance to the local supervisor on maintaining the quality of model implementation, sustaining the integrity of the FFT clinical team, and communicating with stakeholders (within and outside the agency) about team performance. It is important to note that the reliability and validity of the CSS, including the therapist- and supervisor-report data that comprise fidelity reports, have not yet been empirically demonstrated. Nevertheless, the CSS represents an intriguing model of fidelity data management that can (a) readily incorporate the observational fidelity and benchmarking methods described above and (b) facilitate a judicious transfer of fidelity oversight from purveyor to host agency.

**Flexibility in FFT Localization Strategies.** Standard FFT, Inc. training procedures institute a stepped process of decentralization over time, wherein sites become increasingly responsible for supervision and QA/QI activities. However, FFT, Inc. has also been flexible in working with sites and systems to develop localization strategies that more adequately match the diverse organizational features and funding realities of
particular state-level systems. For example, in Florida, FFT, Inc. has partnered with Evidence-Based Associates (EBA) as part of the Department of Juvenile Justice’s Florida Redirections Project. In this project, EBA functions in a primary leadership role in directing the implementation of multiple ESTs across the state, including 10 FFT teams. EBA is directly involved in facilitating funding, linking providers with purveyors, developing consistent and unique QA procedures and targets across models, and monitoring performance data. In this project, FFT, Inc. retains direct oversight of training and QA/QI for FFT teams during the training phase, with sites assuming increased responsibility for these activities in the implementation phase and beyond. Additional external funding supports FFT, Inc. in providing ongoing QA/QI leadership for all sites, irrespective of dissemination phase.

A very different strategy is being used in California, where FFT, Inc. partners with the California Institute of Mental Health (CIMH) to provide training, consultation, and QA/QI to approximately 40 FFT teams throughout the state. Following the community development team model (Chamberlain et al., 2012), CIMH functions in a centralized role at multiple levels, including supporting state-level consultants in later dissemination phases to perform fidelity monitoring functions typically provided by FFT, Inc. experts. CIMH also facilitates local training for new teams and replacement trainings for new therapists throughout the state to reduce training and travel costs for sites that are newly starting or dealing with attrition. More recently, CIMH has arranged partnerships between host providers and government systems (e.g., juvenile probation) to initiate locally housed, state-funded FFT training externships to further reduce costs and localize training opportunities.

Finally, in areas where multiple ESTs are purveyed in the same communities, there is increased recognition of the need to develop cross-model QA/QI procedures for sustaining fidelity to the various ESTs, while also adhering to a consistent set of local monitoring and reporting standards. Developing cross-model QA/QI is no easy task, but increasingly purveyors are working cooperatively with providers and one another to minimize the duplication of paperwork across multiple reporting systems and document model-specific implementation in a manner that fits comfortably with local auditing. Although these types of community-centered cross-model collaborations are in their infancy, host providers and larger systems (such as EBA and CIMH) are playing a critical role in pushing for clarity and consistency in EBP implementation monitoring to facilitate their long-term sustainability.

CONCLUSION: ALIGNING EMPIRICALLY BASED PRINCIPLES AND PROCEDURES TO SUSTAIN EBPS IN LOCAL SETTINGS

An important subtext woven throughout this article, and others in this special issue, is the ongoing maturation of the science of EST dissemination and implementation in the field of clinical psychology. A variety of models, approaches, and pathways toward EST dissemination have evolved over the past decade as part and parcel of the greening of the technology transfer landscape in this area. This article attempts to contribute to this growth by discussing instrumental shifts in operating practices and procedures that can fortify newly adopted EBPs and insulate against protocol drift and erosion that naturally occur in real-world settings (McHugh & Barlow, 2010). Our main procedural recommendations—observational fidelity monitoring, PCB, and intramural expertise development—are by no means a panacea for achieving EBP sustainability. Rather, they are intended to be facilitative components within a larger, locally applied dissemination framework encompassing multiple tiers of coordinated implementation activity (Aarons et al., 2011; Durlak & DuPre, 2008; Torrey, Bond, McHugo, & Swain, 2012). Moreover, these recommendations are intended to be relevant and applicable in any clinical setting regardless of the particular route taken toward EST adoption and sustainability (i.e., purveyor-driven transportation or internally driven and locally implemented adoption).

The sustainability-enhancing practices described above are not specific to any particular dissemination model or approach per se, but instead pertain to adopting a dissemination “mind-set” that espouses an empirical frame of reference—that is, observational and data-based—for service delivery, evaluation, and quality control. Although it is likely that many providers are favorably disposed toward an empirical mind-set, day-to-day operations in community settings
may not reflect this orientation. The innovations discussed in this article are aimed at making standard operating practices more consonant with empirical values and principles, thereby “enriching the soil” and providing fertile ground within which newly adopted EBPs can take root, thrive, and endure (Stirman et al., 2010).

NOTE
1. In this article, we use the term “empirically supported treatment” to refer to behavioral treatment models that are implemented or systematically trained and monitored by model developers, clinical researchers, or associated model purveyors with research-based expertise in the given model. We use the term “evidence-based practice” to refer to behavioral treatment models that are implemented by frontline providers in the course of routine clinical practice. At the risk of oversimplification, the ultimate goal of dissemination is to turn ESTs into EBPs.

REFERENCES


for bulimia nervosa in a service setting. Behavior Research and Therapy, 39, 299–308. doi:10.1016/S0005-7967(00)00004-8.


Received November 11, 2012; accepted December 18, 2012.