Assessing the Field Effectiveness of Acceptance and Commitment Therapy: An Example of the Manipulated Training Research Method

KIRK D. STROSAHL
Group Health Cooperative of Puget Sound, Seattle, Washington

STEVEN C. HAYES
University of Nevada

JOHN BERGAN

PHYLISS ROMANO
Group Health Cooperative of Puget Sound, Seattle, Washington

Health care reform and managed care have produced a growing emphasis on field effectiveness research. The present paper proposes a simple methodological model for conducting such research that can assimilate all of the setting features of effectiveness questions while not requiring that researchers abandon experimental controls in favor of survey methods. This "manipulated training method" is then applied to an analysis of the field effectiveness of Acceptance and Commitment Therapy (ACT). Seventeen master's-level therapists and one psychologist (those not receiving training = 10; receiving training = 8) participated in a 1-year project. The training package consisted of a didactic workshop, an intensive clinical training, and monthly supervision groups. Prior to training and following training the clients of all clinicians in the project were assessed. Among other findings, clients of ACT-trained therapists reported significantly better coping than the clients of untrained therapists and were more likely to have completed treatment in the 5 months following initiation of treatment. A structural equation model indicated that ACT training accounted for differential coping outcomes. There was also evidence that training improved clients' self-ratings of psychological acceptance. At 5 months following the initial session, clients of ACT-trained therapists were more likely than clients of untrained therapists to have concluded therapy, and were more likely to agree with their clinician regarding the ongoing status of therapy. Implications for research on clinical effectiveness and technology transfer of behavior therapy are discussed.

Requests for reprints should be address to Steven C. Hayes, Department of Psychology, University of Nevada, Reno, NV 89557-0062. Preparation of this manuscript was supported in part by a grant from the National Institutes of Health, National Institute on Drug Abuse, grant number DA08634. Other lessons from the ACT training project can be found in Hayes (1995).
The era of health care reform and managed behavioral health care promises to have important effects on psychotherapy research and the role of such research in clinical practice. Managed behavioral health care is moving from an exclusive focus on cost-containment—with all of the problems that involves (Bouie, 1987; Karon, 1995; Strosahl, 1996)—to “quality purchasing” as a guiding management practice (Fitzpatrick, 1992; Howard, Briel, Leuger, Mahoney, & Grissom, 1993; Strosahl, 1995). In this “Generation II” of managed care, competitive advantages accrue to systems that can deliver effective and efficient treatment, and can be accountable for these effects (Cummings, 1995; Strosahl, 1994, 1995).

These changes have led to an explosion of interest in field effectiveness research by researchers (e.g., Howard, Moras, Brill, Martinovich, & Lutz, 1996; Jacobson & Christensen, 1996; Newman & Tejeda, 1996; Seligman, 1995, 1996), federal funding agencies (e.g., Onken, 1997), and managed care organizations (Wagner, 1997). Field effectiveness research is characterized by an interest not so much in the efficacy of treatment compared to no treatment or other forms of treatment in a highly controlled setting, but to the applicability, generalizability, and applied impact of treatment techniques in practical settings. Unfortunately, researchers do not agree about the methods that are most appropriate for field effectiveness studies, and a practical result is that there are few widely accepted studies that address the question of field effectiveness.

Seligman (1995) describes several ways that effectiveness studies differ from clinical efficacy studies: heterogeneous patient samples with multiple problems, open-ended treatment, active selection and deselection of treatment by the patient, self-correcting treatment regimens, and a greater emphasis on general functioning as opposed to symptom reduction. Field-based studies in managed care settings add other features to this list: the use of unscreened, master’s-level practitioners with dominantly nonbehavioral backgrounds and the resulting importance of clinician acceptability, the use of “usual care” as a comparison condition (rather than wait list or placebo treatments), and the use of simple outcome measurement methods that are client and system friendly.

An important, but often overlooked, feature of clinical effectiveness research is that the consumers of such information are more likely to be in the service delivery community, rather than the academic community. This simple difference challenges many of the most tightly held beliefs of psychotherapy researchers since some methodological, conceptual, and statistical traditions that are highly valued in the context of clinical efficacy research may actually hinder the impact of field effectiveness research in the service delivery community. Other beliefs held by efficacy researchers may delay effectiveness research or make it expensive or impractical.
Conceptual and Practical Barriers to Field Effectiveness Research

There has long been a methodological tradition of pitting scientific control against generalizability—to see precision and scope as contradictory forces. Generally, the science community has endorsed a control-focused philosophy and only peripherally addressed generalizability to real-world settings. Achieving a balance between these two is made more difficult by the tendency of the scientific community to think of precision in highly value-laden ways: Research that is highly controlled is said to be “high quality” or show “high scientific standards” even if the research focus is extremely narrow and the questions are addressed in ways that are pragmatically useless.

Discussions of the role of efficacy and effectiveness research play out this historic tension between precision and scope, and often in value-laden terms (e.g., Brock, Green, Reich, & Evans, 1996; Hunt, 1996). It is useful to think of efficacy and effectiveness not in terms of control or quality per se but rather in terms of the fit between the research purpose and the research method. To make this point we will briefly examine five examples of conceptual or methodological traditions that make it hard to conduct field effectiveness studies. We will then describe a method that we believe could advance effectiveness research.

Effectiveness Research Must Follow Efficacy Research

The model of psychotherapy development that has dominated psychotherapy research over the last 20 years is a technological model borrowed from the FDA. In this model, Stage I is technical refinement and pilot research, Stage II is clinical efficacy and research into mechanisms of action, and Stage III is field effectiveness. There are several problems with this model as it applies to psychotherapy (Onken, Blaine, & Battjes, 1997), but what we wish to challenge here is the idea that clinical efficacy must necessarily precede field effectiveness. This belief has delayed field effectiveness research enormously, so that even the most validated psychotherapy methods have been subjected to limited (e.g., Wade, Treat, & Stuart, in press) or even no field effectiveness research.

The rationale for the sequential linkage between efficacy and effectiveness in drug research is that the field effectiveness stage is both expensive and potentially extremely unsafe to consumers, and that only methods that have proven to be safe and effective should be permitted to be disseminated in the field. However, psychotherapy as a rule does not pose the kind of danger to human health and well being that is true of experimental drugs, and it is very difficult if not impossible to gain consensus about which psychotherapy practices actually pose a psychological danger to the recipient of services. Outside of obvious instances of physical, sexual, or psychological abuse to clients, psychotherapy appears to have a relative low “side effect profile.” Moreover,
there is an inordinately wide range of practices that are described as "psychotherapy" without the type of empirical certification that is a hallmark of the FDA model. Unlike experimental drugs, these methods are widely disseminated (e.g., in books, tapes, and workshops) without any evidence of efficacy whatsoever. The demand that effectiveness must follow efficacy means that in these cases the researcher can only bemoan that dissemination occurs, but not actually research the impact of dissemination unless or until someone does the needed efficacy research.

If the impact of dissemination on clinical outcome can be researched directly and in a limited setting, however, there is no logical and compelling reason field effectiveness cannot be an early, indeed even an initial, stage of psychotherapy development and research. This type of field testing can occur within a framework that is both theoretically systematic and guided by basic methodological principles. In this paper, we will both describe and demonstrate one possible framework for research of this kind.

**Dissemination Must Involve Highly Focused Techniques for Axis I Disorders**

The technological model of psychotherapy research emphasizes tightly controlled clinical efficacy studies of highly specialized treatments that are focused on specific Axis I (and more recently, Axis II) conditions. This has many advantages for the clinical efficacy researcher. In theory, this strategy reduces intersubject variability, it makes federal funding more likely since it corresponds with the organization of federal agencies, and allows the use of highly refined outcome and patient selection variables. The field effectiveness question, however, is not just whether a technology *can* work (which clinical efficacy does show), but whether it *will* work in the real-world setting of health care delivery. The methods and strategies appropriate to one question may not generalize in whole cloth to the other, for several reasons.

First, many researchers seem surprised to learn that the “bread and butter” of daily clinical practice in an HMO setting is the patient experiencing life stress or life transition issues that do not fit an Axis I or Axis II condition. As many as 50% to 60% of patients seen in HMO mental health systems only meet criteria for a “V-code” diagnosis (Strosahl, 1994). In fully capitated systems that do not employ medical necessity exclusionary criteria to limit client access to services, “diagnosing up” to obtain insurance coverage or precertification is not needed, leading to a much higher percentage of “nonclinical” behavioral health problems. In general, problems of this kind are too diverse for specialized packages, which in any case often do not now exist for this class of clinical problems. Yet, demonstrations of the efficacy of more generally applicable approaches are sadly lacking with this general population of psychotherapy clients.

Second, while the resulting treatments are specialized, with a limited scope of application (Strosahl, 1994), they often require extensive training.
The cost of this combination for a health care delivery system can be considerable since it means that numerous clinicians must be continuously trained and retrained, but they will be prepared to treat only a limited number of clients as a result. The net effect is that, with the exception of certain common or costly diagnostic areas (e.g., panic disorder, major depression, borderline personality), most managed care systems do not seem to see multiple highly specialized treatment programs as a practical solution to the effectiveness challenges they face.

Third, some of the methods that have been used to define treatment technologically may not apply as readily to effectiveness research. Therapy manuals are an important step forward, but there are many potential problems to be solved before specialized manuals constitute a generally applicable viable strategy for disseminating behavior therapies (Addis & Carpenter, 1996). For example, researchers have limited experience in writing manuals that are user friendly and flexible across situations. Similarly, sophisticated ongoing assessment of therapist adherence and competence is rarely financially, logistically, or politically feasible in the practical health care delivery system.

In the present study we provide an alternative: focus on techniques in a flexible manual linked to broad clinical principles that might be useful across a range of conditions clinicians might treat. If a broad impact could be documented, training costs might be more easily justified.

Effectiveness Research Must Rely on Survey Methods or Gargantuan Multisite Studies

Models of field effectiveness research vary. On the one hand, some researchers claim that given the limitations found in the field (e.g., variable treatment duration, self-correction in the course of treatment, multiple client problems, and so on), it is "hard to imagine how one could ever do a scientifically compelling efficacy study" there (Seligman, 1995, p. 967). The methodological alternatives this view leads to for effectiveness research are much less controlled, such as "a survey of large numbers of persons who have gone through such treatments" (Seligman, p. 967). Others hope to conduct research that links more real-world variation in client, therapist, and treatment characteristics through huge multisite studies that turn such variables into columns and rows of a factorial study (e.g., Project MATCH Research Group, 1997).

The cost of the former approach is scientific precision. The cost of the latter is practicality and expense. The present study presents an example of a middle methodological road. It is relatively controlled and yet is quite practical. In principle, it can come close to meeting Seligman's challenge of a method that has most of the scientific precision of clinical efficacy research while truly addressing the field effectiveness question, with all of the setting features he listed and more.
Reaching the Line Clinician With the Underlying Model

In managed care settings, general psychotherapy increasingly has become the province of the master's-level therapist, as systems shrink their panels of Ph.D. and M.D. providers. Doctoral psychologists are increasingly used to help develop and evaluate therapies, to supervise complex standardized treatments, or to intervene in complex cases when standardized treatments fail (Hayes, Barlow, & Nelson-Grey, in press). In the health care delivery system, field effectiveness must address this question: Is the underlying model (a) straightforward enough to be understood by master's-level providers, without expensive training and ongoing supervision; (b) acceptable to clinicians from a wide variety of philosophical and training backgrounds, and (c) acceptable to clients in terms of being “user friendly,” promoting consumer empowerment and, ultimately, low dropout rates and high levels of participation.

As traditionally presented, behavior therapy seems often to fail on these three criteria in the practical world of mental health delivery systems. Instead, there has been a proliferation of other brief therapies in the managed care setting, often promoted by charismatic proponents. For example, many managed care systems are training line therapists in strategic therapy approaches (e.g., Solution Focused Therapy; deShazer et al., 1986; Narrative Therapy; White, 1995) as an answer to the “time effective/cost-effective” dilemma. There are negligible scientific data to support the clinical effectiveness of these approaches, but they have been widely promoted as generally applicable, they are understandable and appealing to most line clinicians, and they hold out the hope to managed care systems that training in these approaches will have broad impact on a wide range of clinical problems (Strosahl, 1994).

Behavior therapists have been slow to take the obvious next step toward increasing their impact on managed care systems: to develop training and supervision systems that are appealing to line clinicians, that are generally applicable, and that have been shown to improve clinical outcomes. Studies that have examined the degree to which therapists can be trained to be competent in behavior therapy have generally occurred either in the context of tightly controlled clinical trials (Weissman, Rouansville, & Chevron, 1982) or in studies that attempt to differentiate different “active” treatments based upon the in-session behavior of highly trained therapists (DeRubeis, Hollon, Evans & Bemis, 1982; Luborsky, Woody, McLellan, O'Brien, & Rosenzweig, 1982). While important to the success of clinical efficacy research, these investigations do not directly address the most critical decisions faced by modern mental health system administrators. It is not clear, for example, whether the results for highly trained and specially selected therapists will generalize to line therapists in managed care settings, especially at the master's level, whether the results for focused treatment programs will gen-
eralize to other kinds of interventions used with the diverse group of patients clinicians face day to day, and whether these approaches will generalize to clinicians who are nonbehaviorally trained.

The method proposed in this study addresses the question of clinician acceptability directly. Indeed, the essence of the approach is that the focus shifts from what the clinicians are doing with clients and its impact on client functioning, to what the trainer or disseminator is doing with the clinician and its impact on client functioning.

Efficacy Equals Effectiveness

Psychotherapy researchers have taken it as an article of faith that efficacious techniques will be effective in the field, provided only that methods are found to disseminate the technology. This is not necessarily true because there may be features of practice in the field that radically alter how techniques function. Take, for example, the recent “Phen-Fen” controversy involving the common pharmacotherapy practice of combining two separate weight-loss drugs that individually had gone through rigorous Stage I and II testing, prior to dissemination. Neither came out of efficacy testing with an explicit contraindication for use with the other and, were it not for a small group of people convinced that irreversible heart damage was occurring as an “applied” side effect of combining the two drugs, they might still be on the market today.

The Phen-Fen debacle should not be ruled out as an aberration that can't possibly apply to psychotherapy procedures. For example, the vast majority of applied practitioners, regardless of disciplinary background, describe themselves as “eclectic.” Operationally, this means most practitioners routinely mix procedures from different treatments and theoretical orientations to address the client's presenting problems. Like the psychiatrist in a weight loss clinic, psychotherapists will combine different procedures that work, based upon their clinical judgment. According to the rules of efficacy testing, each procedure is legitimate in its own right. At the same time, each procedure could be potentially dangerous or useless when combined with another procedure. Efficacy researchers might like to solve this problem by requiring all therapists to use manual-based procedures trained to a singular level of treatment fidelity, but the professional resistance to this idea and the training, monitoring, and other costs it entails places it into the realm of fantasy in the real world of health care delivery systems.

The lesson is this: Procedures that have marked efficacy may lack effectiveness in the field. Since the whole practical point of efficacy research is altering practices in the field, this observation should make researchers much less certain of the value of even highly efficacious methods. Said another way, successful effectiveness research is a methodological and empirical pinnacle to be sought after, more important in its own way than positive efficacy outcomes. It should not be an afterthought.
The “Manipulated Training Method” of Field Effectiveness Research

The method we will exemplify in the present study we term manipulated training. In the abstract, the core of this approach is as follows:

1. Assemble a group of clinicians and assess the process and impact of their work with a sample of the clients they serve;
2. Divide clinicians into training and no training (or alternative training) conditions;
3. Compare the two (or more) groups on the measures taken in step #1;
4. Provide the training; and
5. Reassess the process and impact of their work with a sample of the clients they serve and compare the two (or more) conditions on these measures.

This simple approach can be fitted into the real-world requirements of field effectiveness studies quite readily, and yet it is sufficiently similar to clinical efficacy research so as to enable researchers to proceed without complete methodological reorientation.

Addressing the Parameters of Effectiveness Research

It is useful to examine the correspondence between the parameters for clinical effectiveness research set forth by Seligman (1995) and the manipulated training method. It goes without saying that no approach, including the one described in this paper, is free from flaws—and it is not our intention to suggest that this method will fit all field effectiveness questions. We do think it is valuable to assess, however, whether this method can rise to the challenging circumstances field effectiveness research must face.

Heterogeneous Patient Samples With Multiple Problems

There is no reason that the manipulated training method cannot be used with heterogeneous patient samples, provided only that the methods trained are generally applicable. This need not be known beforehand. If the methods trained are not generally applicable, they will not produce differences in client outcome. Conversely, the method can also be used if the field setting supplies a flow of more specific kinds of clients, as might be the case in a specialty clinic. In that case, it can successfully test more narrowly focused methods.

Open-Ended Treatment

Since the focus is on the impact of training, not the rigid use of a particular technology, there is no reason not to let treatment vary in length. Length of treatment merely becomes a dependent variable that assesses part of the impact of training on the clinician.
Active Selection and Deselection of Treatment by the Patient

Provided only that training does not fundamentally alter the kind of patients who seek out treatment with a clinician (an issue that can be controlled by the agency or institution), client "therapy shopping" again turns into a dependent variable as reflected in retention statistics for the clinician.

Self-Correcting Treatment Regimens

The point of treatment process is effective treatment outcome; thus, if methods teach clinicians how best to be flexible, they will show beneficial client outcomes. There is no reason to restrict treatment regimens in this approach: It is the training, not the specific technique, that is being evaluated.

Emphasis on General Functioning and the Use of Simple Outcome Measurement Methods

It is the FDA-like technological model that demanded syndrome-specific outcome measures. In the manipulated training method, it is the dissemination process itself that it being evaluated, and there is no necessary reason to limit it to specific syndromes and specific measures.

The Use of Unscreened, Master's-Level Practitioners

The acceptability of training and its impact on line clinicians is necessary in order to show an effect using this method. Any clinical effectiveness paradigm must demonstrate the "trainability" and acceptability of a procedure with the typical provider that is used in a behavioral health delivery system.

The Use of "Usual Care" as a Comparison Condition

The selection of "control" conditions in the manipulated training method depends upon what is being controlled. If practical impact is all that is of issue, for example, "usual care" translates into a control condition of no special additional training. If it is important to control for "training per se," then other control conditions can be selected (e.g., the use of alternative training methods).

The Acceptance and Commitment Therapy Training Project

The clinical framework that was examined in this study was Acceptance and Commitment Therapy (ACT; Hayes, 1987; Hayes, Strosahl, & Wilson, in press). ACT (pronounced as a word, not a sequence of letters) is a treatment that is philosophically based in functional contextualism (Biglan & Hayes, 1996; Hayes, 1993) and theoretically based in contemporary behavior analysis (Hayes & Wilson, 1993). It is composed of a wide variety of specific techniques organized around several key principles.

The ACT model of psychopathology (Hayes & Gifford, 1997; Hayes, Wilson, Gifford, Follette, & Strosahl, 1996; Hayes et al., in press) is based on the idea that many forms of behavioral disorder come from (a) emotional
avoidance, (b) cognitive or verbal "fusion" (the tendency for the derived functions of verbal events to dominate over their direct functions), (c) a failure to distinguish self from behavioral content, (d) the motivational impact of a failure to embrace values, and (e) lack of committed action linked to chosen values. Rather than seeking to change all internal or external sources of personal distress, ACT attempts to help the client distinguish between situations in which direct change is desirable or possible and situations in which psychological acceptance is a more viable coping strategy. ACT encourages the client to "make room" for unavoidable negative private experiences (e.g., historically produced painful emotions, unwanted memories, disturbing physical sensations) and to focus on committed action directed toward valued life ends. The client is taught mindfulness and other skills that increase the sense of distinction between the difficult history or reactions a person has and the person per se as a conscious human being. The client is taught to deliteralize cognition so that private verbal formulations and evaluations are seen as an ongoing process of relating events verbally or cognitively, not a substitution of one event for another (Hayes & Hayes, 1992). More detailed examples of ACT techniques are available elsewhere (e.g., Hayes & Wilson, 1994; see Hayes et al., in press, for a book-length description; Kohlenberg, Hayes, & Tsai, 1993).

ACT is focused on a generally applicable set of phenomena, and seems to have a fairly broad effect. The outcome data on ACT are limited. Small but controlled studies have shown ACT to be an effective treatment for depression (Zettle, 1984; Zettle & Raines, 1989). Less well-controlled research provides some evidence that ACT is helpful in anxiety disorders (Hayes, 1987; Hayes, Afari, McCurry, & Wilson, 1990), chronic drug addiction (Walser & Hayes, 1995), and depression in personality disordered patients (Strosahl, 1991), or parents of disabled children (Biglan, 1990). Process research has shown that these changes occur through processes hypothesized for ACT: changes in emotional avoidance, acceptance, and cognitive defusion (e.g., McCurry, 1991; Khorakiwala, 1991; Zettle & Hayes, 1986; cf. Zettle, 1984; Zettle & Raines, 1989).

In the FDA model of effectiveness research, ACT would probably not have enough efficacy research to yet warrant field effectiveness trials. ACT is just now being tested in federally funded randomized clinical trials, and the Stage II data are quite limited. Yet Group Health Cooperative of Puget Sound, one of the largest HMOs in the northwest, was interested in having their clinicians trained in the approach. This interest seemed to stem from two sources: (a) the broad potential applicability of the method and (b) successes in applying the model to multiproblem, character disordered patients in the system (Strosahl, 1991). It did not seem to make sense to do the training without simultaneously conducting a field effectiveness study since (a) it was not known whether line therapists could be taught to apply basic ACT interventions in a cost-efficient way and that would produce beneficial clinical outcomes, and (b) administrators wanted to verify that they were "getting their
money's worth'' for the expense of training. The manipulated training method permitted a practical piece of effectiveness research to be conducted in the context of training that would have proceeded in any case.

**Method**

**Site**

Group Health Cooperative of Puget Sound (GHC) is a staff model HMO with approximately 380,000 consumers in the Puget Sound region. There are six mental health centers in the staff model portion of the HMO, with a combined staff of approximately 100 mental health providers. Approximately 7% of all GHC consumers seek mental health care in any given year, resulting in approximately 29,000 service episodes annually. Brief problem-focused therapy is the dominant delivery model in the mental health system. Historically, the mean number of visits per episode of care has been approximately 4.5. Prior research has demonstrated that clinical outcomes in the GHC mental health system are comparable to those obtained in the fee-for-service sector (Manning et al., 1984).

**Therapists**

To be eligible for the training study, a therapist had to provide both baseline and posttraining clinical effectiveness data and, in the case of ACT trainees, participate in the workshop and intensive phase of ACT training and attend at least 75% of the monthly ACT consultation groups. Of 38 therapists who either provided certain types of control data or participated in some aspects of the ACT training program, 8 therapists met the ACT training criteria (2 males and 6 females), while 10 provided a complete set of control data (3 males and 7 females). All but one of the final study therapists were master's-level clinicians (there was one doctoral psychologist). The most common reasons for exclusion from the training were leaving employment, attending less than the required number of monthly consultations, or working in roles that did not involve providing general psychotherapy as a primary job responsibility (psychiatrists, nurses, and most psychologists).

The therapists in the final cohort had significant postdegree experience ($X = 5.2$ yr.) in the application of time effective therapies. In a survey of participating therapists, all indicated they were proficient in applying brief therapy approaches to a wide variety of clinical complaints. The most common therapy orientation was solution focused (8), followed by cognitive behavioral (4), narrative therapy (3), problem focused (2), and Ericksonian strategic therapy (1).

As is common in field studies, participation in the training was voluntary, and as a result therapists were not randomly assigned to conditions. This produced initial differences in training orientation between the groups. An extended and intensive training program had been offered in solution-focused therapy during the previous year within the system, quite similar to the ACT
training project in its structure and intensity. Eight solution-focused therapists elected to act as controls, perhaps because of their commitment to an alternative therapy model. These initial differences allowed certain kinds of subsequent comparisons, as will be addressed later, and provided some limited control for the impact of training per se, since so many of the control therapists had recently participated in an intensive training program.

The ACT Training Module

A skill-based training approach was employed that involved a combination of didactic presentations, intensive clinical practice and ongoing monthly consultation groups led by one or both of the senior authors (SCH, KS). The training started with a 2-day ACT workshop led by one of us (SCH), followed by a 3-day “intensive” clinical workshop, and dissemination of a detailed therapy manual (Hayes, McCurry, Afari, & Wilson, 1993). The intensive workshop involved in-depth exposure to specific ACT strategies as well as direct observation of ACT sessions, followed by extended group discussions. Following completion of the intensive phase, two monthly consultation groups (organized by geographic region) were formed, which included the ACT trainees and other therapists who were interested in the consultations but did not meet the study eligibility criteria. Each group met monthly for 3 hours with one of us (SCH or KS) to discuss cases, observe live and videotaped ACT sessions, role play ACT techniques, and discuss critical ACT strategies. The ongoing consultation process lasted 1 year. Because of the popularity of the training, a second identical training regime was employed the following year to a larger group of clinicians. However, the results of the present study are based upon the effects of the first 1-year training program.

Training Effectiveness Studies

Assessment of the impact of training effectiveness was examined during two periods. The baseline study occurred before the initiation of training. For 1 month, all consecutive new intakes seen by training or control therapists were enrolled in the clinical effectiveness study, and pretreatment measures were taken. Because treatment was free to vary in intensity and/or length in the applied setting, the choice was made to pick a standard time frame for posttreatment assessments. Prior utilization research at GHC has indicated that 90% of all clients are finished with a treatment episode within 5 months of initiation. Thus, 5 months after entering treatment, all clients were reassessed to measure their clinical response.

The posttraining clinical effectiveness study was completed after the 1-year training program had been completed. Again, it involved assessing all consecutive intakes for 1 month and obtaining posttreatment data 5 months after entry into therapy. In effect, this amounted to a replication study of the clinical effectiveness of trained and untrained therapists. As is characteristic of field studies, these two cohorts cannot be assumed to be the same, since the system itself continuously evolves (e.g., Medicare patients were enrolled for
the first time during this period; many new employers were added or sub-
tracted, each with certain modal kinds of individuals, and so on). For that 
reason, the two cohorts will be examined separately.

Clinical Effectiveness Measures

To ease demands on clients, providers and clinic service delivery infrastruc-
tures, as well as to address the tremendous heterogeneity of presenting prob-
lems treated in this general effectiveness study, an ideographic clinical assess-
ment protocol was developed that is similar to those used in goal-attainment 
scaling systems in the fee-for-service and community mental health sectors. 
Prior to seeing their therapist for the first time, clients were asked to complete 
a “Problem Identification Survey.” This procedure first involved the client 
writing out a brief statement of his or her presenting problem. Clients then 
rated this problem on a 1 to 5 Likert scale according to its severity (Severity 
Index), how well they were coping with the problem (Coping Index), and 
how well they were accepting emotions, thoughts, memories, and other pri-
ivate reactions to the problem (Acceptance Index). Clients also provided infor-
mation about past counseling experiences and what they hoped to get from 
therapy. Following the initial interview, therapists provided information 
about provisional diagnosis, clinical goals and parallel ratings of pretreat-
ment problem severity and coping efficacy, on similar 1 to 5 Likert scales.

Due to cost constraints associated with the large initial cohort size, post-
treatment client data were obtained by a mail-out “Consumer Satisfaction” 
survey that asked the client again to complete the problem severity, coping, 
and acceptance indices. Prior research at GHC had suggested that there were 
no significant clinical or demographic differences between clients who did 
and did not complete the mail-out posttreatment surveys, suggesting this 
method of data collection produces a representative clinical sample (Stro-
sahl, Mason, & Romano, 1993). At the same time, therapists conducted a 
posttreatment chart review of each patient to determine a final DSM-IV diag-
nosis, service modalities used in treatment (individual, group, family, couples,
medication), the number of service units used in each modality, the current 
status of the case, and the nature of termination, if the case was closed.

Clients

Clients were enlisted into the original sample if they completed a Problem 
Identification Survey prior to their first appointment with a training or con-
trol therapist. Initial appointments with these therapists were made by GHC 
intake workers in the normal fashion (i.e., on the basis of their own judgment 
and open intake slots for available therapists and without regard to the 
training status of therapists). This resulted in an initial sample of 321 clients 
overall, divided almost equally between the baseline (n = 172) and post-
training (n = 149) phases. More clients were seen overall by control ther-
pists (n = 201) than training therapists (n = 120) overall, both in baseline 
(Training = 61; No Training = 111) and posttraining (Training = 57; No
This was due to the fact that the training group had more part-time practitioners, who saw a smaller number of new clients on a weekly basis during the two 1-month enrollment periods. The range of clients seen by training and control therapists over the course of the two studies was 10 to 27.

There were no meaningful differences in the diagnostic composition of the baseline and posttraining samples. Further analyses indicated there were no significant differences in the diagnostic composition of the training and non-training clients, either at baseline or at posttraining. Overall, nearly 55% of all clients received final diagnoses of V-codes. The most common were Partner Relational Problem (22%), Phase of Life Problem (8.5%), Parent-Child Problem (7.5%), and Relational Problem (7.0%). Among DSM-IV Axis I diagnoses, the most common were Adjustment Disorder (16.4%), Affective Disorders (12%), and Anxiety Disorders (7.4%). Eight percent of clients received a final diagnosis of Personality Disorder. A large percentage of clients reported at least one prior episode of counseling (63%). Nearly 50% of these clients had been seen for prior treatment within the GHC mental health system. This is a well-known characteristic of HMO mental health samples, where the clinical emphasis is on brief problem-focused therapy and providing intermittent care over the client's life cycle.

Clients who completed the Problem Identification Survey (at pretreatment) and returned the mail-out “Consumer Satisfaction” survey (at 5 month's post-treatment) were included in the final study sample. One hundred twenty-six clients did so: 59 clients in baseline (Control = 43; Training = 16), or 34% percent of the cohort, and 67 clients in posttraining (Control = 42; Training = 25) or 45% percent of the original cohort.

The mean age of the client cohort in the original sample was 38.5 with a range of 8 to 77, with a sex distribution of 210 females (65.4%) and 111 males (34.6%). Because more turned in questionnaires 5 months after therapy initiation, females were slightly overrepresented as subjects in the baseline sample, but not in the posttraining sample (baseline = 75%; post-training = 65.7%). For the same reason, both baseline and posttraining samples were slightly older than the original sample (baseline = 42.5; post-training = 42.4). Analysis of severity, coping, acceptance, and diagnostic measures failed to indicate any pretreatment level differences between clients who did and did not return posttreatment surveys. This result was consistent with earlier studies.

Results

Baseline Analysis of Service and Clinical Characteristics

Analysis of service utilization characteristics indicated no differences between training and control therapists in the number of individual, couple, group, or family sessions used in the course of treatment. However, at baseline, training therapists were significantly more likely to seek medications
for their clients, referring nearly 65% of their clients for medication, compared to 19% for the control therapists (Cramér's $V = .25, p < .003$).

Chart reviews indicated the vast majority of training (85%) and control (88%) therapist clients had completed treatment by the 5-month posttreatment survey. This was confirmed by client self-report on the posttreatment survey, as only 14% and 19% of training and control therapist clients regarded their treatment as "in progress" at the time of the survey.

**Posttraining Analysis of Service and Clinical Characteristics**

Analysis of service data indicated that the significant difference in medication referrals seen in the baseline sample was not observed in this sample. Both groups were quite similar in their referral patterns (training = 21% vs. no training = 22%), suggesting that training had reduced the therapist's tendency to rely on adjunctive medication treatment.

A significant difference was also observed on whether clients reported they were still in active treatment at the 5-month assessment (Cramér's $V = .23, p < .058$). Clients of trained therapists (86%) were more likely to have completed therapy than clients of untrained therapists (49%), although this effect was not significant when therapist ratings of the same issue were compared for the trained (90% completed) and untrained (78% completed) therapists. Said another way, most clients (86%) and therapists (90%) in the ACT-trained group reported therapy was over at the 5-month follow-up. In the untrained group, however, therapist and clients did not view the situation similarly. Most clients (51%) in this group thought therapy was continuing—and thus presumably expected to request additional services—while relatively few therapists (22%) thought therapy was ongoing. Again, there were no significant differences in the average number of individual, couple, family, or group sessions used by trained and untrained therapists. In the context of the client data, this suggests that trained therapists were finishing their cases more efficiently.

**Baseline Clinical Effectiveness**

Table 1 presents the pre- and posttreatment means and standard deviations of the baseline sample on all clinical effectiveness measures and adjusted posttreatment means for training and no-training patients ($n = 59$). To control for pretreatment level differences and relatively high correlations among the clinical effectiveness measures, each dependent variable was analyzed using a univariate analysis of covariance (ANCOVA), with pretreatment scores on problem severity, coping and acceptance indices functioning as the covariates. Analysis of the problem severity index indicated no differences in clinical effectiveness between training and control therapists as a group. Analysis of the coping index revealed no significant differences between training and control therapists in coping outcomes. Analysis of the acceptance index indicated a significant group effect, $F(1, 54) = 6.42, p < .01$. Clients of control therapists reported significantly better use of acceptance as a coping strategy.
as a function of treatment, perhaps because there were more solution-focused therapists in this condition. The initially superior baseline impact on acceptance is perhaps not surprising, given that the solution-focused approach largely ignores “problem talk,” and emphasizes building on the client’s existing strengths rather than trying to alter perceived weaknesses. In other respects, therapists who elected to not undergo training were performing more effectively as clinicians at baseline since they were also less likely to refer clients for medications and completed their cases equally efficiently in the 5-month study period.

Posttraining Clinical Effectiveness

A similar ANCOVA procedure was used to analyze the clinical effectiveness data from the posttraining sample. Table 2 presents pre- and posttreatment means and standard deviations for each effectiveness variable, and adjusted posttreatment means for training and nontraining groups. Analysis of the problem severity index revealed no significant differences in outcomes among trained and untrained therapists. Analysis of the coping index revealed a significant group effect, $F(1, 62) = 4.05, p < .05$. ACT-trained therapists as a group produced significantly better coping outcomes among their clients.

### TABLE 1
MEANS, STANDARD DEVIATIONS, AND ADJUSTED POSTTREATMENT SCORES FOR THE BASELINE SAMPLE

<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>Pre-Treatment</th>
<th>Standard Deviation</th>
<th>Post-Treatment</th>
<th>Standard Deviation</th>
<th>Adjusted Score Training</th>
<th>Adjusted Score Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severity index</td>
<td>3.61</td>
<td>.98</td>
<td>2.37</td>
<td>1.22</td>
<td>2.52</td>
<td>2.31</td>
</tr>
<tr>
<td>Coping index*</td>
<td>3.22</td>
<td>.89</td>
<td>2.17</td>
<td>.87</td>
<td>2.04</td>
<td>2.31</td>
</tr>
<tr>
<td>Acceptance index</td>
<td>3.63</td>
<td>.91</td>
<td>3.44</td>
<td>.98</td>
<td>3.00</td>
<td>3.60</td>
</tr>
</tbody>
</table>

* The coping index was reverse keyed; thus a lower score indicates higher coping.

### TABLE 2
MEANS, STANDARD DEVIATIONS, AND ADJUSTED POSTTREATMENT SCORES FOR THE POSTTRAINING SAMPLE

<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>Pre-Treatment</th>
<th>Standard Deviation</th>
<th>Post-Treatment</th>
<th>Standard Deviation</th>
<th>Adjusted Score Training</th>
<th>Adjusted Score Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severity index</td>
<td>3.63</td>
<td>1.00</td>
<td>2.57</td>
<td>1.20</td>
<td>2.64</td>
<td>2.44</td>
</tr>
<tr>
<td>Coping index*</td>
<td>3.12</td>
<td>.91</td>
<td>2.39</td>
<td>1.07</td>
<td>2.18</td>
<td>2.48</td>
</tr>
<tr>
<td>Acceptance index</td>
<td>2.81</td>
<td>.94</td>
<td>3.45</td>
<td>1.09</td>
<td>3.33</td>
<td>3.55</td>
</tr>
</tbody>
</table>

* The coping index was reverse keyed; thus a lower score indicates higher coping.
Analysis of the acceptance index now indicated no differences between trained and untrained therapists on this effectiveness measure—reversing a significant deficit noted in the baseline study. These results suggest training had a significant impact upon the performance of therapists. They now produced significantly better coping outcomes and were equally effective in their ability to mobilize acceptance-based coping strategies. Finally, ACT-trained therapists no longer were more likely to refer for medication and were more likely to have completed their cases by the 5-month follow-up assessment, without using more service resources.

**Impact of Training on Clinical Effectiveness**

Because this training study involved multiple therapists surveyed over time, open-ended treatment, a heterogeneous client population, and large differential cell sizes among therapists and between training conditions, a structural equation modeling analysis was used to separately analyze the baseline and posttraining data sets (cf. Joreskog & Sorbom, 1993). This statistical approach is capable of modeling multiple interacting effects of dependent and independent variables in a way that promotes theory testing while retaining statistical precision. The objective of baseline structural path analysis was to model the mechanisms underpinning the clinical effectiveness of therapists as viewed from an ACT perspective of therapeutic change and to assess whether the specified effects could be linked to training versus no training. The correlation matrix of dependent and independent variables for the baseline and posttraining data sets are presented in Tables 3 and 4, respectively.

In preliminary analyses, a variety of potential therapist and client variables were included in the structural equation modeling to act as a potential control

<table>
<thead>
<tr>
<th></th>
<th>Post-acceptance</th>
<th>Post-coping</th>
<th>Post-severity</th>
<th>Pre-acceptance</th>
<th>Pre-coping</th>
<th>Pre-severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-acceptance</td>
<td>.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-coping</td>
<td>-.40</td>
<td>-.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-severity</td>
<td>-.02</td>
<td>-.05</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-acceptance</td>
<td>-.06</td>
<td>.28</td>
<td>-.01</td>
<td>.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-coping</td>
<td>-.01</td>
<td>-.29</td>
<td>.12</td>
<td>-.39</td>
<td>-.66</td>
<td>.17</td>
</tr>
<tr>
<td>Pre-severity</td>
<td>-.44</td>
<td>-.03</td>
<td>.11</td>
<td>.09</td>
<td>-.23</td>
<td>.17</td>
</tr>
</tbody>
</table>

* The original coping index items were reverse keyed (lower scores indicated higher coping) and are presented that way in Tables 1 and 2. In this matrix, the direction of coping scores were re-keyed so that higher scores in all categories equals more presence of the concept listed.

** point bi-serial correlations.
TABLE 4

INTERCORRELATIONS OF INDEPENDENT AND DEPENDENT MEASURES
IN THE POSTTRAINING SAMPLE

<table>
<thead>
<tr>
<th></th>
<th>Post-acceptance</th>
<th>Post-coping</th>
<th>Post-severity</th>
<th>Pre-acceptance</th>
<th>Pre-coping</th>
<th>Pre-severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-acceptance</td>
<td>.63</td>
<td>- .59</td>
<td>-.88</td>
<td>.36</td>
<td>.40</td>
<td>-.41</td>
</tr>
<tr>
<td>Post-coping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-severity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-acceptance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-coping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-severity</td>
<td>-.56</td>
<td>-.58</td>
<td>.57</td>
<td>-.26</td>
<td>-.69</td>
<td></td>
</tr>
<tr>
<td>Training condition**</td>
<td>.00</td>
<td>.27</td>
<td>-.16</td>
<td>-.18</td>
<td>.04</td>
<td>-.14</td>
</tr>
</tbody>
</table>

* The original coping index items were reverse keyed (lower scores indicated higher coping) and are presented that way in Tables 1 and 2. In this matrix the direction of coping scores were re-keyed so that higher scores in all categories equals more presence of the concept listed.

** point bi-serial correlations.

for confounds arising from the fact that patients and therapists were non-randomly assigned to conditions. This included such factors as age, sex, and experience level of therapist and, for clients, age, sex, general class of presenting problem, and whether the client had had a previous episode of mental health care. Also included in these analyses were the main variables of interest, namely, pre- and posttreatment problem severity, coping and acceptance indices, as well as training versus no-training. Including the variables designed to detect the confounding effects of nonrandomization precluded the identification of a statistically acceptable solution (i.e., the resulting model did not characterize the data set), so the control variables were dropped and the main study variables were analyzed. This resulted in a statistically acceptable solution.

Figure 1 presents the structural model for the baseline data set. Prior to training, the model shows that training therapists produced significantly worse outcomes in terms of using acceptance as a coping strategy. Pretreatment coping was a strong predictor of both posttreatment coping and problem severity, and training or nontraining status had no influence on coping outcomes. This model does a good job accounting for variation and covariation existing in the data set (chi-square (16) = 12.37, ns). The Goodness of Fit Index (GFI = .99) indicates that most of the true score variation in this data set is explained by this model. Interestingly, even at this stage of the study, the solution suggests that coping efficacy is a much stronger determinant of outcome than estimates of problem severity.

With the posttraining data set, we modeled a variety of solutions that posited a relationship between ACT training, coping, acceptance, and problem severity. Our main hypothesis was that ACT training would differentially affect acceptance and coping indices. The impact on problem...
Baseline

severity is harder to predict since ACT diminishes a focus on symptomatology, but also attempts to strengthen coping which might reduce symptomatology. As was the case in the baseline data set, we could not integrate potential therapist and client factors without failing statistical criteria for the adequacy of the resulting model. In addition, a model including both differential impacts on both coping and acceptance would not fit the data. Figure 2 presents the best statistical model for the posttraining data set (chi-square \((17) = 27.05, ns\)). Again, the GFI suggests this model accurately accounts for variation in the dependent and independent variables (GFI = .98). This model indicates that there is a direct, significant effect of ACT training on posttreatment coping levels. Therapists who received ACT training were more effective at mobilizing patients' coping skills, but training did not differentially affect posttreatment assessments of problem severity. In fact, estimates of problem severity appear to predict little in either data set, whereas coping is a very dominant mechanism, accounting for five of six outcome indicators. This is consistent with the theoretical and clinical basis of

Fig. 1. Structural equation solution for variables in the baseline study. The original coping index items were reverse keyed but in this model the direction of coping scores were re-keyed so that higher scores in all categories equals more presence of the concept listed.
Post-Training

ACT, which emphasizes use of acceptance strategies to elevate general coping efficacy, generally independent of attempts to directly control and/or eliminate distressing experiences per se.

The fact that coping supersedes acceptance is not surprising since acceptance strategies are but one subset of coping strategies. Acceptance should be significantly related to coping (as it is, see Table 4), but should not supersede it. As for acceptance per se, recall that acceptance was a negatively associated training factor in the baseline data set, but was not a statistically significant factor in the posttraining model. A likely explanation is that therapists in ACT training improved in their ability to mobilize acceptance strategies, such that they were no longer distinguishable from the untrained colleagues, as had been the statistical case in the baseline study.

The average therapeutic effect sizes noted in both clinical effectiveness studies are comparable to those reported in various clinical efficacy studies,
and provide additional evidence that the posttraining cohort was somewhat more difficult (\(d' = 1.12\) in baseline; \(d' = .85\) in posttraining). Average effect sizes for all measures were identical at baseline between the two groups (\(d' = 1.12\)) and were similar after training (\(d' = .87\) for trained and .83 for untrained), indicating that the changes observed were specific to particular measures, not global.

**Discussion**

The present study shows how it is possible to do field effectiveness research on the impact of behavior therapy methods in a direct, inexpensive, and applicable fashion. In the manipulated training research method, the primary relation of importance is between clinician training and client outcomes, not between the use of specific clinical techniques and outcomes. This research method provides a powerful and immediately available approach to field effectiveness that walks a middle ground between client surveys ("Mental Health: Does Therapy Help?", 1995) on the one hand and huge, multisite factorial approaches on the other. Conversely, no approach to clinical effectiveness research is perfect, and the manipulated training method used in this study is no exception.

As we review several limitations encountered in this study, it is important to notice that this method of field effectiveness research has an enormous advantage: The methodological concepts worked out over the last decades in clinical efficacy research apply (albeit sometimes at another level) to this approach. This means that field effectiveness methods need not be a whole new ball game for psychological researchers, as it is in approaches proposed by others (e.g., Seligman, 1995, 1996). Indeed, there is a potential methodological solution to nearly every limitation discussed, and these tactics could be tested in future manipulated training studies.

**Randomized Therapist Assignment**

Probably the biggest experimental limitation to the present study is the lack of random therapist assignment to training conditions. To some degree the lack of pretraining differences in therapy processes or outcomes suggests functional equivalence between groups, but this is not certain since unmeasured processes might interact with training in unknown ways to account for the effects seen.

Logically, there is no reason that random client assignment cannot be used in this approach to field effectiveness research. The problem is a practical one. In the real world of clinical practice, few administrators would compel a specific kind of training on an unwilling clinician. There may be other ways to circumvent this rather common limitation. For example, the *invitation* to participate in training could be random and those who agree or do not agree to participate can be compared.
Randomized Client Assignment to Therapists

In the usual field setting, random client assignment is not practical, though if it is there is nothing in this methodology that would prohibit making use of this experimental control. Nonrandom client assignment is not a problem, as long as the method does not change over time in a fashion that covaries with training participation. The biggest single methodological weakness in the method we are proposing is that training might alter how clients are assigned. For example, suppose a receptionist begins to assign difficult intakes to trained therapists on the grounds that they will know more from their training about how to handle hard cases. The manipulated training method would be confounded by such an effect. In the present study, the triage assignments were removed from immediate clinician control, but it is important to monitor for it when random client assignment is not possible.

Adherence and Competence of Therapists

Logically, the key type of adherence and competence measures needed in this methodological approach are adherence of the trainers to a training protocol. No such measures were taken in this case, which is a clear limitation. However, given the author's experience in developing and refining the ACT treatment model, there is less reason for concern than might be the case when trainers are more "distant" from a treatment model.

It is less clear that adherence and competence measures of a more traditional sort (i.e., regarding the use of ACT techniques by clinicians) are necessary. Without them results can occur that are difficult to explain. For example, while the results of this study reflect the effect of exposure to ACT training, we do not know to what extent trained therapists actually used specific ACT strategies during therapy sessions. That is not a major problem in the present study, in part because client outcomes were shown to be differentially positive due to training.

When positive outcomes are achieved, the absence of adherence and competence measures are not lethal, and may even be desirable when the clients have a broad range of problems. In the current study, clinicians were free with a given case to follow the entire ACT protocol, to use bits and fragments when they seemed appropriate, or to avoid these techniques entirely when they did not seem relevant. It probably would not have made sense (and the study could not have been conducted in a practical setting) had we insisted that a specific ACT protocol be followed with all clients, whether or not the procedures seemed to fit. We do know that therapist actions consistent with the ACT training must have been involved since clearly some of their therapeutic behaviors must have changed in order to produce improved outcomes in their clients.

The biggest potential problem in not collecting traditional adherence measures occurs when client outcomes are not positively influenced. In that case,
we would not know if the training changed therapist actions but the model is incorrect, or the model is correct but the training was not effective in changing therapist behaviors.

**Intent to Train**

In the present study only those therapists who received a fairly intensive course of training were included in the training group. In line with the "intent to treat" approach in clinical efficacy research, a more conservative approach might have been to use an "intent to train" model. For example, clinicians who dropped out of training might have been treated as a member of the training group for program evaluation purposes. This approach speaks to two fundamental issues for managed care administrators. First, how acceptable is the training package to the pool of targeted therapists. A high rate of unexplained therapist dropout might indicate that the training package is too difficult, doesn't generalize to real-world clients, or is competing with other therapist allegiances. In the present study, most of the therapist dropouts were attributable to uncontrollable system factors, rather than to some internal failure in the training system.

A second key issue is the cost of training itself relative to the cohort of potential trainees. When a large number of therapists are enrolled in a training program, delivery systems incur ongoing, fixed costs. These costs have to be reconciled against the training program "hit rate" (i.e., therapists who complete the program) and the "benefit rate" (i.e., the magnitude of improvements in clinical outcome compared to untrained therapists). At the organizational level, the decision whether to embark upon or continue a training program will in part be based upon these two factors.

**Control**

It should be acknowledged that field-based research of this kind is inherently noisy. Many variables are changing at once, and not all can be assessed, much less controlled. For example, the clientele of GHC changes over time, but not in ways that are easily known and understood. This is the main reason we treated each cohort largely as its own study. The uncontrolled factors in field-based research are probably one reason why so few studies have been published on the dissemination or clinical effectiveness of behavior therapy. This is a mistake, we argue, because clinical effectiveness studies ask different questions and have different roles to play in the fabric of empirical analysis of behavioral health approaches as compared to traditional clinical efficacy research.

The key issue is not control in the abstract, but control in the concrete. If this point is not made clearly, any research done in the practical environment is ipso facto "less controlled" in the sense that the quality of the conclusions are lower, and this we believe is clearly false. Consider, for example,
the controls needed to show that improvements in a clinicians' performance are real. In research on psychopathology it is well documented that spontaneous improvements can and do occur regularly, and for that reason mere pre to post client improvements often do not mean very much without other controls. We know of no such data in the area of clinician effectiveness, and in fact it is devilishly difficult to show that training and experience have any positive impact on clinicians at all (Christensen & Jacobson, 1994; Dawes, 1994). Thus, studies showing pre- to post-improvements due to training may be relatively well controlled and believable even if the assignment to training condition is not random, or the control training condition is not ideal. These other controls would, of course, be nice, but it is the adequacy of the total method, relative to the question being asked, that must be evaluated. To be important, field effectiveness studies must indeed be well controlled, but the adequacy of the controls must be evaluated case by case, relative to both the internal and external validity of the specific study.

**Precision Versus Scope of Measurement**

In this study, we opted to use an ideograph measurement system that had great scope, but lacked the type of precision that clinical efficacy researchers typically demand. The psychometric properties of the three outcome indices (i.e., severity, coping, acceptance) is not known. This is a thorny issue for effectiveness researchers. Considering that nearly 50% of the population studied will be assigned a V-code, general measures of psychiatric symptom distress are likely to suffer from serious "floor effects" that could seriously understate the magnitude of a training program. If a floor is approached, clients can say they are benefiting dramatically from treatment, but symptom-based measures suggest very little change is happening. Conversely, if researchers select a narrowly defined disorder where good symptom-based measures exist, then the focus of training is narrowed, leading to increased administrative concerns about the costs of training on a disorder-by-disorder basis, which in most systems would be prohibitively high. One potential solution is to combine both strategies, that is, measure the general effects of a behavior therapy across different presenting problems and also measure one or two more narrowly defined conditions to get a better look at how specific disorders are responding to the general treatment approach.

**Implications**

Demonstrating that field-based training in therapy methods positively impacts subsequent clinical effectiveness is a major challenge in the dialogue between behavior therapists and managed care administrators. It is not enough merely to appeal to the clinical efficacy literature because administrators are well aware of the difficulties in generalizing the results of efficacy studies. They are also all too familiar with the difficulties inherent in imple-
menting highly technical, specialized treatments within a service delivery system. Further, the task of studies evaluating clinical training in the managed care environment is not only to demonstrate a positive effect on clinical outcome, but also to obtain this outcome in a way that is financially and logistically feasible.

The results of the ACT Training Project suggest that training in behavior therapy at the delivery system level can have a positive impact on the clinical effectiveness of line (mostly master's-level) therapists. To our knowledge, this is the first demonstration that training in a behavior therapy approach is associated with better outcomes for clients who are not preselected by diagnosis. A wide variety of general methods exist (e.g., functional analysis, single-case experimental designs, cognitive therapy, a constructional approach to case conceptualization, and so on) that could be subjected to the same kind of manipulated training analysis employed here.

We have cast the result of this study largely in a methodological vein, but the substantive accomplishments also deserve notice. ACT training made a clear difference, which provides important support for the use of this technology and perhaps for the use of other acceptance-based approaches that share some of these principles and techniques (e.g., Greenberg, 1994; Kabat-Zinn, 1991; Koener, Jacobson, & Christensen, 1994; Linehan, 1993; Marlatt, 1994). The treatment effect sizes for the control and training therapists in the present study are comparable to those often reported in the clinical efficacy literature, and the results of ACT training on subsequent clinical effectiveness do not simply reflect a comparison with a weak control treatment or weak treatment effects overall.

Some of the outcomes went beyond changes in symptoms to outcomes of major systems importance. For instance, managed care organizations are particularly interested in the rapid conclusion of therapy with quality outcomes because this has an obvious impact on cost effectiveness in health care delivery systems. The ACT-trained therapists had finished a greater proportion of their cases at a 5-month follow-up, as viewed by client report. In addition, the ACT-trained therapists were much more likely to agree with their clients about the conclusion of therapy, a fact that has obvious long-term implications for client satisfaction and clarity in client-therapist communication.

Perhaps for these reasons, the effect of this training study on administrators at GHC was significant. Based upon these promising data and a generally positive reaction of line clinicians to the ACT model, a decision was made to use ACT as a preferred treatment model in the Intensive Services Program, designed for clients with greater levels of dysfunction and clinical need.

Psychotherapy efficacy researchers may understandably view this administrative effect with a certain degree of concern: While ACT is a new approach gradually building a base of empirical support, there are other techniques with considerably longer empirical track records. Many behavior therapists would surely prefer that these methods, not ACT, be implemented. But that
is precisely the point: Field effectiveness research can have far greater impact than yet another efficacy study in integrating behavior therapy methods into the health care delivery system. Researchers may wish it were otherwise, but that wish seems unlikely to change the realities.

Behavior therapists have had much less of an impact on managed care than one would expect, given the shared values between the two areas. Part of the problem may be that current training technologies in behavior therapy are too committed either to a specialized, procedures-oriented approach on the one hand (e.g., learning a very specific protocol), or an extensive, flexible, and principle-oriented approach on the other (e.g., learning to do functional analysis). Over time, there will be an increasing need for specialized packages for targeted and difficult-to-treat populations, but for the foreseeable future there will always be an even greater need for generally applicable therapeutic models.

The ACT model is behaviorally based and broadly applicable, but is focused on a small set of key concepts. Further, because of its focus on issues of emotions, acceptance, self, commitment, and the like, the ACT model is attractive to many nonbehavioral and master’s-level clinicians who are used to dealing with such concepts. As such, it occupies a kind of middle ground between focused psychotherapy models and broad sets of behavioral principles: a position that makes it particularly well suited for generally applicable training. These same attributes seem important in “marketing” the behavior therapies to the master’s-level community, and in filling an ignored market niche within contemporary managed care systems. In general, managed care clinicians seem to favor clinical models that are not “top heavy” with procedures, but rather use a small number of concepts and that have broad clinical appeal. Indeed, this may be a primary reason why the brief strategic therapies have been so successful (in terms of administrative support, if not clinical impact) in the contemporary managed care marketplace.

Fortunately, several cognitive-behavioral approaches seem to have these general characteristics. With proper marketing and supportive clinical effectiveness research, such approaches may have a major advantage over other therapy alternatives in securing a niche as general therapy approaches in managed care systems. A fairly unique “value added” feature of behavioral approaches is a commitment to verifying clinical effectiveness through research. This is simply not a value shared by most other therapies (including the strategic therapies), but it is increasingly a value held by managed care administrators who are looking for empirically validated models of care that provide the best clinical outcomes at the lowest possible cost. Field tests such as the kind reported here have the potential to show administrators that field-based training in behavioral and cognitive therapies not only can improve the general clinical effectiveness of mental health care, but may also be superior to other forms of therapy that claim to be time effective but do not back up these claims with data.
FIELD EFFECTIVENESS OF ACT

References


Received: October 1, 1996

Accepted: October 28, 1997