Improving the Transition From Basic Efficacy Research to Effectiveness Studies: Methodological Issues and Procedures

Gregory N. Clarke
Oregon Health Sciences University

This article proposes methodological strategies that, if used in treatment outcome research, may help in the transition of efficacy research findings into effectiveness trials in clinical and service delivery settings. Alternative methodologies are proposed to examine how treatment effectiveness may vary as a function of degree of treatment structure, treatment protocol compliance, psychotherapy integration into an overall treatment regimen, participant selection and composition, and variations in treatment parameters. The discussion focuses on encouraging the retention of experimental control while stretching psychotherapy outcome research designs to encompass effectiveness issues.

Recent proposed changes in the financing and organization of health service systems provide a compelling background for discussion of the relevance and applicability of findings from controlled laboratory studies (efficacy investigations) to broader mental health services systems (effectiveness research). If certain interventions prove effective in both laboratories and clinics, this may result in meaningful health care reforms with potentially improved services for consumers. Brook and Lohr (1985) advocate this perspective, noting that health care systems are improved not just by data from a single perspective but from an integration of data from efficacy, effectiveness, and quality-of-care perspectives.

However, several barriers limit consensus on the effectiveness and usefulness of mental health treatments in service system settings. First, empirical treatment outcome information is often incomplete and, for many disorders, all but unavailable. This lack of advanced treatment efficacy data is particularly true in child and adolescent mental health. Lacking controlled outcome research for a given disorder or problem area, it is difficult to resolve whether a given intervention will work as well in the clinical setting as it may in the laboratory. This should not be seen as a call to exhaust all avenues of efficacy research before initiating treatment effectiveness trials. Instead, the goal of this article is to encourage concurrent advances in both efficacy and effectiveness research, even to the point of addressing both perspectives within the same investigation.

Moving beyond the availability of research, however, there are several design and methodological shortcomings in the existing literature that limit the integration of efficacy methods within effectiveness trials. These shortcomings exist, in large part, because mental health outcome research is often designed to answer efficacy or theoretical questions. Scant attention is generally paid to generalization beyond the research paradigm (e.g., how well could this intervention be carried out in a clinical setting?).

This article attempts to address this issue by proposing methodological strategies that, if used in treatment outcome research, may help shift efficacy research findings and methods into effectiveness trials conducted in clinical and service delivery settings. Whenever possible, these issues are examined in the context of our own research in school- and clinic-based interventions for adolescent depression (Clarke, Hawkins, Murphy, Sheeber, Lewinsohn, & Seeley, 1995; Lewinsohn, Clarke, Hops, & Andrews, 1990), with examples of how the generalization of these studies may be improved.

Periodic reviews of methodological design and strategy in psychotherapy outcome research are nothing new (e.g., Kazdin, 1986). However, most previous discussions have been in the efficacy research literature, often weighted in favor of greater experimental control at the expense of generalization to real-world settings. This article revisits these same issues, but with the perspective of stretching research designs to encompass effectiveness issues. Bear in mind that this is not meant to preclude scientific rigor. Instead, I advocate retaining as much experimental control as possible while using greater creativity in the methodology and issues studied in treatment outcome research.

This article was inspired by Weisz, Weiss, and Donenberg's (1992) examination of the positive child psychotherapy effects in research studies, compared with the general absence of such effects in clinic-based studies. In their conclusion, Weisz et al. suggested that "...a key task for researchers [is]...identifying those proper conditions under which effects of child therapy may be optimized." The present article responds to this issue by considering how study design features might help identify these optimizing conditions.

Methodological Issues

Weisz and Weiss (1989) provide a detailed review of aspects of controlled experimental methodology that differ enough between research and clinical settings to limit the generalization of positive findings from the former to the latter. A circumscribed set of these issues is summarized in Table 1. Each of these is briefly reviewed in turn, followed by suggestions for methodological changes to enhance generalizability.

Correspondence concerning this article should be addressed to Gregory N. Clarke, Department of Psychiatry, Oregon Health Sciences University, Portland, Oregon 97221.
Table 1
Potential Mediating Factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Therapist training; degree of treatment structure; monitoring, protocol compliance</td>
</tr>
<tr>
<td>2</td>
<td>Combined or multiple treatments (e.g., pharmacotherapy and psychotherapy)</td>
</tr>
<tr>
<td>3</td>
<td>Multiple roles vs. single role for therapist</td>
</tr>
<tr>
<td>4</td>
<td>Participant selection (homogeneity vs. heterogeneity; comorbidity)</td>
</tr>
<tr>
<td>5</td>
<td>Control group (no treatment vs. attention placebo vs. usual care)</td>
</tr>
<tr>
<td>6</td>
<td>Treatment parameters (duration, dose, modality, location)</td>
</tr>
</tbody>
</table>

Other issues identified by these and other authors (Kazdin, 1978), but not addressed here, include participant recruitment methods; professional versus nonprofessional therapists; measurement technology; participant, therapist, and assessor masking to therapy condition; service setting; and participant assignment. This article does not attempt to exhaustively catalog method variants to address each of these parameters. Instead, a sampler of design features is proposed to motivate investigators who may be interested in broadening the scope of treatment outcome paradigms under consideration or in the planning stages.

Degree of Therapy Structure

Increasingly, controlled outcome trials of psychotherapy treatments provide intensive, specialized training in the specific research intervention protocol (Luborsky & DeRubeis, 1984). To simplify replication of the intervention across studies, researchers often “manually” tailor treatments by providing scripted therapist guidebooks (e.g., Clarke, Lewinsohn, & Hop, 1990; Moreau, Mufson, Weissman, & Klerman, 1991; see also Lambert & Ogles, 1988). Close compliance with these treatment manuals is often encouraged by audio or videotaping therapy sessions for later compliance review by research staff (Clarke et al., 1995; Lewinsohn et al., 1990; Hollon, 1988).

These training and implementation methods result in interventions that are highly regimented and very reproducible, with high adherence to a predefined protocol. From a pure efficacy perspective, these are desirable features because they control for extraneous contributors to treatment outcome. However, these controls are unlikely to be used in nonresearch settings because of the increased effort and burden they require, as well as a lack of interest in enforcing a reproducible treatment regime in many clinical or service settings.

These therapy structure methodologies impede effectiveness trials of research treatments because, in my experience, real-world therapists often resist following rigid or uni-modality interventions. Many clinicians prefer to be responsive to client-by-session presentations with a blended or “eclectic” therapy model, borrowing pieces of interventions as they seem relevant rather than using a scripted but potentially more cohesive protocol. Research interventions are often viewed as too regimented, leading clinicians (the interventionists in effectiveness trials) to resist using them out of concern that they may reduce psychotherapy to an automated “cookbook” approach, lacking responsiveness to individual client presentation.

How might outcome researchers design their studies differently to address this issue? One approach, which elaborates on a suggestion by Kendall and Lipman (1991), is to conduct treatment efficacy studies with several levels of experimental control over the implementation of therapy protocols to examine the effect on treatment outcome. Experimental conditions with less structured therapy content and implementation would resemble real-life clinical settings, although in a limited fashion. Of course, this does not mean that therapists would be allowed to deliver interventions completely unobserved or without limits. It is important to measure how much and in what way therapists deviate from a planned intervention (e.g., by using expert raters, a therapy content coding system, and videotaped intervention sessions).

An illustration of this proposal may examine the hypothetical effects of several potential mediators in the research on adolescent depression prevention. In a previous investigation (Clarke et al., 1995), my colleagues and I provided high school counselors with 40 hr of supervised training in the use of a scripted manual of cognitive intervention to prevent depression in at-risk youths. These counselors were strongly discouraged from deviating from the intervention protocol. Audiotaped reliability checks revealed that they were very compliant (94%) with the scripted protocol. Although satisfactory effects were obtained with the manual-specified intervention, it would be interesting to conduct this study again with the intent of exploring “therapy structure” issues. For example, a semi-factorial design may be used to examine the extent to which planned and measured variations in (a) therapist deviation from a regimented treatment protocol and (b) the use of structured intervention manuals and specialized training had an effect on outcome. Three different implementation versions of the same cognitive therapy intervention could be used:

1. Rigorous therapist training with a structured manual and minimal protocol deviation (enforced by means of videotaped monitoring). This is essentially the study as it was originally conducted, and it represents typical efficacy research design.
2. Similar therapist training with a structured manual but with minor to moderate protocol deviation permitted on the basis of therapists’ clinical judgment. For example, anger management, although not part of the original protocol, might be offered to a depressed adolescent with comorbid conduct disorder. Therapists would be monitored by means of videotaped sessions, and deviations would be assessed and coded, but not corrected through supervision.
3. General therapist training in the same theoretical approach as espoused in the manual, but with no structured manual provided. Treatment course and planning would be based on the therapists’ best clinical judgment. Similar to the second condition, therapist deviations would be assessed but not corrected.

Measuring how experienced clinicians elect to deviate from

---

1 Note that the remaining cell in this design (no manual, but no deviation permitted) cannot be implemented, because where a manual is not provided it is not possible to ascertain whether protocol has been followed.
protocols is a difficult task but necessary in this design. If certain so-called deviations from protocol are associated with good client outcome, structured research interventions might be improved by the incorporation of these techniques or activities in the protocol. In this way, unstructured clinical practices might more systematically influence the development of effective structured research interventions, contributing to a two-way exchange of information between researchers and clinicians. At the very least, researchers may better understand which components of psychotherapy approaches are fairly robust and tolerant of individual therapist deviations, and which aspects are relatively fragile and should be carried out in a fairly uniform manner.

One may argue that this design introduces too many opportunities for unexplained results. For example, what if the least structured condition was associated with the best outcome? How would one know what aspects of the treatment accounted for this finding? Furthermore, can these positive findings be replicated, or were they just a function of unique therapist characteristics? Certainly therapy content data (from coded videotaped sessions) must be examined to determine whether systematic therapist deviations from established protocol were associated with better outcome. If so, subsequent studies could be conducted with controlled variations of these deviations to further test their impact on outcome. Regarding replication, a single study of this type is not meant to answer all questions. Replication is still as important to this type of design as it is in efficacy research, and failure to replicate would indeed raise questions as to whether uncontrolled therapy implementation is best.

Too few controlled psychotherapy studies have specifically tested these issues to predict what results would be obtained. However, Weiss and Weisz’s (1990) meta-analysis of methodological factor effects on child psychotherapy outcome research suggests that increased methodological rigor is generally (although not uniformly) associated with more beneficial treatment outcome across unrelated studies. Although Weiss and Weisz (1990) did not specifically examine the mediating effects of intervention implementation rigor, extrapolating from their data suggests that more positive results may be obtained when treatment protocol is carefully carried out.

Although studies examining treatment implementation issues in a controlled paradigm are virtually unavailable in psychotherapy outcome research, these issues are an increasing focus of school-based prevention and health promotion research (Felner, Phillips, DuBois, & Lease, 1991). For example, Rohrbach, Graham, and Hansen (1993) examined the relationship between integrity of program delivery and outcomes of a school-based, psychosocial substance abuse prevention program. School districts were randomly assigned to either intensive or brief teacher training in the program; schools within districts were randomly assigned to have the principal involved or not involved in the intervention. Not only was program implementation highly variable but it also faded from the first to the second year. Delivery of the program in a rigorous manner was predictive of positive student outcomes. These results are similar to those reported in other studies of school-based substance abuse prevention programs (e.g., Botvin et al., 1989; Pentz et al., 1990). Although these findings are not directly applicable to traditional psychotherapy, many design and hypothesis issues examined by Rohrbach et al. (1993) are relevant to effectiveness studies that examine the implementation and real-life usefulness of psychotherapy services.

If future studies find that greater adherence to psychotherapy protocols predicts better client outcome, what implications would this have? Such a conclusion might be unpopular with clinicians, especially those who favor therapist autonomy. This likely fallout is an effect that researchers should acknowledge, because concerns regarding this use of effectiveness data is a major contributor to clinician resistance to the adoption of research interventions.

### Integrated Versus Isolated Services

Efficacy studies typically offer their treatments in isolation, often to focus on the disorder of interest or to remove or control extraneous mediating factors (e.g., other treatments). This isolation is often twofold: first, research treatments often target only one disorder, diagnosis, or problem domain, without addressing psychiatric or general medical comorbidity in individuals. A second and related form of treatment isolation arises when efficacy studies provide only the research intervention, without the frequent clinical requirement to offer an integrated array of multiple different assessment, intervention, referral, and advocacy services in addition to the single research focus. This insularity means that research projects have a generally greater capacity to devote more time, resources, and follow-up to their limited number of participants. In contrast, effectiveness trials conducted at sites such as schools and community mental health programs can typically provide only a limited number of visits and contact hours to each client.

Although for theory testing it is often desirable to pare interventions down to a single, internally cohesive treatment component, this may not reflect mental health services as often provided in the community. For at least some populations and settings, services are often provided in an integrated intervention "package," consisting of several interconnected parts. For example, community-based services for individuals with chronic mentally illnesses such as schizophrenia often include a physical examination, psychotropic medications, psychological services, case management, housing support, and social welfare services, with other components added as required (e.g., Solomon, 1992). Researchers must acknowledge that mental health interventions are often embedded in a larger context of general health and social services, and intervention trials of "atomized" psychotherapy components may have little external validity. Initial outcome studies of psychotherapy services may be geared toward testing the larger, integrated intervention package, with subsequent "component analyses" conducted only after the composite intervention is associated with beneficial outcome.

How could methodological variations in outcome research address this issue? Although daunting, one obvious solution is to develop an integrated intervention regime that offers research-based intervention tracks for a variety of common disorders, with clear triage rules regarding assignment to intervention (or interventions) on the basis of assessment findings. Furthermore, interventions must be integrated to address those instances in which participants have more than one disorder,
and thus, they must enroll in more than one intervention track. The control condition for this design may be the locally evolved intervention standards for addressing the heterogeneous clinical presentation seen in the community. Defining these local standards for the purposes of an effectiveness trial requires substantial preliminary meetings with representative local providers to clarify and codify the usual and customary services provided for a given client population.

Such a protocol might also introduce realistic personnel and fiscal budgets to both research and control conditions, within which both conditions must provide all services required by the client sample. Budgetary caps of this sort simulate service system limits and may address the perception that research interventions are overly enriched. Caseloads and service burden would be comparable, increasing confidence in the generalizability of outcome findings.

Examples of research-based treatment parameters for psychiatric disorders are rare, but becoming less so. For example, the Journal of the American Academy of Child and Adolescent Psychiatry recently published practice parameters for several common child mental health disorders, including anxiety disorders (Bernstein & Shaw, 1993), attention-deficit hyperactivity disorder (Jaffe, 1991), schizophrenia (McClellan & Werry, 1994), and conduct disorder (Jaffe, 1992). Similarly, this journal has recently published several articles on the treatment of common combinations of comorbid disorders (e.g., Mueser, Bellack, & Blanchard, 1992; Shea, Widiger, & Klein, 1992). Parameters such as these may form the basis for developing the "integrated intervention regime" advocated earlier.

Of course, mounting a comprehensive study of this type might prove so costly that it would quickly exceed traditional research budget limits. Although easier said than done, the only realistic solution is to forge an alliance between traditional research funding bodies such as the National Institute of Mental Health and the public and private agencies that already provide some version of these services to the general population. This cooperation is the essence of successful services research, and readers interested in developing these unlikely yet necessary alliances will find that much has already been written about the process (e.g., Attkisson et al., 1992).

A less ambitious option would be to select participants with predefined comorbidity combinations (e.g., depression and substance abuse) and require research therapists to address the treatment and associated clinical issues important to both problem areas. Developing an assessment, triage, and intervention protocol for this circumscribed sample would be considerably easier. Although less realistic and less integrated than the totally comprehensive approach described first, this nonetheless represents a significant advance over existing efficacy studies.

Despite the difficulty of mounting an integrated research protocol, some investigators have made initial inroads toward developing such a model. For instance, the FAST (Families and Schools Together) Track Program (Conduct Problems Prevention Research Group, 1992) provides a research-based intervention model for the prevention of conduct disorder in youths by integrating family, school, peer group, and child intervention components. However, even this model is limited in terms of encompassing other disorders that may be comorbid with conduct disorder (e.g., depression).

Historically, my own investigations of adolescent depression treatment and prevention (see Clarke et al., 1995) have similarly taken a relatively narrow focus. However, future investigations could broaden both the sample and the intervention to address common comorbid diagnosis combinations, following the less ambitious of the two paradigms suggested earlier. Because alcohol and drug abuse—dependence is the most common comorbid DSM-III-R (Diagnostic and Statistical Manual of Mental Disorders [3rd ed., rev.; American Psychiatric Association, 1987]) diagnosis for adolescent depression (Rhode, Lewinsohn, & Seeley, 1991), focusing on this combination is a reasonable starting point. An omnibus, school-based, adolescent depression and substance abuse preventive intervention might combine the best features of the depression prevention program (Clarke et al., 1995) with aspects of successful substance abuse prevention programs such as the Midwestern Prevention Project (Pentz et al., 1990). One interesting study design among many possibilities is the implementation of both the separate and the combined prevention programs in the high school setting, with at-risk youths randomly assigned to depression-only, substance abuse-only, or the combined program. This design could be crossed with an "implementation" independent variable, with the programs administered either by school counselors (who would be obliged to provide all other services required by these youths) or by "single-purpose" research therapists (with no obligation to provide associated services). This design would contribute to the theoretical question of shared versus separate etiologies and treatments of depression and substance abuse, as well as address pragmatic issues of outcome as a function of therapist treatment obligations and their degree of integration in the service setting.

As stated earlier, this design example is not meant to capture all elements related to the issue of integrated versus isolated interventions but provides just one concrete example of how efficacy trials could extend their methodology to address effectiveness issues. Rather than representing a relaxing of research standards, this design is meant to be an example of extending efficacy research rigor to topics that previously have been studied (if at all) with less careful and controlled methods.

**Usual Care Versus No-Treatment or Placebo-Attention Controls**

The sine qua non (Parloff, 1986) of psychotherapy efficacy research is the randomly assigned control condition. Several variations exist, but the most common are the enforced no-treatment control, the waiting-list control, and the placebo-attention control. In the no-treatment and waiting-list control conditions, participants are prohibited from obtaining an active intervention altogether or for some predefined period, respectively. Participants enrolled in a placebo-attention control condition are provided with some structured activity believed to be therapeutically neutral (at least with respect to the theoretical model underlying the experimental treatment), in an attempt to control for the nonspecific aspects of interpersonal contact and the number of service hours provided in the experimental condition.

Although these control conditions help resolve theoretical issues and control threats to internal validity, they are usually
unteleable in clinical settings appropriate for effectiveness trials (Weisz, Weiss, & Donenberg, 1992). Furthermore, they do not provide outcome information specifically applicable to the extension of the treatment to a clinical care setting; that is, none of these control conditions represent what typically happens to clients who seek treatment but are not provided with it. If the problem is severe and chronic enough, these individuals often go elsewhere to obtain meaningful treatment.

These control conditions generate other problems. Participants drop from enforced no-treatment and waiting-list control conditions can often be more pronounced than dropout from the active intervention, contributing to potential bias in comparisons of the retained control sample and the experimental condition. The placebo-attention control condition is also problematic in that it may not be as therapeutically neutral as advertised. Elkin et al. (Elkin, Parloff, Hadley, and Autry, 1985), Strupp (1977), and Parloff (1986) argued that most placebo-attention conditions include elements (e.g., a feeling of being understood, an opportunity for social contacts) which would be considered therapeutically active by at least some intervention modalities (e.g., client-centered therapies).

In our recent outcome studies (Clarke et al., 1995; Clarke & Hornbrook, 1994) we have shifted away from these traditional control conditions to a randomized usual-care control condition, identified as a “minimal-treatment” control by Weiss and Weisz (1990). In this condition, subjects are provided with mental health services typically offered in the service setting. A similar control condition, the “best alternative treatment,” compares the experimental therapy against the best available treatment, if such exists (O’Leary & Borkovec, 1978). These designs are all subsumed under the comparative outcome study design, the relative advantages and disadvantages of which are discussed by Basham (1986) and Kazdin (1986).

What impact would the use of this control condition have on outcome research? Weiss and Weisz (1990) reported in their meta-analysis of child psychotherapy studies that the outcome effect sizes associated with the minimal-treatment control group is indistinguishable from that associated with other, more traditional control conditions. From this data at least, using this type of control condition appears to neither endanger the internal validity of controlled outcome studies nor alter estimates of beneficial effect associated with the experimental intervention.

Given these cautions, what (if any) benefits are associated with a usual-care control condition that make it superior to traditional controls in effectiveness trials? I believe that the most important benefit of a usual-care control condition is that it represents a more generalizable test of the intervention. It has the greatest ecological validity of all common control conditions, with the greatest likelihood of corresponding to a real-world counterpart. Although a waiting-list control seems realistic, in my experience substantial numbers of these participants may surreptitiously seek other treatment and eventually drop out of the study. Enforced no-treatment and attention-placebo controls also have limited real-life counterparts.

More pragmatically, I agree with Weisz et al. (1992) that a comparative treatment design is more likely to be tolerated in clinical settings and, thus, overcome clinic staff resistance to effectiveness studies on their premises. Providentially, a usual-care control provides a more conservative test of the experimental intervention, as the new intervention must exceed the benefits associated with usual mental health care to emerge as successful, with the presumption, of course, that usual care imparts at least a minimum benefit. However, a control or comparison condition of this type should not be used uncritically. For example, the usual-care condition makes it much more difficult to characterize the services received by control participants, and in many situations this is important to assess. Kazdin (1986) provides a thorough review of the advantages and disadvantages of each type of control condition, and investigators should carefully examine these issues during planning for any outcome trial.

Sample Representativeness, Heterogeneity Versus Homogeneity

As noted by Weisz and Weiss (1989), controlled outcome research studies typically use very strict participant inclusion criteria, resulting in a sample that may have only a limited resemblance to the usual cases of that disorder served in the community. For example, in previous adolescent depression treatment research (Lewinsohn et al., 1990) my colleagues and I limited the sample by restricting (but not eliminating) psychiatric comorbidity, requiring minimum reading levels, placing a moratorium on all other mental health treatment (or treatments), and placing other restrictions on sample characteristics.

Although there are compelling conceptual reasons to select a highly homogeneous patient sample (see Kendall & Lipman, 1991), the downside of homogeneity is that it yields patients that may be very different from their nonresearch counterparts, making generalization of results suspect. Pragmatically, oversampling may also make participant recruitment more difficult. For example, local referral sources reported having many depressed adolescents to send to depression treatment studies, but only a few that met the stringent selection criteria.

In contrast, effectiveness studies must deal with the viscidities of the clinical world, where comorbidity is common, connection with multiple providers of therapy or social services (or both) is the norm among certain populations (e.g., children; severely mentally ill individuals), placing limits on other treatments is usually neither possible nor ethical, and participant characteristics are generally much more heterogeneous.

How might researchers design different studies to satisfy the goal of a homogeneous sample and also examine the intervention for effects among a more realistic and heterogeneous population? Although it is hardly the only approach, I recommend broadening recruitment for most all intervention outcome studies with a two-tiered participant recruitment strategy; I call this the “donut” model. A highly selected, homogeneous core sample (the “donut hole”) could be recruited for testing basic theo-

---

2 This differs somewhat in our prevention trials (Clarke et al., 1995), where participants not yet meeting a clinical diagnosis are less likely to seek enrollment in treatment. Under these circumstances, usual-care participants are free to continue with any preexisting intervention or to seek any new assistance during the study period if they so desire. To equate the base level of noneffectual intervention across both conditions, participants enrolled in the active prevention program are also permitted to continue any preexisting treatment and to seek out any additional treatment.
retical issues regarding outcome, not dissimilar from efficacy study recruitment as presently practiced. However, to this I recommend adding a relatively unselected, comorbid, and heterogeneous sample (the donut ring) would be recruited to examine generalizability and real-world effectiveness. Participant heterogeneity or homogeneity could be used as a dichotomous blocking variable in a factorial design, crossed with experimental condition, or it could be examined in the full sample by means of post hoc multivariate analyses of the mediating effects of several client and environment variables that differ across the two subsamples.

Increasing the heterogeneity of samples may be justified by the increased generalizability of the results; it also rests on the acknowledgment that many psychotherapies are potentially applicable with more than just one diagnosis. For example, cognitive therapy has been successfully used with depression (Beck, 1991), anxiety disorders (Butler, Fennell, Robson, & Gelder, 1991; Chambless & Gillis, 1993), and eating disorders (Fairburn et al., 1991; Wilson & Fairburn, 1993). Effectiveness studies of cognitive therapy may justifiably recruit more broadly within the broadband categorizations of overcontrolled or internalizing diagnoses, such as those listed earlier, and still hypothesize successful outcomes. The same approach could be used with the undercontrolled or externalizing disorders. This strategy not only eases recruitment, a stumbling block to developing the large samples required in outcome studies, but simultaneously broadens sample heterogeneity. As long as diagnostic and other participant characteristic data is carefully assessed and recorded, post hoc regression analyses can be used to examine the effects of these characteristics on outcome.

Sample heterogeneity may also be increased by including different severity levels within the same problem area or symptom constellation, including individuals who may not meet a DSM-III-R diagnostic category but who have some subdiagnostic syndrome consisting of a reduced set of symptoms within the same category. Expanding samples in this manner may be important, as many individuals seeking service for mental health problems do not appear to qualify for a full DSM-III-R diagnosis yet may still be clearly impaired. For instance, Johnson, Weissman, and Klerman (1992) reported that in the general adult population, as much service burden and health impairment (or more) was associated with subdiagnostic depressive symptoms as with the clinical diagnoses of major depression or dysthymia. These data suggest that it is important to collect outcome data for these individuals as well as those who qualify for the full diagnostic categories.

Mark Hornbrook and I (Clarke & Hornbrook, 1994) are currently conducting just such a study. Adolescents at risk for depression by virtue of having a parent being treated for depression (see review by Downey & Coyne, 1990) are carefully assessed and then triaged to one of three severity levels: (a) clinically depressed adolescents (major depression, dysthymia, or both); (b) at-risk adolescents (elevated but subdiagnostic depressive symptom levels, past depressive episodes, or both); and (c) resilient adolescents (no current depressive symptoms or history). Randomized outcome trials are conducted at each severity level, with increasingly more intensive psychotherapeutic interventions (corresponding to clinical severity) contrasted against a usual-care control condition. Because this study is conducted within a large health maintenance organization (HMO), the costs of delivering experimental interventions are relatively easily measured and compared with the costs of all health care services consumed by the control group. Although the aims of the study go beyond the issues raised in the present article, the participant recruitment and triage methodology has been influenced by a broadened perspective of whom might benefit from psychotherapeutic interventions.

My experience working with schools and public agencies in psychotherapy outcome trials suggests that a broadening of the eligible pool of participants would simplify and enhance recruitment rather than complicate it. Because referring agencies do not have to cul out just the "pure" cases, the referral process is a less time-consuming and frustrating task. Another advantage of extending the sample in this way is that client and psychopathological characteristics, process variables, and other potential outcome mediators may have greater variability than is typically the case with a more homogeneous sample. Up to a point, this increased variability may potentiate multivariate analyses examining the effect of these mediators on treatment outcome across the combined sample, an often desirable post hoc analytical strategy to help identify important therapy issues and directions for future research.

Treatment Parameters: Dosage, Modality, Location, Implementation

This may be the most neglected yet potentially most important area of effectiveness research. After posing the general question "Does psychotherapy work?" most legislators and policy makers in the current national health care debate focus on the effectiveness of multiple variations in the delivery of efficacious intervention. For example, are patients with severe major depression better treated in inpatient or outpatient facilities, and at what costs? For how long? In groups or individual therapy, or both? Delivered by professionals or paraprofessionals? At present, there are few clear answers to these questions.

These are not just pragmatic (read "nonscientific") issues; they represent exciting questions that can have important and fundamental theoretical implications. For example, the relative benefit of group versus individual psychotherapy is obviously relevant from an effectiveness perspective; in service systems such as HMOs such a finding could have major implications for mental health service delivery (see Budman, 1992). However, studies examining this seemingly pragmatic issue can also address numerous theoretical issues related to therapeutic change. For instance, in several models of therapeutic process and change (Orlinsky & Howard, 1986) it is hypothesized that working through the dynamics of a developing therapist-client relationship is a significant contributor to positive treatment outcome. This relationship is presumably optimized in individual therapy and could be argued to be proportionally much weaker in group therapy, in which the therapist's attention is divided across many group participants.

---

3 Child and adolescent examples of overcontrolled disorders include depression, anxiety, and eating disorders; examples of undercontrolled conditions include attention-deficit hyperactivity disorder, conduct disorder, and oppositional defiant disorders.
However, suppose that future research were to find that individual and group versions of the same treatment approach produce roughly equivalent beneficial outcome (Tillitski [1990] and Weisz, Weiss, Alicke, & Klotz [1987] review these issues in child and adolescent psychotherapy). This hypothetical finding might suggest that a combination of client-to-client relationships in combination with a weaker therapist–client relationship is equivalent to an intensive therapist–client relationship, at least as far as psychotherapeutic benefit is concerned.

The key issue here is not whether this is a correct interpretation (this is, after all, only an example based on hypothetical findings), but that investigations of these more pragmatic implementation issues may also lead to a more thorough understanding of theoretical treatment models. In short, studies need not address only basic research or pragmatic issues; they may be designed to address both.

Conclusion

In summary, this article calls for the inclusion of methodological features to transfer desirable aspects of efficacy research (e.g., greater independent variable control) into combined efficacy–effectiveness trials. This message parallels several earlier calls for accelerated study of psychotherapy process variables (summarized by Marmar, 1990). However, the articles in this special section are proposing a different set of variables than those that have been the major focus of process researchers such as Orlinsky and Howard (1986). Traditionally, psychotherapy process research has focused on client and therapist characteristics and interactions, which is reasonable given the focus on how psychotherapy works. In contrast, the blending of psychotherapy efficacy and effectiveness approaches is better served by detailed study of the mediating effects of variables such as the setting in which services are delivered (e.g., school vs. clinic vs. home), the type of clinician who delivers these services, and other issues addressed earlier.

Marmar (1990), in a review of psychotherapy process research, argued for the value of embedding substudies of process variables within larger clinical (efficacy) trials. I agree with this position, but I argue that investigators must look beyond traditional therapeutic relationship variables and include what Kazdin (1986) calls treatment parameter variables, such as the frequency and duration of sessions, the setting in which the treatment is offered, therapist training and profession, and other parameters discussed in this article.

Such studies may manipulate these factors as independent variables or allow them to vary naturally and examine their effect on outcome with post hoc multivariate analyses. The first, more controlled approach is more likely to appeal to efficacy researchers, whereas the latter approach may be more acceptable to confirmed effectiveness investigators. Regardless, both groups (if researchers do self-identify into one or the other group) should be encouraged to generate hybrid studies, broadening the sample and issues under study to address effectiveness concerns while still maintaining as much rigor and experimental control as possible to eliminate or minimize competing explanatory hypotheses. In short, efficacy versus effectiveness is a somewhat artificial distinction, an unnatural dichotomy that a new generation of hybrid studies may help to break down.

References


Clarke, G. N., & Hornbrook, M. (1994). Prevention of depression in adolescent offspring of parents enrolled in a HMO. Grant application funded by the Services Research Branch of the National Institute of Mental Health [R01-MH51318-01A1].


Received November 20, 1993
Revision received December 22, 1994
Accepted December 22, 1994