

A Stage Model of Behavioral Therapies Research: Getting Started and Moving on From Stage I

Bruce J. Rounsaville and Kathleen M. Carroll, Yale University

Lisa S. Onken, National Institute on Drug Abuse

The progressively rigorous methodological requirements of conducting clinical trials of behavioral treatments has placed a large burden on individual investigators, as treatment manuals, methods of evaluating treatment quality and fidelity, and persuasive evidence of the treatment's promise are now virtual requirements of receiving support for conducting a clinical trial of a new or adapted treatment. A Stage Model of Behavioral Therapies research, by articulating the progressive stages of development and evaluation for behavioral treatments, recognizes the scientific merit and need for support for treatment development and initial evaluation designated as stage I. This article describes the conduct of stage I research, including issues addressed in stage I research, major design decisions confronted by investigators, the close relationship of stage I to stage II research and proposes a time line for stage I research.

Key words: clinical trials, behavioral therapies, stage model, study design, methodology. [*Clin Psychol Sci Prac* 8:133–142, 2001]

WHY DO WE NEED A STAGE MODEL?

Research on efficacy and active components of behavioral treatments was revolutionized 20 years ago by the articulation of the technology model of research as exemplified by the National Institute of Mental Health Treatment of Depression Collaborative Research Program (Carroll & Rounsaville, 1990; Waskow, 1984). Akin to specifying the formulation and dosage of medications in Food and Drug

Administration standard pharmacotherapy trials, this approach has generated methods for specifying the behavioral techniques to be evaluated, for training therapists to use techniques consistently, and for monitoring the actual delivery of these techniques over the course of clinical trials.

Although the technology model approach has increased internal validity, generalizability, and replicability of psychotherapy studies, it has placed a large burden on individual investigators whose goal is to improve available treatments by refining current methods or developing new therapies. For example, before fully developed efficacy studies can be undertaken, investigators must have on hand a number of elements now considered requirements of efficacy research, including (a) training manuals specifying techniques to be used and excluded, (b) training programs to impart the requisite skills to study clinicians, (c) process measures to evaluate therapists' competence and adherence to manualized guidelines, and (d) preliminary findings on the acceptability and promise of the new or revised therapy.

Recognizing the large amount of preparation required to conduct full-scale efficacy trials, Onken and colleagues (1997) proposed a Stage Model of Behavioral Therapies research, demarcating three divisions in a rigorous scientific process that leads from initial clinical innovation through efficacy research to effectiveness research. Stage I consists of pilot/feasibility testing, manual writing, training program development, and adherence/competence measure development for new and untested treatments. Stage I can also involve the incorporation of basic behavioral research into research on the development of new behavioral interventions. Such research may have the dual goals of understanding behavioral change process as well as developing interventions to promote positive change

Address correspondence to Bruce J. Rounsaville, M.D., VA Connecticut Healthcare, 950 Campbell Ave., Mail Route 151D, West Haven, CT 06516. Electronic mail may be sent to bruce.rounsaville@yale.edu.

processes. Stage II initially consists of randomized clinical trials (RCTs) to evaluate efficacy of manualized and pilot-tested treatments which have shown promise or efficacy in earlier studies. Stage II research can also address mechanisms of action or effective components of treatment for those approaches with evidence of efficacy derived from RCTs. Stage III consists of studies to evaluate transportability of treatments for which efficacy has been demonstrated in at least two RCTs. Key stage III research issues revolve around generalizability (i.e., will this treatment maintain effectiveness with different practitioners, patients and settings?); implementation issues (i.e., what kinds of training, by what kinds of trainers are necessary to train what kinds of clinicians to learn a new technique?); cost effectiveness issues (i.e., compared with the costs of learning and implementing this treatment, what are the savings, particularly in comparison to existing methods?); and consumer/marketing issues (i.e., how acceptable is a new treatment to clinicians, patients and payers outside of research settings?). This sequence of stages was designed to facilitate the systematic development of promising treatments from the point where they are merely good ideas to one where they are capable of being disseminated to the clinical field as validated, effective, well-defined treatments with guidelines for choosing the patients, providers, and settings most associated with optimal outcomes.

Perhaps the most innovative aspect of this stage model is the recognition the systematic, scientific study of behavioral therapies does not begin or end with the conduct of internally valid RCTs conducted in stage II. Stage I activities of manual development, therapist training, assessment of therapist adherence, and competence and feasibility testing are research tasks that are guided by scientific standards (e.g., development of process measures following psychometric principles). Likewise, stage III research issues are crucial for the process of bridging the much-noted gap between research findings and clinical practice. Most important, research support is required to encourage therapeutic innovators to prepare new treatments for rigorous efficacy testing. This is significant because the rigorous methodological requirements inherent in the technology model had the inadvertent effect of discouraging novel approaches and narrowing the range of therapies being evaluated to those for which manuals, training programs, and process measures were either already available such as cognitive therapy (Beck, Rush, Shaw, & Emery,

1979) or interpersonal therapy (Klerman, Weissmann, Rounsaville, & Chevron, 1984) for depression or could be readily operationalized such as contingency management treatments (e.g., Stitzer, Iguchi, & Felch, 1992). If financial support could be obtained for stage I work, the field would be opened to a richer, more diverse range of new behavioral treatment approaches because fully developed manuals, training programs, and adherence measures would no longer be part of the minimum design criteria. As such, support is reserved only for research projects with scientific merit; the articulation of the stage model is intended to clarify the scientific nature of work conducted at stages I and III. Following early definitions of the stage model, the National Institute on Drug Abuse (NIDA) and the National Institute on Mental Health (NIMH) have both announced the availability of research support for what is designated as stage I research.

Although the stage model seeks laudable goals of encouraging innovation and facilitating more widespread use of empirically validated behavioral treatments, many questions remain about the standards by which stage I proposals should be judged, the nature of the work to be completed in stage I and the criteria for moving out of stage I by either beginning stage II, declaring the new treatment unsuccessful or continuing stage I research with a revised treatment. To address these questions, the NIDA's Treatment Research Branch held a series of workshops on January 24–25, 1995, October 21–22, 1996, and July 12–13, 1998. All grantees conducting stage I NIDA projects were invited to attend as well as expert consultants and NIDA staff.¹ Emerging from these meetings was a listing (displayed in Table 1) of key elements of behavioral therapies needed to begin work at stages I–II. The current article is drawn from discussions at these meetings and is intended to describe guidelines for work taking place at stage I, to delineate optional substages within stage I research, and to describe the transition from stage I to stage II. We believe that these guidelines are not relevant only to development of behavioral therapies for drug abuse but apply generally to the early stages of behavioral therapies development and testing generally.

HOW STAGE II DEFINES STAGE I

As with the evaluation of new pharmacological treatments or medical procedures before approving widespread use, the centerpiece of behavioral therapies development is formal testing of efficacy and safety most

Table 1. Steps in new treatment development

When submitting a proposal	for Stage ^a		
	Ia	Ib	II
Treatment method			
Specify theoretical rationale (theory of the disorder)	✓	✓	✓
Specify hypothesized causal chain (theory of change mechanisms)	✓	✓	✓
Demonstrate feasibility/describe feasibility plan (e.g. acceptability, safety)		✓	✓
Specify process measures (operationalize causal chain)			✓
Provide a provisional therapist manual specifying procedures		✓	✓
Specify pilot testing procedures (if applicable)		✓	✓
Provide a completed therapist manual specifying procedures			✓
Therapist			
Specify and justify inclusion criteria (requirements) and how measured			✓
Establish availability of needed therapists		✓	✓
Specify procedures for assigning cases to therapists		✓	✓
Specify procedures for training and certifying therapists		✓	✓
Specify procedures for supervising and monitoring performance	✓	✓	✓
Establish feasibility of training (replicability)			✓
Specify therapist adherence measures			✓
Establish reliability of therapist adherence measures			✓
Specify therapist competency measures and their reliability			✓
Client/participants			
Identify target population (heterogeneity/homogeneity)	✓	✓	✓
Specify and justify, inclusion/exclusion criteria		✓	✓
Establish capability to recruit the needed sample			✓
Document retention rate in treatment			✓
Design and analysis			
Specify measures	✓	✓	✓
Establish reliability of new outcome measures			✓
Report outcome data			✓
Specify control conditions (experimental of quasi-experimental)		✓	✓
Specify procedures for assigning clients to treatment		✓	✓
Demonstrate feasibility of control			✓
Specify a priori hypotheses (outcome, predictors, process, etc.)		✓	✓
Specify analysis plan to test hypotheses		✓	✓
Demonstrate necessary statistical power for analyses			✓
Specify efficacy/safety monitoring point for continue/stop		✓	✓

^aStages of research: Ia = feasibility testing, early development and refinement of treatment procedures; Ib = “tinkered” pilot testing of treatment outcome; II = efficacy testing; may include a tailored pilot efficacy-testing phase.

typically by means of RCTs. Internal validity is the major emphasis of such trials and the impetus for methodological requirements of the technology model of behavioral therapies research (Carroll & Rounsaville, 1990; Waskow, 1984). The stage model represents an orderly progression from feasibility to efficacy to transportability research (Table 1 provides a summary of research emphases in stages I–II). Establishment of efficacy is the criterion by which it can be determined for a new treatment that fur-

ther research efforts are warranted. Thus, many important aspects of research on a new treatment are given lower priority until efficacy has been established. Once efficacy is demonstrated, research on a new treatment’s “effective ingredients” or applicability in a range of diverse populations and settings can then receive major attention and support.

It should be noted that the pivotal position of RCTs in behavioral therapies research has been criticized as piecemeal (Garfield, 1998), insufficiently attentive to evaluating clinical theory and mechanisms of therapeutic change (Persons & Silberschatz, 1998) and inapplicable in real world settings (Seligman, 1995). These criticisms would be warranted if efficacy testing in RCTs were seen as the end product of the research. Instead, RCTs represent an essential hurdle a treatment must clear to justify subsequent research on its transportability, robustness, and mechanisms of action. Thus, questions regarding how and why a treatment is effective are not ignored in the model but are addressed only for treatments with established efficacy.

Given the central role of stage II research, the ultimate goal or product of stage I research is to develop the elements required to test the efficacy of a new behavioral therapy in an RCT. Of course, an alternative and likely outcome of much stage I research are pilot trial findings which fail to demonstrate the feasibility or promise of the new treatment being developed. In such circumstances, the investigator must choose between discontinuing work on the new treatment or modifying it for further stage I evaluation on the basis of emerging observations and findings.

THE SEQUENCE OF STAGE I RESEARCH

The list in Table 1 of the 32 design elements needed (or at least addressed) to begin a stage II trial is likely to appear daunting and raises questions about an efficient sequence of work within stage I that would allow the transition to stage II within a timeframe justified by the untested nature of the new treatment under study. We address the following general issues for sequencing stage I research: substages of stage I, and timing of attention to transportability issues.

Substages of Stage I

All of the proposed products of stage I research can be narrowed down to two key phases of work: stage Ia, focused on therapy development and manual writing, and stage Ib, focused on pilot testing of a final or nearly final

version of the therapy. All other activities hinge on the completion of a working version of the treatment manual. As long as major aspects of the treatment (e.g., length, session content, group vs. individual format) remain undeveloped or unspecified, definitive training of therapists, development of adherence measures, or pilot efficacy testing cannot take place.

Stage Ia: Therapy Development/Manual Writing. Therapy/manual development is at the heart of stage I work and must precede other steps. This is the most creative aspect of psychotherapy development work and one for which it is difficult to specify procedures, guidelines, or even general parameters of this work, such as length of the therapy development phase or the number of cases to be evaluated. The therapy development stage of behavioral therapies research is analogous to the hypothesis-generating stage of experimental scientific research. Although hypothesis testing is generally guided by formalized rules and procedures, hypothesis generation can be based on insights drawn from any source, such as basic behavioral research, cognitive neuroscience, or clinical observations.

Likewise, new treatment approaches can arise from incremental modifications of existing methods or from completely novel conceptualizations of the therapeutic change process. The open-ended nature of early therapy development activities is underscored by the small number of minimal entry requirements for stage Ia research listed in Table 1. As noted there, the requirements for starting stage Ia research consist only of a theoretical rationale for the disorder and for the new treatment's change process, identification of a target population for which the treatment is indicated, and specification of measures to be used in initial evaluation of the treatment. Given widespread availability of numerous standard treatment approaches, an implicit additional requirement is a description of how the new treatment differs from and may improve upon available treatments.

As noted above, the major essential end product of stage Ia is a provisional or working version of a therapist manual specifying treatment procedures. Carroll and Nuro (1997) have provided guidelines for the many areas to be addressed in a fully developed training manual. These include specifying the rationale for the new treatment, specification of unique and common elements of the therapy, decision rules for choice among alternative interventions, description of interventions excluded from

the new approach, and specification of key treatment parameters such as frequency and duration of treatment. The content of the manual is typically arrived at through a process of initially outlining key elements of the new treatment and then evaluating and elaborating the treatment process and outcome of an open series of patients. Typically, the therapists in this early phase are the investigators/manual developers themselves, and the decisions regarding what methods to keep or discard are based on clinical judgment about what is achieving desired results on a small series of what are hoped to be paradigmatic cases. Although later stages of therapy development are subject to empirical verification, there is no substitute for clinical judgment at this stage because the number of treatment elements to be specified (e.g., group vs. individual treatment, session length, topics addressed, sequence of sessions) greatly exceeds the number that can be systematically varied within a reasonable time frame.

While the therapy is being tried out and modified in an initial series of cases, work on other elements required for stage II testing can begin. Specifically, training materials can be prepared by videotaping sessions with the aim of selecting sessions exemplifying successful or unsuccessful attempts to apply key therapeutic techniques and strategies. Therapist adherence and competence measures can be developed in draft form, and initial reliability testing can take place by performing ratings on videotapes of these early cases.

The duration and scope of work to be completed in stage Ia depends in large part on the level of the therapy's development at the commencement of grant-supported efforts. Many stage I studies begin with fully developed treatments based on years of nonexperimental clinical experience or on comparatively incremental changes required to adopt existing treatments for a particular target population. Among the current group of stage I investigators, some have sought support at a point where a published manual is already available, while others began stage I work with only an outline of the treatment procedures in hand.

Stage Ia activities may also entail extensive features other than therapy/manual development when such features as treatment-specific outcome measures are not available. For example, when a new treatment targets novel aspects of functioning or addresses a treatment group that is not well-characterized, stage Ia work may take the form of needs assessment surveys or instrument

development projects. Some stage I projects have included work on instrument development for outcome measures that have not been systematically evaluated in substance-abusing populations. Alternatively, numerous stage I projects have included initial use of focus groups composed of the target patient groups to provide feedback about the potential feasibility and acceptability of proposed new treatment methods.

Stage Ib: Pilot Trial. Most typically, the major specific aim of an initial stage II trial is to provide a strong test (usually involving a randomized clinical trial) of a new treatment's efficacy. Because of the substantial resources entailed in supporting a full-scale stage II behavioral therapies trial, some empirical promise is essential. This most typically takes the form of a pilot trial of a nearly final version of the new treatment. The aims of such a study are to demonstrate (a) patient acceptance of the new treatment (e.g., retention), (b) the investigators' ability to recruit sufficient numbers of the target population, (c) feasibility of treatment delivery with the proposed types of therapists, patients, and treatment settings, (d) clinically significant patient improvement over the course of treatment in at least one important outcome domain, and (e) the likely effect size to be obtained contrasting the new treatment with a comparison group to be used to determine the sample size for a stage II trial. At the same time, the necessity for a pilot study raises many questions about the scope and design of this aspect of stage I. For example, what should the sample size be? How much can the treatment continue to evolve during the course of the pilot study? Are random assignment and a control condition required? How broad or narrow should patient and therapist inclusion/exclusion criteria be made in a Ib pilot study?

There are no hard and fast rules for making these decisions, but a broad guideline is recognition of the major differences between a pilot and a major trial, which generally relate to the scale of the work and the amount of preparation needed. For example, regarding sample size, although no fixed number can be recommended, a rough guide is that 15–30 subjects per cell is expected, but should be determined by the investigator based on considerations such as choice of control condition and anticipated levels of therapist and subject heterogeneity. Regarding preparation, the contrast between areas checked on Table 1 as desirable for stage Ib and II studies is apparent. For example, for approval of a stage II study,

prior demonstration of patient acceptance, a nearly finalized therapist manual and procedures for training, supervising, and evaluating therapists are virtual requirements. In addition, decisions about the targeted sample (i.e., inclusion/exclusion criteria), acceptable therapist characteristics, comparison group, selection of assessments, and primary outcome measures and data analytic plan need to have been made.

Decisions around ongoing revision of the new therapy during the stage Ib pilot study involve a trade-off between the clinician's desire to make the treatment ever more powerful and the investigator's desire to arrive at an effect size estimate on the treatment as it will be conducted in the stage II trial. Optimally, to provide the most accurate assessment of a treatment's likely effect size, the procedures specified in the training manual should not be substantially modified in the Ib pilot trial (e.g., frequency of use of specific interventions may change, but major defining characteristics of the treatment should not). In other words, treatment manuals should be principle-driven, so that if the treatment is found to be effective, the principles and theories supporting it can (and should) be subject to empirical validation.

The choice of, or necessity for, a control group in a stage Ib trial is complex and reflects the lack of clear-cut standards for choice of comparison conditions in behavioral therapies clinical trials of any size. Investigators using the stage model have used a variety of control conditions for stage Ib pilot trials including historical controls (Pocock, 1976), no treatment/wait list controls, minimal treatment comparisons (Kazdin, 1986), treatment as usual (Borkovec, 1990), a dismantled version of the treatment under study (Kazdin, 1995), or a standard active reference condition (Kazdin & Bass, 1989). Unfortunately, there is no standard behavioral comparison group analogous to placebo treatment used to evaluate new pharmacological agents. Moreover, there are no standard, widely accepted active behavioral treatments of proven (if limited) efficacy against which a new treatment can be compared. Instead, comparison groups, in both the pilot (stage Ib) and major (stage II) efficacy trials need to be tailored to the particular type of behavioral treatment being evaluated and the research questions to be addressed. For example, if the therapy under study represents an enhancement of treatment as usual (e.g., post-hospitalization booster sessions), then initial efficacy testing might involve a contrast between therapy as usual versus therapy as usual plus

enhanced care. However, if a new treatment is conceived of as stand alone care, then the comparison group might involve therapy as usual, a minimal treatment contrast group, or a manualized, clearly defined alternative therapy that the new treatment is designed to improve upon. Choice of a comparison condition in stage Ib is a crucial one because, optimally, the same control condition should be used when efficacy is more definitively tested in a subsequent stage II trial. Investigators should also note that if a manualized active or minimal treatment is to be used as a control condition in the stage II trial, manuals describing this treatment need to be prepared (or obtained) along with therapist training materials and adherence/competence assessments.

Because of the comparatively small scale and attendant small sample size for most stage I research, the issue of patient and therapist heterogeneity must be addressed more on the basis of theoretical considerations than on empirical evidence. In fact, it could be argued that, except for considerations of resource allocation, sample sizes for stage I research should exceed those for stage II because the treatment itself is not at its optimal level of development, therapists' skill in a new treatment is likely to be more variable or reduced by limited experience, and the optimal targeted patient group has not yet been identified. Many potential losses can be imagined in a therapy development program that starts with small, unrepresentative samples as a required antecedent to larger, more costly, more powerful trials. For example, if a treatment is effective only with a subgroup of patients (e.g., cocaine abusers vs. opioid addicts) then exclusion of the optimal patient sample would falsely fail to demonstrate the treatment's promise. A similar false conclusion could be made if the treatment sample is heterogeneous and contains too few subjects in the optimal group to allow a significant effect (or substantial effect size) to be detected for the entire sample. Of course, investigator (and therapist) investment in and enthusiasm for new treatments reduce the power of these factors to mask a treatment effect in early trials. At any rate, because resources for major stage II trials are typically allocated only for those treatments demonstrating initial promise, the current sequence of progressively larger sample size is not likely to change.

Given the limitation of sample size for a pilot study, one commonly used approach entails choosing therapists, treatment setting, and patient sample narrowly to optimize the possibility of detecting a treatment effect and

maximize the power available in a small pilot trial. In addition, a more homogeneous population reduces the risk of imbalance between treatment conditions in major patient prognostic variables, which is particularly likely and highly problematic when randomization is used with small samples (Howard, Cox, & Saunders, 1990; Stout, Wirtz, Carbonari, & DelBoca, 1994). Choice of these optimal conditions can be based on experience with early cases during stage Ia manual development or hypotheses about the mechanisms of action for the new treatment. If the treatment is shown to be effective under optimal conditions, then evaluation of its utility under less restricted circumstances is justified (which fits the stage model's inherent sequence of emphasis from development to efficacy to transportability). Conversely, if treatment effects are not detected even under what are believed to be optimal conditions, then the investigator can have confidence that the new approach has failed to pass a fair test of its efficacy. Of course, most investigators are understandably reluctant to abandon a new treatment approach after investing years of effort in manual development, development of therapy specific process measures, and training therapists for the pilot trial. However, if continued support for this effort is to be sought, the investigator is charged with presenting compelling evidence or reasons regarding negative results from a Ib pilot trial and a substantially altered treatment or study design to be used in further testing.

Weighing against a narrowly focused pilot trial, part of a therapy's promise entails its potentially broad applicability with many different kinds of patients and its use by therapists with a wide range of background and experience. That is, limited initial enthusiasm is likely to be generated for a proposed new therapy that targets a highly restricted patient group and that uses therapists with unusual attributes or extensive training, unless the investigators can make a compelling, theoretically grounded argument otherwise (e.g., a very powerful treatment that is limited to a very small patient group may, in time, lead to understanding of general principles of behavior change). Hence, whenever possible, therapists with the minimal appropriate level of training and experience should be used, and exclusion criteria for patients should be limited to those which are absolutely necessary. Regarding principal drug of abuse, the attendant programs, prognosis, and so on, are quite different for major agents such as opioids, alcohol, cocaine, marijuana,

and nicotine. Moreover, substantially different patterns and intensities of use across drug types (e.g., daily for opioids and marijuana, episodic for cocaine) complicates choice of outcome measures, reduces sensitivity of outcome measures, increases noise, and may lead to masking treatment effects. Hence, at stage Ib treatment across different drug groups may not make sense unless some other type of behavior (e.g., compliance with HIV medications) is the major target of treatment. Applicability of a new treatment to different groups of drug abusers is thus usually an appropriate question for stage III transportability research.

Another issue faced by investigators planning stage Ib and II research is the decision around whether therapists should be nested (i.e., each therapist performs only one treatment) or crossed (i.e., all therapists perform all treatments) across treatment conditions. The advantages and disadvantages of each model have been discussed elsewhere (Crits-Christoph & Mintz, 1991). In stage Ib research, however, this decision usually depends on the choice of control condition selected. In studies where the new treatment is to be compared to treatment as usual or another active treatment, the treatments compared will usually be substantially different in style and technique, and most therapists could not be expected to perform both at high levels of adherence and competence. This usually entails designs where therapists are nested within conditions. In stage Ib studies where the new treatment is a variant of one already used and thus where therapists can be expected to perform both competently, a crossed design has several advantages, including minimization of therapist effects and possible confounding of therapist characteristics with treatment outcome.

Also, regarding treatment setting, it is desirable for a stage Ib trial to take place at a single site to eliminate the potential loss of power involved in statistically managing site differences. Thus, even if few exclusion criteria are named, use of a single site and a single major drug of abuse will most likely place limits on the generalizability of stage Ib trial results. Another type of exclusion criterion pertains to the restrictions placed on allowable additional treatments that the patient can receive while participating in the pilot trial. As we discuss elsewhere (Rounsaville, Weiss, & Carroll, 1999), considerations of internal validity dictate more stringent restrictions on allowable ancillary treatments, whereas less restrictive criteria enhance external validity. In general, given the restricted sample size and

power involved in stage Ib pilot trials, restricting ancillary treatments is desirable in most instances.

Selection of outcome measures is another major issue. As noted above, a major criterion for moving from stage Ib to stage II is evidence of the novel treatment's promise regarding two key outcome variables: retention and drug use. This priority, together with the small sample size and power that characterizes stage Ib work, typically dictates a very focused, pared-down assessment battery, limited to key drug use variables (e.g., urine toxicology screens, frequency/intensity self-report measures) plus a small number of secondary outcomes (e.g., Addiction Severity Index composite scores). At the same time, this approach risks failing to demonstrate promise of the new treatment if a substantially large effect size does not emerge from the one or two primary outcomes selected a priori. In the initial stages of evaluation of a new treatment, it is often not clear which outcomes the treatment may effect. Moreover, a limited assessment battery would preclude process research or detection of changes associated with the theoretic mechanisms of action of the new approach (e.g., acquisition of coping skills in cognitive behavior therapy), which is a major aim of stage I research. Thus, for investigators to understand what outcomes the new treatment may effect as well as whether or not the treatment worked as the investigators hypothesized, a commonly-used strategy is to identify one or two major outcomes measures a priori, but also to include a larger number of exploratory process and outcome measures as a means of permitting a fuller understanding of the new treatments potential effects and mechanisms of action. A related point is that data analytic plans for stage Ib studies are typically modest, limited to simple comparisons of the new treatment and control group on key variables such as retention and drug use as a means of estimating effect size and preliminary process analyses (e.g., demonstrating that the new treatment can be implemented with adequate levels of adherence and competence and can be discriminated from the comparison condition).

A PROPOSED TIME LINE FOR STAGE I RESEARCH

Given the preliminary nature of stage I research, the duration of support likely to be available is restricted to 3 years or less for any given grant proposal (although the investigator is free to propose up to 5 years under the existing grant mechanisms). Within this time frame, the investigator has two general options: to complete both stage Ia and

Ib work together in the same three year study, or initially to propose stage Ia work and subsequently to seek additional support to conduct a Ib pilot trial. The choice between these options can be based on a number of considerations such as the length the treatment being proposed, the amount of prior experience with the new treatment, and the degree to which the new treatment represents a major departure from currently available treatments. For a number of treatments, members of our research group (e.g., Luthar & Suchman, 1999) have been able to successfully complete stage I work in 36 months by dividing the effort into three phases: a 12-month manual development phase; a 6-month therapist training phase; and a 12- to 18-month pilot Ib clinical trial. We briefly describe the activities conducted at each of these phases.

Stage Ia: Phase I Manual Development

A manual development phase can take place in 12 months only if the proposed treatment is highly circumscribed, brief, or straightforward or if the investigators have extensive experience with preliminary versions of the treatment. Typically the work of this phase entails the use, by the senior investigators, of an outline version of the treatment manual with a small number of cases chosen openly and without a control condition. All sessions are audiotaped or videotaped, and patients are evaluated with a preliminary version of an assessment battery to be used in subsequent efficacy evaluations. During this phase, major decisions must be made about content, duration, format (e.g., group vs. individual), therapist credentials, targeted patient groups, treatment context, and countless other features of the proposed treatment. Clinical judgment must be the basis of most of these decisions, although some attempt at systematic variation of a small number of variables (e.g., heroin vs. cocaine abusing patients, group vs. individual format) may be made during this phase. In addition to guiding manual development, clinical work at this phase is important for establishing the initial feasibility and promise of the new treatment. Moving on to therapist training and a stage Ib pilot trial is contingent on demonstrating patient improvement in a portion of pilot cases that equals or exceeds that seen with standard treatments.

While early pilot cases are being conducted, the investigators can complete important methods development work. For example, assessment batteries can be piloted and refined. Items for rating therapist competence and

adherence (Carroll et al., 2000; Waltz, Addis, Koerner, & Jacobson, 1993) can be written and initial psychometric properties (e.g., reliability, internal consistency) can be determined by rating recordings of early treatment settings. Again, because of the need to conduct at least minimal process assessment of both the new treatment as well as the comparison condition, investigators should consider using or adapting scales already developed, (e.g., Barber, Krakauer, Calvo, Badgio, & Faude, 1997; Barber, Mercer, Krakauer, & Calco, 1996) or using more generic scales which tap elements of a number of treatments for the addictions (e.g., Carroll et al., 2000). Training materials can be developed including taped examples of key treatment techniques and strategies and methods for assessing therapist knowledge of the new treatment.

Although development of process measures and training materials can be seen as standard during early stage I, a wide range of other types of research may also be conducted to guide manual development or study design. These include needs assessment activities, evaluating rates of the condition under study, acceptability of the proposed treatment to prospective patients, or methods for detecting and recruiting targeted patients. If the targeted outcomes or change mechanisms of the treatment involve new domains, new instruments may need to be developed to assess these constructs. Laboratory studies may be used to guide choice of particular therapeutic techniques.

Stage Ia: Phase II Therapist Training

Up to this point, a new treatment's developers are typically the only ones who have experience with the therapy. In such cases, it is desirable to demonstrate that the method can be successfully used by a wider group of therapists. As with the pilot trial, this pilot-training program is typically small in scale, with five or fewer therapist trainees. As described elsewhere (Carroll, Kadden, Donovan, Zweben, & Rounsaville, 1994; Crits-Christoph et al., 1998; Rounsaville, O'Malley, Foley, & Weissman, 1988; Weissman, Rounsaville, & Chevron, 1982), essential components of training consist of a nearly final manual, a formal didactic training seminar, and at least one closely supervised training case. A high level of involvement by the investigators is important at this phase, as therapist trainee's experience in using the manual and techniques is invaluable in understanding and further refining the treatment. Therapist self-monitoring of techniques used and not used in the context of training cases (e.g., Carroll,

Nich, & Rounsaville, 1998) can be extremely helpful in focusing supervision sessions as well as bringing to investigators' attention the techniques and interventions which therapist trainees have difficulty using or mastering.

Stage Ib: Phase III Pilot Trial

The aims and characteristics of a stage Ib pilot trial are described above. For treatments of 12 weeks' duration or less and in settings with a patient flow of one or more eligible patient weekly, this trial can usually be completed in 12 months, leaving 6 months to complete data analysis and to finish post-treatment follow-up assessments. Regarding the overall time frame of 3 years for completion of both stage Ia and stage Ib work, practical issues arise if the research team depends on continuous external grant support to remain together. For National Institutes of Health extramural research, applications for renewed support must typically be submitted at least 1 year before anticipated receipt of an award. This means sending in a renewal application only 6 months into a 12-month stage Ib pilot trial, at a point when only around half of the sample is recruited or treated. Alternatively, if the application submission is delayed until the completion of the pilot trial, a hiatus in support of at least 6 months must be managed even if the application is reviewed with a superior priority score.

NOTE

1. Workshop participants: Samuel Ball, Ph.D., Robert Batjes, D.S.W., Alan Bellack, Ph.D., Jack Blaine, M.D., Kathleen Brady, M.D., Ph.D., John Boren, Ph.D., Kate Carey, Ph.D., Kathleen Carroll, Ph.D., John Clarkin, Ph.D., John Curry, Ph.D., Gayle Dakof, Ph.D., Dennis Daley, M.S.W., Bonnie Dansky, Ph.D., John Docherty, M.D., Marc Galanter, M.D., T. Mark Harwood, Ph.D., Steven Hayes, Ph.D., Adam Jaffe, Ph.D., Karen Johnson, Ph.D., Yifreh Kaminer, M.D., Daniel Keller, Ph.D., Herbert Kleber, M.D., Judith Landau, M.B. Ch.B, DPM, William Latimer, Ph.D., Carl Leukefeld, D.S.W., Howard Liddle, Ed.D., Marsha Linehan, Ph.D., Suniya Luthar, Ph.D., Neil McGillicuddy, Ph.D., William Miller, Ph.D., Karla Moras, Ph.D., Lisa Najavits, Ph.D., Lisa Onken, Ph.D., Susan Penava, Ph.D., Paul Pilkonis, Ph.D., Patrick Reilly, Ph.D., Alan Richard, Ph.D., Jami Rothenberg, Ph.D., Lisa Roberts, Ph.D., Bruce Rounsaville, M.D., Daniel Santisteban, Ph.D., Andrew Shaner, M.D., Michael Thase, M.D., Elisa Triffleman, M.D., Roger Weiss, M.D., Elizabeth Wells, Ph.D., David Wetter, Ph.D., Kelly Wilson, Douglas Ziedonis, M.D.

ACKNOWLEDGMENT

Support was provided by NIDA grants P50-DA09241, KO2 DA00248 (to K.M.C.), KO5-DA00089 (to B.J.R.) and the Veterans Administration New England Mental Illness Research Education and Clinical Center.

REFERENCES

- Barber, J., Krakauer, I., Calco, N., Badgio, P., & Faude, J. (1997). Measuring adherence and competence of dynamic therapists in the treatment of cocaine dependence. *The Journal of Psychotherapy Practice and Research*, 6, 12–24.
- Barber, J., Mercer, D., Krakauer, I., & Calco, N. (1996). Development of an adherence/competence rating scale for individual drug counseling. *Drug and Alcohol Dependence*, 43, 125–132.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1990). *Cognitive therapy of depression*. New York: Guilford.
- Borkovec, T. (1990). Control groups and comparison groups in psychotherapy outcome research. In L. S. Onken & J. D. Blaine (Eds.), *Psychotherapy and counseling in the treatment of drug abuse*. (NIDA Research Monograph Series No. 104, DHHS Publication No. ADM 91-1722, pp. 50–65). Rockville, MD: National Institute on Drug Abuse.
- Carroll, K., Kadden, R., Donovan, D., Zweben, A., & Rounsaville, B. (1994). Implementing treatment and protecting the validity of the independent variable in treatment matching studies. *Journal of Studies on Alcohol*, 12 (Suppl.), 149–155.
- Carroll, K., Nich, C., & Rounsaville, B. (1998). Use of observer and therapist ratings to monitor delivery of coping skills treatment for cocaine abusers: Utility of therapist session checklists. *Psychotherapy Research*, 8, 307–320.
- Carroll, K., Nich, C., Sifrey, R., Frankforter, T., Nuro, K., & Rounsaville, B. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*, 57, 225–238.
- Carroll, K., & Nuro, K. (1997). The use and development of treatment manuals. In K. Carroll (Ed.), *Improving compliance with alcoholism treatment* (pp. 53–72). Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.
- Carroll, K., & Rounsaville, B. (1990). Can a technology model be applied to psychotherapy research in cocaine abuse treatment? In L. S. Onken & J. D. Blaine (Eds.), *Psychotherapy and counseling in the treatment of drug abuse*. (NIDA Research Monograph Series No. 104, DHHS Publication No. ADM 91-1722, pp. 91–104). Rockville, MD: National Institute on Drug Abuse.
- Crits-Christoph, P., & Mintz, J. (1991). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *Journal of Consulting and Clinical Psychology*, 59, 20–26.

- Crits-Christoph, P., Siqueland, L., Chittams, J., Barber, J., Beck, A., Frank, A., Liese, B., Luborsky, L., Mark, D., Mercer, D., Onken, L., Najavits, L., Thase, M., & Woody, G. (1998). Training in cognitive, supportive-expressive, and drug counseling therapies for cocaine dependence. *Journal of Consulting and Clinical Psychology, 66*, 484–492.
- Garfield, S. L. (1998). Some comments on empirically supported treatments. *Journal of Consulting and Clinical Psychology, 66*, 121–125.
- Howard, K., Cox, W., & Saunders, S. (1990). Attrition in substance abuse comparative treatment research: The illusion of randomization. In L. S. Onken & J. D. Blaine (Eds.), *Psychotherapy and counseling in the treatment of drug abuse*. (NIDA Research Monograph Series No. 104, DHHS Publication No. ADM 91-1722, pp. 66–79). Rockville, MD: National Institute on Drug Abuse.
- Kazdin, A. (1986). Comparative outcome studies of psychotherapy: Methodological issues and strategies. *Journal of Consulting and Clinical Psychology, 54*, 95–105.
- Kazdin, A. (1995). Methods of psychotherapy research. In B. Bongar & L. E. Beutler (Eds.), *Comprehensive textbook of psychotherapy: Theory and practice* (pp. 405–433). New York: Oxford University Press.
- Kazdin, A., & Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of Consulting and Clinical Psychology, 57*, 138–147.
- Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. S. (1984). *Interpersonal psychotherapy of depression*. New York: Basic Books.
- Luthar, S. S., & Suchman, N. E. (1999). Developmentally informed parenting interventions: The Relational psychotherapy mothers' group. In D. Cichetti & S. L. Toth (Eds.), *Rochester symposium on developmental psychopathology*, Vol. 9, *Developmental approaches to prevention and intervention* (pp. 271–309). Rochester, NY: University of Rochester Press.
- Onken, L. S., Blaine, J. D., & Battjes, R. (1997). Behavioral therapy research: A conceptualization of a process. In S. W. Henggler & R. Ament (Eds.), *Innovative approaches from difficult to treat populations* (pp. 477–485). Washington, DC: American Psychiatric Press.
- Persons, J. B., & Silberschatz, G. (1998). Are results of randomized controlled trials useful to psychotherapists? *Journal of Consulting and Clinical Psychology, 66*, 126–135.
- Pocock, S. (1976). The combination of randomized and historical controls in clinical trials. *Journal of Chronic Disease, 29*, 175–188.
- Rounsaville, B., O'Malley, S., Foley, S., & Weissman, M. (1988). Role of manual-guided training in the conduct and efficacy of interpersonal psychotherapy for depression. *Journal of Consulting and Clinical Psychology, 56*, 681–688.
- Rounsaville, B., Weiss, R., & Carroll, K. (1999). Options for managing psychotropic medications in drug-abusing patients participating in behavioral therapies clinical trials. *American Journal on the Addictions, 8*, 300–311.
- Seligman, M. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist, 50*, 965–974.
- Stitzer, M. L., Iguchi, M. Y., & Felch, L. J. (1992). Contingent take-home incentive: Effects on drug use of methadone maintenance patients. *Journal of Consulting and Clinical Psychology, 60*, 927–934.
- Stout, R., Wirtz, P., Carbonari, J., & DelBoca, F. (1994). Ensuring balanced distribution of prognostic factors in treatment outcome research. *Journal of Studies on Alcohol, 12* (Suppl.), 70–75.
- Waltz, J., Addis, M., Koerner, K., & Jacobson, N. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61*, 620–630.
- Waskow, I. (1984). Specification of the technique variable in the NIMH Treatment of Depression Collaborative Research Program. In J. B. W. Williams and R. L. Spitzer (Eds.), *Psychotherapy research: Where are we and where should we go* (pp. 32–50). New York: Guilford.
- Weissman, M., Rounsaville, B., & Chevron, E. (1982). Training psychotherapists to participate in psychotherapy outcome studies. *American Journal of Psychiatry, 139*, 1442–1446.

Received May 15, 2000; accepted May 23, 2000.