

Adv. Behav. Res. Ther. Vol. 16, pp. 1-29, 1994.
Printed in Great Britain. All rights reserved.

0146-6402/94 \$24.00
© 1993 Pergamon Press Ltd

FAIR TESTS OF CLINICAL TRIALS: A TREATMENT IMPLEMENTATION MODEL

Kenneth L. Lichstein, Brant W. Riedel, and Rick Grieve

Department of Psychology, Memphis State University, Memphis, TN 38152, U.S.A.

Abstract — We propose a psychotherapy treatment implementation model whereby adequate levels of independent treatment components (delivery, receipt, and enactment) are prerequisite to asserting whether a valid clinical trial has been conducted. The delivery component refers to the accuracy of treatment presentation, receipt refers to the accuracy of the client's comprehension of treatment, and enactment refers to the extent of out of session application initiated by the client. Clinical scientists regularly address one or two of these components, but rarely all three, according to a survey we report. Sources and effects of model deficits, i.e., inadequate levels of treatment components, as well as methods of component assessment and induction, are discussed. We conclude that faults in any one of the components drain validity proportional to the degree of deficit, and that clinical trials have often incorrectly been considered fair tests, resulting in biased efficacy judgments.

INTRODUCTION

Clinical scientists investigating psychotherapy have devoted well-justified attention to matters of research design, subject selection and assignment, dependent measure characteristics, and the like, in their quest to purify methodology and to elevate the confidence due the results of their labors. Motivated by these same goals, a series of articles that appeared about 10 years ago issued a methodological alert concerning an allied but novel facet of clinical outcome trials (Billingsley, White, & Munson, 1980; Quay, 1977; Sechrest, West, Phillips, Redner, & Yeaton, 1979; Sechrest & Yeaton, 1981; Yeaton & Sechrest, 1981). These authors illuminated the critical distinction between the treatment that was intended to be delivered and the actual treatment delivered. Sechrest and his colleagues argued that in the absence of ascertaining that the treatment was delivered as intended, i.e., establishing treatment integrity, partial treatment administration could prevail and lowered judgments of treatment efficacy might correctly belong to some variation of the intended treatment, but would nevertheless be awarded to the intended treatment. Since then, there has been an increasing willingness by investigators to assess treatment integrity and to adopt measures that secure high levels of it, though this has not attained universal status. In many clinical trials we must still guess what treatment was actually tested.

This new line of methodological inquiry shifts attention from experiment-wide concern with molar design issues to a more molecular examination of the conduct of therapist and client within a given treatment to ascertain

if the treatment of interest was given a fair test. Treatment integrity is but one aspect of the administration and consumption of a treatment. The present paper expands the notion of treatment integrity into a full treatment implementation model by tracking the presentation of the treatment from its therapist source to its utilization by the client. Having established that treatment was delivered as intended, final judgments of treatment efficacy are still premature until two additional standards are satisfied: treatment receipt and treatment enactment.

Drawing upon a medical model to dramatize the point, a physician prescribing medicine for a patient's hypertension has satisfied the standard of treatment integrity if the prescription he/she wrote was the one he/she had in mind, and we call this stage of treatment *delivery*. However, if the patient did not (a) fill the prescription, they never received the treatment, and (b) consume the medicine as instructed, they may not have achieved adequate treatment exposure. We will employ the terms *receipt* and *enactment* to refer to steps (a) and (b) above, respectively.

Establishing only treatment integrity does not necessarily justify conclusions as to medicine efficacy. Concerning receipt of treatment, had the patient decided to ignore the physician's advice, but instead gained motivation to more wisely regulate dietary and exercise habits, hypertension improvements might be unjustly attributed to the medicine if the patient's deviation from the plan was not revealed. Concerning enactment of treatment, had the patient actually filled the prescription but then took it irregularly, we might unjustly attribute poor therapeutic response to the medicine.

In clinical research, analogous processes occur, though sometimes these are less distinct. Consider research on the outcome of psychodynamic psychotherapy and failure to correctly implement it, referred to as component deficits. Should the therapist interject other methods of therapy into treatment, i.e., delivery deficits, we are no longer assessing only psychodynamic psychotherapy whether we have knowledge of this alteration in treatment or not. But if the patient was daydreaming, intellectually incompetent to appreciate therapist communications, or distracted by physical discomfort, i.e., receipt deficits, then even had we established that orthodox psychodynamic psychotherapy had been presented, it would be unfair to blame an unsuccessful outcome on the treatment. But still, had the therapy been properly delivered and the patient adequately comprehended it, treatment exposure is so far limited to 1 h out of a 7-day period. We can expect that the patient who dwells upon issues addressed in psychotherapy, who continues to examine and reevaluate personal concerns during the week, and who experiments with new ways of relating to others would differ from a patient who vanquished all thoughts of psychotherapy outside the boundaries of the session with

Treatment Implementation Model

3

respect to rate and magnitude of therapeutic response, were the treatment in fact a potent one.

In more focal treatments, the relevance of this treatment implementation model is less ambiguous. In the case of assertiveness training, if the method is improperly delivered, if it is not understood by the patient, or if it is not rehearsed during the week, an attenuated treatment response is expected.

In the past decade, the psychotherapy methodology literature has made gains mainly with respect to treatment integrity, but has been slow to recognize the other components of the proposed model. Excellent reviews have focused our attention on the importance of establishing the integrity of the treatment (e.g., Kazdin, 1986a,b; Moncher & Prinz, 1991), but do not address receipt or enactment components. Other monographs on clinical research methodology do not discuss any aspect of this model (e.g., Bellack & Hersen, 1984; Garfield & Bergin, 1986).

Kazdin (1980) provided a notable exception to the attention afforded the treatment implementation components of present interest. Under the term manipulation check, he devoted a chapter to delivery and receipt components in which he squarely illuminated the threat to internal validity they pose. Smith and Sechrest (1991) recently reiterated these concerns, as the published psychotherapy research has been slow to take heed. The National Institute of Mental Health Treatment of Depression Collaborative Research Program (Elkin, Parloff, Hadley, & Autry, 1985) provided another methodological landmark. This highly publicized investigation paid meticulous attention to treatment manual preparation, therapist training, and therapist monitoring, and has probably been influential in guiding psychotherapy researchers.

Before proceeding, a brief discussion of the topics of efficacy and of validity will help clarify concerns raised repeatedly in this article. Starfield (1977) and others have emphasized the distinction between therapeutic potency under optimal and suboptimal conditions. Efficacy refers to treatment administration under optimal conditions and defines the full therapeutic potential of the treatment. Effectiveness refers to the same treatment administered under more routine, suboptimal conditions and reflects typical effects that can reasonably be expected. Clinical research usually strives to attain efficacy trials, and these results usually exceed the effectiveness obtained in clinical practice. The present article is, of course, interested mainly in treatment efficacy and one set of factors that threaten it.

According to Cook and Campbell (1979), "one could invoke many types of validity," (p. 37) though most are subsumed under the umbrella terms internal and external validity. The former refers to validity in asserting that a causal relationship exists between covariates, and the latter to the

validity in asserting the generalizability of findings across a continuum of salient independent and dependent variables under varying subject, time, and setting conditions. Treatment implementation model faults may corrupt both internal and external validity, depending on the nature of the fault. For example, undisciplined therapists in Study A may depart from treatment protocols, but therapists in replication Study B do not. Study A delivery faults may thus alter both internal and external validity. Alternatively, homework assignments in Study A, though requisite for satisfactory outcome, are so burdensome as to discourage adequate compliance in all but the most motivated patients. Internal validity will suffer, but not so external validity, since all replication attempts will encounter the same obstacle. In each of these varied circumstances, erosion of validity will escape the perception of even the most trained eye unless component-specific assessments are performed. In the remainder of this article, we shall use the term validity to refer collectively to the family of challenges to scientific inference.

In order to fairly judge treatment efficacy, treatment must be fully implemented. This is not to say that deficits occurring in one or more of the three components of this treatment implementation model would necessarily result in diminished treatment effects. At all times, such deficits would lead to either positive or negative bias in treatment appraisal, save for the random occurrence of therapeutic equivalence between the original component and its revision.

GENERAL CHARACTERISTICS OF THE TREATMENT IMPLEMENTATION MODEL

We postulate that three components must be present as a prerequisite to valid conclusions in clinical trials. Implementation of the three components (delivery, receipt, and enactment) may be linked, but most often are orthogonal. Satisfactory performance of any one of these components provides little insight into the status of the others. Fault in any one of these components, even when the remaining two are fully implemented, may disable therapy and unfairly degrade therapy appraisals. Like a chain pulling its weight, fault need not arise in all three links to compromise its integrity. Such relationships presume a potent treatment is being tested. Less often, departures from planned treatment procedures can unfairly inflate therapy appraisals, as may occur when creative therapists embellish treatment protocols by inserting new treatments. All three components are easily susceptible to diminished execution. It is the responsibility of the researcher to demonstrate adequacy of the full treatment implementation model, and failure to do so renders indeterminate the question of whether or not a fair test of a particular treatment was conducted.

The remaining sections of this paper will elaborate on this treatment implementation model. We will survey respected clinical journals to assess current practice with respect to our methodological recommendations. We will discuss a wide range of factors contributing to treatment implementation deficits and the effects of these. We will examine methods of component assessment needed to determine model adherence and methods of component induction intended to ensure high component adherence or to correct faults when discovered. Finally, summary observations will be offered.

INCORPORATION OF THE TREATMENT IMPLEMENTATION MODEL IN CURRENT METHODOLOGIES

A number of surveys have estimated the incidence of treatment integrity promotion and assessment (Billingsley et al., 1980; Moncher & Prinz, 1991; Peterson, Homer, & Wonderlich, 1982), and these have yielded discouraging findings. Over a 13-year period in the *Journal of Applied Behavior Analysis*, about 20% of the studies reported an attempt to assess the procedural accuracy of the independent variable (Peterson et al., 1982), and this rate dropped (inexplicably) to 5.6% in a 2-year period in the same journal combined with a 1-year period in *Behavior Modification* (Billingsley et al., 1980). More recent data reveal a positive but still unsatisfactory trend (Moncher & Prinz, 1991). Eight journals surveyed over a 9-year period for three aspects of treatment integrity promotion and assessment showed increasing use of these procedures. For example, one of the strongest findings was an increase in the use of treatment manuals to guide therapist performance in the *Journal of Consulting and Clinical Psychology* from 39 to 49.1% over this period. Comparable surveys concerning receipt and enactment assessment are scarce. The only one we discovered surveyed studies of progressive relaxation training in four journals for nearly 3 years to determine if researchers assessed the actual level of relaxation produced by the procedure (Luiselli, Marholin, Steinman, & Steinman, 1979). This corresponds to our receipt component. Counting either self-report, physiological, or behavioral measures, 70% of the studies made no attempt to measure if a relaxation effect was induced.

In order to determine concordance between current methodological practice and the proposed treatment implementation model, we surveyed the most recent full year (1990) of *Behavior Therapy* and of the *Journal of Consulting and Clinical Psychology*. These journals were selected because they are highly respected outlets for clinical outcome research, and because they are regarded as being among the most methodologically rigorous clinical journals.

Table 1. Occurrence of Treatment Implementation Model Components in *Behavior Therapy*, 1990

Article	Delivery				Receipt				Enactment			
	Induct ^a		Assess ^b		Induct		Assess		Induct		Assess	
	Fo ^c	In ^d	Di ^e	In ^f	Fo	In	Di	In	Fo	In	Di	In
Agras et al.	x				x					x		x
Abramowitz & O'Leary ^g	x		x									
Appelbaum et al.	x								x			x
Azrin & Peterson	x				x				x			
Behrens et al.	x											
Boudawyns & Hyer		x										
Carey et al.	x				x		x					
Chambless	x			x								
Christensen & Burrows	x				x				x		x	x
Church et al.	x				x		x					
Keeffe et al.	x		x	x								x
Lerner & Clum	x		x	x			x	x				
Lewinsohn et al.	x		x	x	x			x	x			x
Lowe & Chadwick	x			x ^h								x
Murphy et al.	x											x
Nelms & Feitas	x		x				x		x			x
Renneberg et al.	x								x			x
Roberts & Powers ⁱ	x		x									
Rosen et al.	x			x					x		x	x
Sullivan & O'Leary ^g	x		x									
Viegner et al.	x			x ⁱ	x				x			x
Webster-Stratton & Hammond	x											
Wells	x			x ^j								
Williams & Chambless		x ^k										
Wurtele	x				x		x					

^aInduct records whether the author employed induction strategies to increase the likelihood that satisfactory delivery, receipt, or enactment occurred. ^bAssess records whether the author conducted an assessment of the implementation of delivery, receipt, or enactment. ^cFo records if the author employed formal methods of delivery, receipt, or enactment induction. Examples are written treatment manuals or mock therapy training sessions for delivery, written handouts summarizing key therapy points for receipt, and written homework assignments for enactment. ^dIn records if the author employed informal methods of delivery, receipt, or enactment induction. Examples are verbal reminders to therapists to include all treatment procedures for delivery, verbal prompts to patients to attend to salient therapy ingredients for receipt, and verbal requests to patients to practice therapy techniques for enactment. ^eDi indicates delivery, receipt, or enactment was assessed directly, leaving little doubt as to the certainty of its findings. ^fIn indicates delivery, receipt, or enactment was assessed indirectly, and thereby is unable to erase all doubt as to the certainty of its findings. ^gThese studies involved children who were not expected to assume an active role in the treatment process. Therefore, the delivery component is relevant to the present inquiry, but not the receipt or enactment components. ^hAlthough there was no independent assessment of treatment delivery, all treatment was delivered by the second author. Therapists participated in weekly training sessions, but it is not clear if supervision of actual therapist-subject interaction (which would qualify as treatment delivery assessment) occurred. ⁱAlthough treatment delivery was not independently assessed, all treatment was delivered by the first author. ^jThe article mentions that therapists were "trained and supervised." Insufficient detail is provided to justify scoring this study positive for assessment of delivery.

Treatment Implementation Model

7

Table 2. Occurrence of Treatment Implementation Model Components in the *Journal of Consulting and Clinical Psychology*, 1990

Article	Delivery				Receipt				Enactment			
	Induct ^a		Assess ^b		Induct		Assess		Induct		Assess	
	Fo ^c	In ^d	Di ^e	In ^f	Fo	In	Di	In	Fo	In	Di	In
Baggs & Spence					x					x		
Baucorn et al.	x		x									
Blanchard et al., a	x		x				x		x ^g	x ^h		x ^h
Blanchard et al., b	x		x				x		x			x ^h
Botvin et al.		x	x		x				x ⁱ			
Danforth et al.	x		x ^j				x					
Davis et al.	x				x							
Fonagy & Moran		x		x								
Hartmon et al.		x							x			x
Imber et al.		x										
Jay & Elliott	x ^k				x				x			
Johnson et al.		x			x				x			
Killen et al.	x				x			x				
Kivlahan et al.		x					x			x		x ^l
Klosko et al.			x									
Malgady et al., a		x	x									
Malgady et al., b		x		x								
Manne et al.												
Schinke et al.		x ^m			x		x					
Telch et al.	x									x		x
Turner et al.	x			x	x					x ⁿ		x
Weisz et al.												

^aInduct records whether the author employed induction strategies to increase the likelihood that satisfactory delivery, receipt, or enactment occurred. ^bAssess records whether the author conducted an assessment of the implementation of delivery, receipt, or enactment. ^cFo records if the author employed formal methods of delivery, receipt, or enactment induction. Examples are written treatment manuals or mock therapy training sessions for delivery, written handouts summarizing key therapy points for receipt, and written homework assignments for enactment. ^dIn records if the author employed informal methods of delivery, receipt, or enactment induction. Examples are verbal reminders to therapists to include all treatment procedures for delivery, verbal prompts to patients to attend to salient therapy ingredients for receipt, and verbal requests to patients to practice therapy techniques for enactment. ^eDi indicates delivery, receipt, or enactment was assessed directly, leaving little doubt as to the certainty of its findings. ^fIn indicates delivery, receipt, or enactment was assessed indirectly, and thereby is unable to erase all doubt as to the certainty of its findings. ^gSubjects were given audio tapes to help facilitate home enactment in two of the three treatment groups. ^hOf the three treatment groups, two assessed homework indirectly, and one just reminded subjects to practice homework. ⁱThe treatment was implemented in elementary and junior high class rooms. It is assumed that some type of homework was assigned, but there was no report of how it was assessed. ^jTreatment was assumed to be directly measured; however, there was not enough information reported in the article to be certain. The treatment administered consisted of an exercise program. Experimenters monitored the children utilizing the ergometers, but it was not reported whether they determined if the treatment was delivered correctly. ^kThe treatment was divided into three components. The first was a videotape, which is a formal induction. However, the other two components had no special induction associated with them. ^lOne of the two treatment groups used a self-report measure; the other had no reported homework. ^mOne of two treatment groups consisted of reading a comic book. While this comic book itself was a formal induction, guidelines to run the other group were informal. ⁿSubjects kept logs of homework activity in two of the three treatment groups.

We identified all clinical outcome trials (see Tables 1 and 2), including case studies, single-subject experimental designs, and group studies. Studies were scored for whether or not they took steps to ensure adequate levels of treatment delivery, receipt, and enactment, termed induction, and whether or not they assessed resulting levels of the treatment implementation components.

If the article reported no special effort to attain accurate treatment delivery, adequate treatment receipt by the patient, or adequate out of session enactment, the Induction columns in Tables 1 and 2 were left blank. When induction efforts were reported, they were scored as either Formal or Informal. Formal methods were characterized by written documents or audio/visual aids. Informal methods were characterized by verbal instructions that left more room for variation and ambiguity. Examples of formal induction methods for delivery, receipt, and enactment, respectively, are written treatment manuals (Viegner et al., 1990), comic books to emphasize therapy principles (Schinke, Gordon, & Weston, 1990) and relaxation tapes to guide home practice (Blanchard et al., 1990a). Examples of informal induction methods for delivery, receipt, and enactment, respectively, are role play training of therapists (Lerner & Clum, 1990), verbal prompts to remind clients of key therapeutic responsibilities (Azrin & Peterson, 1990), and encouragement to comply with homework assignments (Kivlahan, Marlatt, Fromme, Coppel, & Williams, 1990).

If the article reported no attempt at assessing the accuracy of treatment delivery, the adequacy of treatment receipt, or the extent of out of session enactment, the Assessment columns in Tables 1 and 2 were left blank. When assessment efforts were reported, they were scored as either Direct or Indirect. Direct methods were characterized by direct observation or objective measurement which yielded findings of high certainty. Indirect methods were characterized by reliance on verbal reports or self-report questionnaires which could not erase doubt as to their accuracy. Examples of direct assessment methods for delivery, receipt, and enactment, respectively, are scoring tapes of therapy sessions (Lewinsohn, Clarke, Hops, & Andrews, 1990), written tests of clients' change in knowledge following therapy (Kivlahan et al., 1990), and saliva samples (Christensen & Burrows, 1990). Examples of indirect assessment of delivery, receipt, and enactment, respectively, are verbal supervision of therapy sessions (Rosen, Cado, Silberg, Srebnik, & Wendt, 1990), inquiry as to clients' confidence in applying therapy skills (Lerner & Clum, 1990), and self-report records of homework performance (Turner, Clancy, McQuade, & Cardenas, 1990).

Behavior Therapy (BT)

BT 1990 contained 25 clinical outcome studies. Of these, the design of 20 fit our full treatment implementation model. For example, some were classroom interventions for which there was no homework component. Only 1 of 20 studies induced and assessed all three treatment implementation components (Lewinsohn et al., 1990).

Considering treatment implementation induction, 5 of 20 studies took steps to induce all three components, though only 4 (20%) of these exclusively employed formal methods (Agras, Taylor, Feldman, Losch, & Burnett, 1990; Christensen & Burrows, 1990; Lewinsohn et al., 1990; Viegner et al., 1990). Delivery induction was routine, occurring in 24 of 25 studies, and performed formally in 23 (92%) of these. Receipt induction was evident in 8 of 22 studies, with 7 (31.8%) studies incorporating formal receipt induction. Enactment induction existed in 9 (45%) of 20 studies, and all of these nine studies utilized formal enactment induction techniques.

Considering treatment implementation assessment, 2 (10%) of 20 studies (Lewinsohn et al., 1990; Neimeyer & Feixas, 1990) assessed all three components, but neither assessed all three components directly. About half the studies (12 of 25) assessed treatment delivery, and 9 (36%) of these assessed it directly. The concept of receipt as utilized in our model was applicable to 22 studies. Six assessed receipt, with five (22.7%) assessing it directly. Enactment was applicable to 20 studies. In 14 of these studies, a specific homework assignment was given to the subjects. The six remaining studies lent themselves to treatment enactment by subjects and such was implied, but explicit homework assignments were not reported in the articles. Eleven (55%) of these 20 studies assessed enactment, though only two (10%) assessed it directly.

Journal of Consulting and Clinical Psychology (JCCP)

JCCP 1900 contained 22 clinical outcome studies. The design of all 22 studies fit our full treatment implementation model. Only two (Blanchard et al., 1990a,b) of 22 studies, 9.1%, assessed delivery, receipt, and enactment, but these studies did not report efforts to induce all three components of the model. Thus, no single study assessed and induced all components of the treatment implementation model.

Considering model induction, 4 (18.2%) of 22 studies induced all three components. Of these, only one (Jay & Elliott, 1990) relied on formal techniques exclusively. Delivery induction was common, occurring in 18 of 22 studies, but was performed formally in only 9 (40.9%) studies. Receipt induction occurred in 8 (36.4%) of 22 studies, and all were formal.

Enactment induction occurred in 10 of 22 studies, and 5 (22.7%) of these utilized formal induction techniques.

Considering model assessment, 2 (9.1%) of the 22 studies assessed all three components (see reference to Blanchard above), though neither of these studies exclusively employed direct assessment. Ten studies assessed treatment delivery, and 7 (31.8%) of these were direct. Receipt of treatment was assessed in 6 (26.1%) of 23 studies, with 5 of the 6 studies using direct measures. Enactment was also assessed in 6 (27.3%) of 22 studies, but none were direct.

Summary Findings

The utilization patterns of the treatment implementation model in the above journals are subject to alternative interpretations. Few of the studies induced and/or assessed all three components, but nearly all of the studies addressed at least one of the components, which indicated awareness, albeit inconsistently, by clinical scientists that these dimensions are worthy of attention. Induction, and to a somewhat lesser extent assessment, of treatment delivery has achieved conventional status and is far more established in methodology practice than are induction or assessment of receipt and enactment, and these conclusions apply roughly equally to the two journals. Apparently, treatment integrity recommendations of 10 years ago exerted an important influence on the course of clinical methodology, and this is certainly heartening.

When assessments of delivery, receipt, and enactment were reported, the findings of the first two were uniformly favorable and the third usually favorable but more variable. We can conjecture four equally plausible explanations for these positive findings. First, researchers who are sufficiently sophisticated to plan this degree of meticulous research are better researchers across the range of research dimensions and simply produce high quality research. Second, given the decision to assess implementation components, researchers strive to produce high quality research that will measure up to conventional standards. Third, when assessments reveal implementation deficits, such as when a treatment other than the one planned is actually delivered, researchers elect not to submit the study for publication. Fourth, journal editors will not publish manuscripts when model deficits are announced. Whatever the reason, the presence of assessments of treatment implementation components is strongly associated with a higher quality of published clinical outcome research.

A final brief comment is worth noting. Scoring articles to compile the data for Tables 1 and 2 was in many cases a difficult task because procedural information was either not reported or provided with in-

sufficient detail. It is possible that our data underestimate the prevalence of our treatment implementation model as a function of concise reporting styles reflecting editorial pressure to conserve journal space and/or disinterest in reporting research in sufficient detail as to permit replication.

SOURCES AND EFFECTS OF MODEL DEFICITS

Departures from satisfactory component implementation, i.e., inaccurate treatment delivery, poor patient comprehension, or poor patient enactment, represent model deficits. The following sections alert the reader to factors promoting model deficits and interpretive problems arising from their presence.

Deficits in Treatment Delivery

When the actual treatment delivered differs from the intended treatment by virtue of additions or deletions, this is said to be a deficit in treatment delivery. The term *treatment integrity* has focused mainly on partial treatment delivery or errors of omission. Kazdin (1986a) employed the term *treatment differentiation* to specify the differences between treatments being compared. Integrity and differentiation are not completely independent, and have recently been joined under an umbrella term *treatment fidelity* (Moncher & Prinz, 1991). As integrity suffers depending on what part of treatment is omitted or what new ingredients are added, distinctions between treatments may fade. Experimenter/therapists may depart from the planned treatment for a multitude of reasons that may be subsumed under the categories of treatment and therapist characteristics.

Sechrest et al. (1979) reasoned that vaguely defined or highly complex treatments invite lapses in treatment integrity more so than treatments undistinguished by these characteristics. Similarly, burdensome or taxing treatments may inspire therapist modification. Therapist preparation and personality characteristics may also undermine integrity. The training provided for the therapist may be inadequate. The absence of a specific training manual and inadequate mock trials could all result in treatment modification. Therapists with an inadequate educational background or who are poorly motivated (Sechrest et al., 1979) for the tasks assigned to them may alter treatments intentionally or accidentally.

Sechrest and his colleagues (Sechrest et al., 1979; Yeaton & Sechrest, 1981) warned of shrinking treatment potency as the delivered treatment departs from planned protocols, but outcome may also exceed that of the planned treatment when partial treatment is presented. Omission of well-conceived portions of treatment could surely drain treatment efficacy, but therapists may also omit unimportant treatment components. Indeed, treatment components will sometimes dilute or even detract from efficacy

and were these omitted, the diminished treatment integrity could produce comparable or even better results than the original treatment.

Treatment integrity may be compromised by still another device. Therapists may insert unplanned treatment components to supplement the intended treatment. In this case, the actual treatment delivered is composed of the entirety of the intended treatment plus additional unintended components, and the supplement could conceivably improve or detract from the original treatment. Kazdin (1980, p. 39) related the case of a comparison group study for patients in psychiatric day care. The therapist inserted techniques from the behavior therapy condition into the milieu therapy condition, and thereby diffused differences between them.

Finally, partial and supplemented treatment may co-occur, wherein the therapist omits part of the intended treatment to create room for his/her innovations. In sum, treatment variations of omission or commission may deplete or enhance efficacy, and final efficacy conclusions may contain a positive or negative bias when treatment is altered. The magnitude of efficacy error is proportional to the degree of departure from the stated treatment, assuming that receipt, enactment and other therapy outcome influences are not deficient. In the absence of treatment integrity assessment, doubt persists in identifying exactly to which treatment to attribute obtained results.

Deficits in Treatment Receipt

Receipt deficits can result from either therapist or patient shortcomings, or from poor communication between the two for which they share responsibility. Therapists may be so handicapped in their ability to communicate that even the most earnestly motivated patient would fail to master substantial portions of the treatment. Patients may be poorly motivated, inattentive, or of diminished intellectual capacity, thereby failing to comprehend the treatment delivered by the most able therapists. At other times, personality factors muddle the interaction between a therapist who usually communicates effectively and a patient who would master the treatment presented by most therapists. In this latter case, receipt deficits would mount, perhaps surprisingly, given the ill-fated match of competent therapists and patients.

Except when a treatment has zero therapeutic potential, receipt deficits will always shrink outcome. If the treatment is completely ineffectual, no additional therapeutic losses will accrue when the treatment goes unperceived. Excepting for the case when it is the nature of the treatment itself that precludes adequate client mastery, as with highly laborious, complex treatments, treatment delivery and treatment receipt are independent phenomena.

Deficits in Treatment Enactment

The client assumes the role of primary change agent in place of the therapist for the entirety of the week outside the customary 1-h therapy session. According to this view, the professional therapist delivers a small fraction of the potential change process and the client the remainder. Noncompliant clients receive abbreviated treatment exposure, and the degree of the abbreviation will vary depending on the nature of the treatment. A number of authors have focused our attention on the crucial nature of this matter. Feinstein (1979) coined the phrase, "compliance bias," referring to the often overlooked impact of home compliance on therapy outcome. He reasoned that comparable therapies may be judged unequal due to differential client enactment. Goldsmith (1979) argued that noncompliant subjects are, in essence, absent from the clinical trial, and statistical power to detect clinical effects will suffer unless the researcher runs extra subjects to compensate for anticipated noncompliers.

The emphasis on the patient as the primary change agent varies greatly across therapies, but must exist to some extent in all therapies. For example, Primakoff, Epstein, and Covi (1986) lamented that homework activities in cognitive therapy for depression are universally recognized as a critically important therapeutic ingredient, but that the salient literature is lax in reporting specific homework assignments or compliance rates. Vermilyea, Barlow, and O'Brien (1984) dramatically demonstrated the relationship between enactment and clinical outcome. They tracked rated anxiety and compliance to anxiety treatment home instructions, and they found differential treatment response in three patients paralleled varying levels of adherence both within and between patients. Other investigators have reported comparable findings (Hoelscher, Lichstein, & Rosenthal, 1984; Kazdin & Mascitelli, 1982; Persons, Burns, & Perloff, 1988).

Relaxation therapy may serve as the prototype for this process. Typically, the therapist induction is explicitly recognized as 1/7 of the therapy, and the prescribed daily home practice of the patient defines 6/7 of treatment exposure. Clearly, poor homework compliance will usually drain therapeutic potency.

Lichstein (1988) identified six common factors contributing to lowered patient relaxation practice. These likely invade practice compliance in other therapies as well:

- (1) *Not enough time.* Busy schedules compete with time needed to practice treatment.
- (2) *Forget.* Patients lament that their good intentions are frustrated by poor memory.

(3) *Unsure of correct procedure.* Outwitted by the complexity of the procedure, patients will abandon practice as frustration mounts.

(4) *Lack of success.* When benefits derived from home practice fall short of expectations, discouragement may depress practice.

(5) *Unsatisfactory setting.* Treatment practice may require setting characteristics or assistance from others not conveniently available.

(6) *Declining interest in therapy.* Innumerable factors will contribute to general disinterest in therapy, and poor homework follow through may be one of the first signs of this development.

ASSESSMENT OF TREATMENT IMPLEMENTATION

In the absence of evaluating the degree to which treatment exposure included implementation components, we cannot determine if the clinical outcome in a particular trial fairly judges the efficacy of the treatment being tested. Thus, not only must a clinical trial include a dependent variable assessing treatment, it must also include additional therapist and client variables to determine what treatment was actually presented, was it adequately perceived, and was it adequately enacted?

Delivery Assessment

This component of our model has received the most attention in the literature. The experimenter must list the key therapeutic components of the treatment being tested and must also identify the particular ingredients of associated treatments that are specifically not part of the treatment in question. From this basis, the experimenter can operationalize those therapist behaviors that must be present in order to certify adequate delivery and those therapist behaviors judged taboo in order to avoid confusing the treatment of interest with others. This process of determining what to assess is itself potentially instructive. It will perhaps stimulate clarification in the minds of some experimenters as to the active components of their treatment.

Treatment delivery ratings of taped therapy sessions are probably the most trustworthy means of assessing delivery. Taping provides a method whereby an independent observer can directly monitor the adherence of the actual treatment delivery process to the intended treatment protocol. Also, taping sessions erases any assessment errors due to recall lapses. Tapes of treatment sessions can be utilized in a number of ways to assess treatment delivery. For example, Keefe et al. (1990) had independent raters listen to audiotapes of treatment sessions and then measured the ability of those raters to distinguish between the two treatment conditions included in the study, an index of treatment differentiation

(Kazdin, 1986a). Vallis, Shaw, and Dobson (1986) employed a 7-point Likert rating for 11 items specifically designed to judge cognitive therapists. Lewinsohn et al. (1990) specifically outlined therapist behaviors that were intended to occur during treatment sessions and then used videotapes to assess the compliance of therapists with the intended treatment delivery protocol. Using a three-point scale (2 = perfect adherence; 1 = partial adherence; 0 = no adherence), independent raters assessed the degree of therapist adherence to each of the outlined behaviors of the delivery protocol. Kazdin (1980, p. 293) had a clever suggestion, though we have not encountered its use since. He employed unidentified, confederate subjects who participated in the experiment solely for the purpose of reporting back to the experimenter on the performance of the therapist.

Indirect methods of delivery assessment rely on therapist or subject self-report. Although indirect assessment through therapist self-report may be preferable to no assessment, therapists may intentionally or unintentionally misrepresent the degree of protocol adherence. Unintentional inaccuracies in memory may be compounded by pressure to satisfy the expectations of a supervisor, thus providing an additional source of error. Using subject self-reports instead of therapist reports of treatment session content avoids assessment error due to intentional exaggeration of compliance. However, subject reports, like therapist reports, are prone to assessment error due to recall difficulty.

Figure 1 presents an example of a delivery assessment form adapted from our own research that expands upon the approach employed by Lewinsohn et al. (1990). In this treatment session, a multi-component method of relaxation is used to treat insomnia. Key treatment ingredients are listed under *positive behaviors* and are rated 0, .5, or 1 for adherence along the lines of Lewinsohn et al., but two additional features are added. First, plausible alternative treatments that may or may not be present in competing conditions within the same study are listed under *negative behaviors*. Second, ingredients are weighted for their estimated therapeutic potency, and weights of the positive behaviors will always sum to 100% and the negative behaviors to 50%.

This approach to treatment delivery assessment offers several advantages. It alerts therapists to matters of both integrity and differentiation, and the total assessment score is sensitive to the dimensions of each. If we can partition treatment into component pieces, some parts may be more important than others and their omission more costly. Therefore, weighting the parts allows more precise quantification of treatment delivery. If the entirety of treatment is presented as intended and no extra treatments are inserted, the total delivery score will be 100%. However, if intended treatments are partially presented or extra treatments are added, the total score will shrink from 100% proportional to the degree of departure from

Procedure	Delivered			Weighted %	Score
	None	Part	Full		
	<i>Positive Behaviors</i>				
Review past week	0	.5	1	15	
Relaxation:					
relaxed attitude	0	.5	1	20	
breath meditation	0	.5	1	15	
passive relaxation	0	.5	1	15	
autogenic phrases	0	.5	1	15	
Emphasize home practice	0	.5	1	20	
	Sum Positive Behaviors				
	<i>Negative Behaviors</i>				
Stimulus Control Treatment	0	.5	1	20	
Social and Recreational Activity	0	.5	1	15	
Depression Treatment	0	.5	1	15	
	Sum Negative Behaviors				
	<i>Total Assessment Score</i>				

FIG. 1 Example of form employed for delivery assessment. The degree of presence of weighted negative behaviors, i.e., treatments not intended, is subtracted from the degree of presence of weighted intended behaviors to quantify the accuracy of treatment delivery.

protocol. We decided to award a sum percentage of 50% to the negative behaviors for the following rationale. If the entirety of the intended treatment was delivered, we reasoned that the total delivery score should be high and there would be relatively little time remaining for inserting extra treatments. In the unusual event that an efficient therapist was able to add extra treatments while still completing delivery of the required treatment or took advantage of an extended session to add treatments, we did not want the total delivery score to approach zero if the entirety of the intended treatment had in fact been presented. By the present method, if the entirety of treatment is delivered and delivery is saturated by invading treatments as well, then the lowest total score possible is 50%. However, when omissions of the intended treatment are mixed with unwelcome treatment insertions, the total delivery score can range well below 100% all the way to -50%.

Treatment Implementation Model

17

Receipt Assessment

Presentation of a treatment, irrespective of whether it is the treatment that was intended, does not assure it will be accurately perceived and comprehended by the client. Treatment receipt speaks more to the mechanism or process of clinical influence than to clinical outcome, and independent measures must be employed in assessing the two.

Based upon our survey of JCCP and BT above, clinical scientists have neglected receipt assessment, but it has not escaped the attention of social psychologists. The method of manipulation checks parallels receipt assessment in clinical outcome studies, and the former has achieved widespread use. A manipulation check determines whether the subject was influenced by the manipulation in the manner which the experimenter wished (Aronson & Carlsmith, 1968). We surveyed the most recent year (1990) of the *Journal of Personality and Social Psychology*, and found that about 2/3 (42 of 63) of the studies utilizing manipulations included a manipulation check.

Manipulation checks are normally performed following the delivery of a manipulation but prior to the assessment of the dependent variable. One caution to be noted when conducting a manipulation check (or receipt assessment) is that the insertion of this assessment procedure could alter the dependent variable (Kidd, 1976). Kazdin (1980, pp. 209-210) recommends manipulation checks, but is also wary of possible artifacts they may introduce. If the experimenter suspects such problems, Kazdin recommends establishing adequate receipt in pilot work or administering the receipt assessment after final dependent measure evaluation is completed. Though these alternatives avoid tainting the dependent measure, they introduce their own shortcomings in interpretation. In our opinion, conventional manipulation checks should be omitted only when there is high expectation of dependent measure contamination. The risks created by brief, nondisruptive manipulation checks are heavily outweighed by the benefits of ascertaining that the independent variable successfully fulfilled its mission.

In clinical research, as in social psychology, receipt assessment is crucial to establishing that a fair test of the independent variable is being conducted, for preserving construct validity, and for elucidating causal mechanisms (Cook & Campbell, 1979, p. 60; Smith & Sechrest, 1991). The most desirable form of receipt assessment is direct observation of the changes effected in the patient by the delivery of the treatment. For example, Carey, Carey, and Meisler (1990) had subjects recall the five problem-solving steps they had learned in treatment. Kazdin and Mascitelli (1982) required subjects to describe their imagery in order to confirm that it corresponded to the treatment protocol. Lichstein and

Eakin (1985) conducted within-session physiological monitoring (heart rate and skin temperature) to determine if relaxation inductions produced relaxation effects. Finally, Kivlahan et al. (1990) used questionnaires to ascertain whether didactic treatments produced increments in alcohol-related knowledge.

Indirect evaluations of receipt of treatment may be used, but their findings yield less clarity. For example, in a problem-solving treatment for suicidal young adults, Lerner and Clum (1990) asked clients to complete a Problem-Solving Inventory assessing clients' perceived self-efficacy or confidence in solving interpersonal problems. Were this the sole receipt assessment tool, we could not judge how closely perceived efficacy matched the intended manipulation, actual problem-solving skill level.

If treatment does not lend itself to direct receipt assessment, indirect methods are preferable to none. Agreement between a set of indirect measures can bolster confidence that treatment was adequately received. When receipt assessment findings are positive, this adds one more piece to the methodological puzzle. Confirmation that treatment was adequately received does not by itself certify that treatment caused outcome. Other methodological controls safeguarding internal validity must also be in place. However, when receipt assessment is omitted entirely, the link between treatment and outcome can only be guessed at.

Enactment Assessment

Enactment of the treatment outside the therapy room confines, often referred to as homework, is perhaps the most difficult component of treatment implementation to assess. Depending on the nature of the treatment, subject-generated information on enactment may be all that is feasibly attainable, and under these circumstances we expect that subject report is better than none at all. Based upon our survey (see Tables 1 and 2), the incidence of indirect assessment dwarfs that of direct assessment. Most often, clients provide a written or verbal accounting of their home behavior (e.g., Appelbaum et al., 1990; Neimeyer & Feixas, 1990).

The rarity of direct enactment assessment is particularly troublesome in light of the questionable validity of indirect enactment assessment suggested by the small amount of available data comparing direct and indirect methods. A series of studies conducted by the first author and his colleagues illustrates this point. We contrasted self-reported levels of home relaxation practice with unobtrusive, objective data on the same behavior derived from a cumulative stopwatch hidden within the tape player delivering the relaxation instructions. In relaxation therapy for community volunteers diagnosed with generalized anxiety disorder or

hypertension, subjects uniformly overestimated practice time by their self-reports (Lichstein & Hoelscher, 1986). On the average, self-reports doubled actual relaxation practice. Among anxious subjects, when relying on self-reports of practice, no significant relationship was found between the amount of home practice and clinical outcome, but objective practice data did significantly correlate with outcome (Hoelscher et al., 1984). Both self-reported and objective practice levels significantly predicted systolic blood pressure changes (Hoelscher, Lichstein, & Rosenthal, 1986).

Examples of direct assessment or higher quality indirect assessments than simple self-report to monitor out of session performance can be found. Some experimenters have employed detailed interviews with subjects to corroborate reports of home behavior (Kazdin & Mascitelli, 1982), some have retrieved an independent accounting of subject enactment from household members (Lichstein & Eakin, 1985), and others have relied on biological assay of bodily fluids or expired air as a compliance index (Hill, 1988).

Requesting subjects to monitor out of session performance is a valuable means of emphasizing the importance of such behavior and instigating higher compliance, even if the resulting data are imperfect. Further, though absolute values between indirect and direct enactment data may differ, change over time between them may be correlated, and trends in indirect data across experimental phases may accurately reflect relative change.

Direct assessment of treatment enactment outside the therapy session is, of course, preferred and satisfies our curiosity about its level of occurrence. If direct assessment is impractical, indirect assessment will have to suffice. If the experimenter neglects to conduct even indirect assessment of enactment, then there should be documentation that subjects were at least urged to implement therapeutic processes on their own. In the absence of this minimal indicator, we have a weakened basis on which to assume the independent variable was given a fair test.

INDUCTION OF TREATMENT IMPLEMENTATION

Having conducted an assessment of the components of the treatment implementation model, model adherence will be fixed at a point on a continuum ranging from confirmed to disconfirmed for each component. Unless a pilot project is conducted or the experiment of interest is conducted in stages and each stage is assessed prior to commencing the next, the results of the treatment implementation assessment are available at the same time that the study is completed and the results of the study itself are first seen. Should the assessment find implementation faults, then the validity of the study's findings are degraded proportionately.

Remediation of methodological deficits in completed studies is limited to statistical procedures that isolate deficit effects. For example, level of homework practice may be correlated with treatment outcome to gauge the importance of this factor, or subjects may be classified as high, medium, or low on homework practice and then contrasted on outcome. However, solutions of this sort are inadequate compared to uniform high treatment implementation, because (a) they divert attention away from the main purpose of the experiment, (b) in the absence of random assignment of subjects to high vs. low implementation levels, subject self-selection biases heavily muddy efforts to extrapolate high implementation effects, and (c) they at best permit us to guess what the outcome might have been had high treatment implementation prevailed, but confident conclusions must still be delayed until the clinical trial is properly executed. When implementation faults surface in more than one component, statistical solutions to salvage trials grow increasingly inaccurate.

Treatment induction strategies are aimed at ensuring that high implementation of each component occurs, and the best role for such strategies is prophylactic. To avert predictable implementation lapses, experimenters should consider routine employment of methods that induce treatment implementation model components.

Delivery Induction

The treatment manual has become the standard for ensuring accurate treatment delivery, and we consider this a formal induction method. Treatment manuals minimally consist of an outline of the procedures to be followed in order to properly deliver a treatment. In addition, manuals may include background information on the treatment or the targeted disorder and often identify taboo subjects, such as competing treatments, as well. Manuals will of course vary with respect to detail and operational specificity, but descriptions of such variables are rarely reported. Behavioral therapies readily lend themselves to manualization, but treatment manuals have also been prepared for traditional therapies (Luborsky & DeRubeis, 1984).

Given their frequent occurrence and endorsement by reviewers in the area (Kazdin, 1986b; Luborsky & DeRubeis, 1984; Moncher & Prinz, 1991), it is noteworthy that the value of treatment manuals has never been empirically demonstrated. Indeed, Dobson and Shaw (1988) observed that virtually every cognitive therapy treatment publication is manual based, but data supporting their use has not been reported. Specifically, data do not exist to answer the question, when therapists are guided by treatment manuals, are the treatments delivered more accurately than when therapists are trained by other means? Confidence

in treatment manuals and other methods of inducing delivery, receipt, and enactment are often based more on faith than evidence.

Therapist training can take other avenues. Informal methods such as supervising pilot cases (Moncher & Prinz, 1991) and role-playing treatment techniques (Lewinsohn et al., 1990; Roberts & Powers, 1990) are sometimes used. In the absence of specific delivery induction procedures, assessment of delivery through viewing taped sessions or verbal supervision may itself promote treatment integrity. Adherence to treatment protocol may increase if delivery agents know they are being monitored (Peterson et al., 1982). Irrespective of the method of pre-treatment training, "booster sessions" can be conducted in the midst of treatment to maintain therapist performance (McMahon, 1987; Weissman et al., 1982).

For those who place a very high premium on treatment integrity, elimination of human change agents promises supreme confidence in knowing exactly what treatment was delivered. Though automated treatments guarantee integrity, they may sacrifice clinical potency. Taking the progressive relaxation literature as an example, contrasts of live vs. taped presentations usually find superior physiological, experiential, and clinical effects accruing to the former (Lichstein, 1988).

Receipt Induction

Communication between therapist and patient will deposit therapy knowledge with the latter, and receipt induction strategies aim to maximize communication effectiveness. Although empirical evidence guiding our choice of induction method is scarce, the clinical literature does offer examples of inventive techniques. Davis, Olmsted, and Rockert (1990) used slides during the presentation of treatment to facilitate patient comprehension. Turner et al. (1990) used role playing with feedback to ensure clients understood how to avoid reinforcing others' pain behaviors.

Informal receipt induction, i.e., simple verbal cues, instructions, reminders, etc., can supplement or function independently of the formal methods exemplified above, and informal methods by themselves can be helpful. For example, repetition of the message in treatment can facilitate receipt and retention of key ideas, as recommended by Rosenthal and Downs (1985) in their review, and this assertion is bolstered by basic research (e.g., Rose, 1984). Similarly, Kazdin (1979) demonstrated that using summaries of the important points of treatment increased memory of treatment content. Lastly, questioning the patient's understanding of material already presented will sometimes cause patients to thoughtfully analyze and reorganize new information. According to Rosenthal and

Downs (1985), this process, which they term actively reworking, is particularly effective in increasing retention.

The design of the treatment itself and its presentation style may enhance treatment comprehension. Personally relevant treatments or methods of presentation are more likely to be remembered and to stimulate further thought by the patient (Sorrentino, Bobocel, Gitta, Olson, & Hewitt, 1988). Relevance may be advanced through metaphorical explanations. Not only have metaphors been shown to improve memory, they also promote problem-solving abilities (Evans, 1988).

Receipt induction techniques should be tailored to the special characteristics of their audience. For example, metaphors may not be understood by children. Role-play methods may be ineffective with shy people. Complex treatments may be confusing to all but the brightest patients. Temporary characteristics, such as the patient's state of mind, merit similar consideration. The issue of tailoring highlights the wisdom of a fluid interplay between induction and assessment. If one induction strategy does not achieve adequate receipt, perhaps another one would.

Poor comprehension of potent treatments will yield treatment gains comparable to well-comprehended weaker treatments. When receipt of treatment suffers, treatment efficacy is confounded with communication deficits, and the contaminated clinical trial outcome cannot be deemed a fair test of the identified treatment.

Enactment Induction

Securing client cooperation with out of session assignments can be a difficult task, but a very important one. To illustrate the potential severity of this problem, when highly respected health authorities, i.e., physicians, advise their patients to engage in a nondemanding home task bearing strong health consequences, i.e., taking medication, long term compliance averages about 50% (Sackett & Snow, 1979). Therapists prescribing more complex and time consuming home activities may encounter even less success. Under the rubrics of adherence to treatment regimen or homework compliance, much attention has been paid to this dimension, and excellent reviews are available (Epstein & Cluss, 1982; Haynes, Taylor, & Sackett, 1979).

Numerous instances of inventive strategies to induce enactment can be found in the literature, though their utility is infrequently confirmed experimentally. We have referred to the use of concrete props as formal induction, and this is exemplified by reminder/motivational cards given to the patient (Azrin & Peterson, 1990), forms for recording home practice (Rosen et al., 1990), and distributing relaxation tapes for home practice (Blanchard et al., 1990a). Informal induction, the use of exhortation to

persuade patients of the importance of home behavior and to motivate their cooperation, has also been adopted and certainly boasts of convenience (Kivlahan et al., 1990; Neimeyer & Feixas, 1990).

Mild to moderate success in elevating compliance has been reported for using reminder cues, individually tailored homework assignments, contracting, self-monitoring, reinforcement, and family/peer support (Dunbar, Marshall, & Hovell, 1979). Epstein and Cluss (1982) concluded that strategies that employed reinforcement for compliance or feedback on level of compliance achieved the most success.

Without supportive data, caution should be exercised in assuming that enactment induction methods which convey confidence in their efficacy are indeed so. To illustrate this point, a study cited in the paragraph above employed relaxation tapes to aid home practice, and this is a very common strategy. Three studies have contrasted home relaxation practice with and without tapes and found no advantage for taped practice among insomniacs (Lick & Heffler, 1977), anxious individuals (Mayer, Frederiksen, & Scanlon, 1984), and hypertensives (Hoelscher, Lichstein, Fischer, & Hegarty, 1987).

In many treatments, the home is the primary arena of therapeutic exposure and the value of efforts to obtain high home compliance cannot be overstated. Opinions about this may be so strong, that some authors have recommended the next therapy session itself not be scheduled until the patient completes home assignments (Shelton & Ackerman, 1974). Although we cannot attest to the power of exhortation to induce enactment, the absence of even this minimal effort to evoke patients home participation appears to us as a flagrant methodological oversight. Judging from our survey of BT and JCCP, many would not agree with this conclusion. Less than half of the clinical outcome studies in both journals reported formal or informal enactment induction.

DISCUSSION

Efficacy judgments rendered on partially implemented treatments will be unjustly construed as fair conclusions when implementation assessment is absent. If assessment is omitted, the question "Was a fair clinical trial conducted?" is indeterminate. When assessment is conducted and should partial implementation be confirmed, the validity of the clinical trial is undermined. Failing to take specific steps to secure satisfactory implementation heightens the risk of partial implementation. In light of the great cost in time and money demanded by most clinical trials, proper treatment implementation induction and assessment are prudent investments to minimize erosion of validity.

In our survey of *Behavior Therapy* and the *Journal of Consulting and*

Clinical Psychology, we found strong interest in ensuring the adequacy of the delivery component, but scattered interest in the other two, and this data is generally consistent with the few other surveys we have cited. Further, it may be fair to conclude that methodological rigor in the large number of clinical journals suffers in comparison to BT and JCCP, and the untested great majority of clinical studies have done even less that is relevant to our treatment implementation model. If we believe that adequate levels of delivery, receipt, and enactment are required for valid clinical trials and that shortcomings in any one of these will potentially sabotage fair tests, can we dispel suspicion cast on the enormity of the extant clinical outcome literature? Dare we have the courage to even admit that such a question is legitimate?

If psychology is a young science, then clinical psychology is a toddler still struggling to articulate its methodological foundation. The recommendations advanced herein should be considered to be but one more step towards the distant conclusion of this long process.

Acknowledgements — Preparation of this article was supported, in part, by a grant from the H. W. Durham Foundation and by the Department of Psychology's Center for Applied Psychological Research, part of the State of Tennessee's Center of Excellence Grant program. We wish to thank Robert Klesges and William Shadish for their helpful comments on an earlier draft.

REFERENCES

- Abramowitz, A. J., & O'Leary, S. G. (1990). Effectiveness of delayed punishment in an applied setting. *Behavior Therapy*, 21, 231-239.
- Agras, W. S., Taylor, C. B., Feldman, D. E., Losch, M., & Burnett, K. F. (1990). Developing computer-assisted therapy for the treatment of obesity. *Behavior Therapy*, 21, 99-109.
- Appelbaum, K. A., Blanchard, E. B., Nicholson, N. L., Radnitz, C., Kirsch, C., Michultka, D., Attanasio, W., Andrasik, F., & Dentinger, M. P. (1990). Controlled evaluation of the addition of cognitive strategies to a home-based relaxation protocol for tension headache. *Behavior Therapy*, 21, 293-303.
- Aronson, E., & Carlsmith, J. M. (1968). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology*; Vol. 2 (2nd ed., pp. 1-79). Reading, MA: Addison-Wesley.
- Azrin, N. H., & Peterson, A. L. (1990). Treatment of Tourette Syndrome by habit reversal: A waiting-list control group comparison. *Behavior Therapy*, 21, 305-318.
- Baggs, K., & Spence, S. H. (1990). Effectiveness of booster sessions in the maintenance and enhancement of treatment gains following assertion training. *Journal of Consulting and Clinical Psychology*, 58, 845-854.
- Baucom, D. H., Sayers, S. L., & Sher, T. G. (1990). Supplementing behavioral marital therapy with cognitive restructuring and emotional expressiveness training: An outcome investigation. *Journal of Consulting and Clinical Psychology*, 58, 636-645.
- Behrens, B. C., Sanders, M. R., & Halford, W. K. (1990). Behavioral marital therapy: An evaluation of treatment effects across high and low risk settings. *Behavior Therapy*, 21, 423-433.
- Bellack, A. S., & Hersen, M. (Eds.) (1984). *Research methods in clinical psychology*. New York: Pergamon.

Treatment Implementation Model

25

- Billingsley, F., White, O. R., & Munson, R. (1980). Procedural reliability: A rationale and an example. *Behavioral Assessment, 2*, 229-241.
- Blanchard, E. B., Appelbaum, K. A., Radnitz, C. L., Michultka, D., Morrill, B., Kirsch, C., Hillhouse, J., Evans, D. D., Guarnieri, P., Attanasio, V., Andrasik, F., Jaccard, J., & Dentinger, M. P. (1990a). Placebo-controlled evaluation of abbreviated progressive muscle relaxation therapy and of relaxation therapy combined with cognitive therapy in the treatment of tension headache. *Journal of Consulting and Clinical Psychology, 58*, 210-215.
- Blanchard, E. B., Appelbaum, K. A., Radnitz, C. L., Morrill, B., Michultka, D., Kirsch, C., Guarnieri, P., Hillhouse, J., Evans, D. D., Jaccard, J., & Barron, K. D. (1990b). A controlled evaluation of thermal biofeedback and thermal biofeedback combined with cognitive therapy in the treatment of vascular headache. *Journal of Consulting and Clinical Psychology, 58*, 216-224.
- Botvin, G. J., Baker, E., Dusenbury, L., Tortu, S., & Botvin, E. M. (1990). Preventing adolescent drug abuse through a multimodal cognitive-behavioral approach: Results of a 3-year study. *Journal of Consulting and Clinical Psychology, 58*, 437-446.
- Boudewyns, P. A., & Hyer, L. (1990). Physiological response to combat memories and preliminary treatment outcome in Vietnam veteran PTSD patients treated with direct therapeutic exposure. *Behavior Therapy, 21*, 63-87.
- Carey, M. P., Carey, K. B., & Meisler, A. W. (1990). Training mentally ill chemical abusers in social problem solving. *Behavior Therapy, 21*, 511-518.
- Chambless, D. L. (1990). Spacing of exposure sessions in treatment of agoraphobia and simple phobia. *Behavior Therapy, 21*, 217-229.
- Christensen, L., & Burrows, R. (1990). Dietary treatment of depression. *Behavior Therapy, 21*, 183-193.
- Church, P., Forehand, R., Brown, C., & Holmes, T. (1990). Prevention of drug abuse: Examination of the effectiveness of a program with elementary school children. *Behavior Therapy, 21*, 339-347.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago: Rand McNally.
- Danforth, J. S., Allen, K. D., Fitterling, J. M., Danforth, J. A., Farrar, D., Brown, M., & Drabman, R. S. (1990). Exercise as a treatment for hypertension in low-sociometric-status black children. *Journal of Consulting and Clinical Psychology, 58*, 237-239.
- Davis, R., Olmsted, M. F., & Rockert, W. (1990). Brief group psychoeducation for bulimia nervosa: Assessing the clinical significance of change. *Journal of Consulting and Clinical Psychology, 58*, 882-885.
- Dobson, K. S., & Shaw, B. F. (1988). The use of treatment manuals in cognitive therapy: Experience and issues. *Journal of Consulting and Clinical Psychology, 56*, 673-680.
- Dunbar, J. M., Marshall, G. D., & Hovell, M. F. (1979). Behavioral strategies for improving compliance. In R. B. Haynes, D. W. Taylor, & D. L. Sackett (Eds.), *Compliance in health care* (pp. 174-190). Baltimore: Johns Hopkins University Press.
- Eikin, I., Parloff, M. B., Hadley, S. W., & Autry, J. H. (1985). NIMH treatment of Depression Collaborative Research Program: Background and research plan. *Archives of General Psychiatry, 42*, 305-316.
- Epstein, L. H., & Cluss, P. A. (1982). A behavioral medicine perspective on adherence to longterm medical regimens. *Journal of Consulting and Clinical Psychology, 50*, 950-971.
- Evans, G. E. (1988). Metaphors as learning aids. *Journal of Experimental Education, 56*, 91-99.
- Feinstein, A. R. (1979). "Compliance bias" and the interpretation of therapeutic trials. In R. B. Haynes, D. W. Taylor, & D. L. Sackett (Eds.), *Compliance in health care* (pp. 309-322). Baltimore: Johns Hopkins University Press.
- Fonagy, P., & Moran, G. S. (1990). Studies on the efficacy of child psychoanalysis. *Journal of Consulting and Clinical Psychology, 58*, 684-695.

- Garfield, S. L., & Bergin, A. E. (Eds.) (1986). *Handbook of psychotherapy and behavior change*. New York: Wiley.
- Goldsmith, C. H. (1979). The effect of compliance distributions on therapeutic trials. In R. B. Haynes, D. W. Taylor, & D. L. Sackett (Eds.), *Compliance in health care* (pp. 297-308). Baltimore: Johns Hopkins University Press.
- Harmon, T. M., Hynan, M. T., & Tyre, T. E. (1990). Improved obstetric outcomes using hypnotic analgesia and skill mastery combined with childbirth education. *Journal of Consulting and Clinical Psychology, 58*, 525-530.
- Haynes, R. B., Taylor, D. W., & Sackett, D. L. (Eds.) (1979). *Compliance in health care*. Baltimore: Johns Hopkins University Press.
- Hill, R. D. (1988). Prescribing aversive relapse to enhance nonsmoking treatment gains: A pilot study. *Behavior Therapy, 19*, 35-43.
- Hoelscher, T. J., Lichstein, K. L., & Rosenthal, T. L. (1984). Objective vs subjective assessment of relaxation compliance among anxious individuals. *Behaviour Research and Therapy, 22*, 187-193.
- Hoelscher, T. J., Lichstein, K. L., & Rosenthal, T. L. (1986). Home relaxation practice in hypertension treatment: Objective assessment and compliance induction. *Journal of Consulting and Clinical Psychology, 54*, 217-221.
- Hoelscher, T. J., Lichstein, K. L., Fischer, S., & Hegarty, T. B. (1987). Relaxation treatment of hypertension: Do home relaxation tapes enhance treatment outcome? *Behavior Therapy, 18*, 33-37.
- Imber, S. D., Piskonis, P. A., Sotsky, S. M., Elkin, I., Watkins, J. T., Collins, J. F., Shea, M. T., Lober, W. R., & Glass, D. R. (1990). Mode-specific effects among three treatments for depression. *Journal of Consulting and Clinical Psychology, 58*, 352-359.
- Jay, S. M., & Elliott, C. H. (1990). A stress inoculation program for parents whose children are undergoing painful medical procedures. *Journal of Consulting and Clinical Psychology, 58*, 799-804.
- Johnson, C. A., Pentz, M. A., Weber, M. D., Dwyer, J. H., Baer, N., Mackinnon, D. P., Hansen, W. B., & Flay, B. R. (1990). Relative effectiveness of comprehensive community programming for drug abuse prevention with high-risk and low-risk adolescents. *Journal of Consulting and Clinical Psychology, 58*, 447-456.
- Kazdin, A. E. (1979). Imagery elaboration and self efficacy in the covert modeling treatment of unassertive behavior. *Journal of Consulting and Clinical Psychology, 47*, 725-733.
- Kazdin, A. E. (1980). *Research design in clinical psychology*. New York: Harper & Row.
- Kazdin, A. E. (1986a). Comparative outcome studies of psychotherapy: Methodological issues and strategies. *Journal of Consulting and Clinical Psychology, 54*, 95-105.
- Kazdin, A. E. (1986b). The evaluation of psychotherapy: Research design and methodology. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (pp. 23-68). New York: Wiley.
- Kazdin, A. E., & Maccitelli, S. (1982). Covert and overt rehearsal and homework practice in developing assertiveness. *Journal of Consulting and Clinical Psychology, 50*, 250-258.
- Keefe, F. J., Caldwell, D. S., Williams, D. A., Gil, K. M., Mitchell, D., Robertson, C., Martinez, S., Nunley, J., Beckham, J. C., Crisson, J. E., & Helms, M. (1990). Pain coping skills training in the management of osteoarthritic knee pain: A comparative study. *Behavior Therapy, 21*, 49-62.
- Kidd, R. F. (1976). Manipulation checks: Advantage or disadvantage? *Representative Research in Social Psychology, 7*, 160-165.
- Killed, J. D., Fortmann, S. P., Newman, B., & Varady, A. (1990). Evaluation of a treatment approach combining nicotine gum with self-guided behavioral treatments for smoking relapse prevention. *Journal of Consulting and Clinical Psychology, 58*, 85-92.
- Kivlahan, D. R., Marlatt, G. A., Fromme, K., Coppel, D. B., & Williams, E. (1990).

Treatment Implementation Model

27

- Secondary prevention with college drinkers: Evaluation of an alcohol skills training program. *Journal of Consulting and Clinical Psychology*, *58*, 805-810.
- Klosko, J. S., Bartow, D. H., Tassinari, R., & Cerny, J. A. (1990). A comparison of alprazolam and behavior therapy in treatment of panic disorder. *Journal of Consulting and Clinical Psychology*, *58*, 77-84.
- Lerner, M. S., & Clum, G. A. (1990). Treatment of suicide ideators: A problem-solving approach. *Behavior Therapy*, *21*, 403-411.
- Lewinsohn, P. M., Clarke, G. N., Hops, H., & Andrews, J. (1990). Cognitive-behavioral treatment for depressed adolescents. *Behavior Therapy*, *21*, 385-401.
- Lichstein, K. L. (1988). *Clinical relaxation strategies*. New York: Wiley.
- Lichstein, K. L., & Eakin, T. L. (1985). Progressive versus self-control relaxation to reduce spontaneous bleeding in hemophiliacs. *Journal of Behavioral Medicine*, *8*, 149-162.
- Lichstein, K. L., & Hoelscher, T. J. (1986). A device for unobtrusive surveillance of home relaxation practice. *Behavior Modification*, *10*, 219-233.
- Lick, J. R., & Heffler, D. (1977). Relaxation training and attention placebo in the treatment of severe insomnia. *Journal of Consulting and Clinical Psychology*, *45*, 153-161.
- Lowe, C. F., & Chadwick, P. D. J. (1990). Verbal control of delusions. *Behavior Therapy*, *21*, 461-479.
- Luborsky, L., & DeRubeis, R. J. (1984). The use of psychotherapy treatment manuals: A small revolution in psychotherapy research style. *Clinical Psychology Review*, *4*, 5-14.
- Luiselli, J. K., Marholin, D., III, Steinman, D. L., & Steinman, W. M. (1979). Assessing the effects of relaxation training. *Behavior Therapy*, *10*, 663-668.
- Malgady, R. G., Rogler, L. H., & Constantino, G. (1990a). Culturally sensitive psychotherapy for Puerto Rican children and adolescents: A program of treatment outcome research. *Journal of Consulting and Clinical Psychology*, *58*, 704-712.
- Malgady, R. G., Rogler, L. H., & Constantino, G. (1990b). Hero/heroine modeling for Puerto Rican adolescents: A preventive mental health intervention. *Journal of Consulting and Clinical Psychology*, *58*, 469-474.
- Manne, S. L., Redd, W. H., Jacobsen, P. B., Gorfinkle, K., Schorr, O., & Rapkin, B. (1990). Behavioral intervention to reduce child and parent distress during venipuncture. *Journal of Consulting and Clinical Psychology*, *58*, 565-572.
- Mayer, J. A., Frederiksen, L. W., & Scanlon, G. J. (1984, March). *Relaxation training: The effect of take-home audio tapes on frequency of practice*. Paper presented at the meeting of the Southeastern Psychological Association, New Orleans.
- McMahon, P. M. (1987). Shifts in intervention procedures: A problem in evaluating human service interventions. *Social Work Research and Abstracts*, *23*, 13-16.
- Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, *11*, 247-266.
- Murphy, A. I., Lehrer, P. M., & Jurish, S. (1990). Cognitive coping skills training and relaxation training as treatments for tension headaches. *Behavior Therapy*, *21*, 89-98.
- Neimeyer, R. A., & Feixas, G. (1990). The role of homework and skill acquisition in the outcome of group cognitive therapy for depression. *Behavior Therapy*, *21*, 281-292.
- Persons, J. B., Burns, D. D., & Perloff, J. M. (1988). Predictors of dropout and outcome in cognitive therapy for depression in a private practice setting. *Cognitive Therapy and Research*, *12*, 557-575.
- Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis*, *15*, 477-492.
- Primakoff, L., Epstein, N., & Covi, L. (1986). Homework compliance: An uncontrolled variable in cognitive therapy outcome research. *Behavior Therapy*, *17*, 433-446.
- Quay, H. C. (1977). The three faces of evaluation: What can be expected to work. *Criminal Justice and Behavior*, *4*, 341-354.
- Renneberg, B., Goldstein, A. J., Phillips, D., & Chambless, D. L. (1990). Intensive

- behavioral group treatment of avoidant personality disorder. *Behavior Therapy*, 21, 363-377.
- Roberts, M. W., & Powers, S. W. (1990). Adjusting chair timeout enforcement procedures for oppositional children. *Behavior Therapy*, 21, 257-271.
- Rose, R. J. (1984). Processing time for repetitions and the spacing effect. *Canadian Journal of Psychology*, 38, 537-550.
- Rosen, J. C., Cado, S., Silberg, N. T., Srebnik, D., & Wendt, S. (1990). Cognitive behavior therapy with and without size perception training for women with body image disturbance. *Behavior Therapy*, 21, 481-498.
- Rosenthal, T. L., & Downs, A. (1985). Cognitive aids in teaching and treating. *Advances in Behaviour Research and Therapy*, 7, 1-53.
- Sackett, D. L., & Snow, J. C. (1979). The magnitude of compliance and noncompliance. In R. B. Haynes, D. W. Taylor, & D. L. Sackett (Eds.), *Compliance in health care* (pp. 11-22). Baltimore: Johns Hopkins University Press.
- Schinke, S. P., Gordon, A. N., & Weston, R. E. (1990). Self-instruction to prevent HIV infection among African-American and Hispanic-American adolescents. *Journal of Consulting and Clinical Psychology*, 58, 432-436.
- Sechrest, L., West, S. G., Phillips, M. A., Redner, R., & Yeaton, W. (1979). Some neglected problems in evaluation research: Strength and integrity of treatments. In L. Sechrest, S. G. West, M. A. Phillips, R. Redner, & W. Yeaton (Eds.), *Evaluation studies review annual* (Vol. 4, pp. 15-35). Beverly Hills, CA: Sage.
- Sechrest, L., & Yeaton, W. E. (1981). Assessing the effectiveness of social programs: Methodological and conceptual issues. In S. Bail (Ed.), *New directions for program evaluation: Assessing and interpreting outcomes* (pp. 41-56). San Francisco: Jossey-Bass.
- Shelton, J. L., & Ackerman, J. M. (1974). *Homework in counseling and psychotherapy*. Springfield, IL: Thomas.
- Smith, B., & Sechrest, L. (1991). Treatment of aptitude x treatment interactions. *Journal of Consulting and Clinical Psychology*, 59, 233-244.
- Sorrentino, R. M., Bobocel, D. R., Gitta, M. Z., Olson, J. M., & Hewitt, E. C. (1988). Uncertainty orientation and persuasion: Individual differences in the effects of personal relevance on social judgements. *Journal of Personality and Social Psychology*, 55, 357-371.
- Starfield, B. (1977). Efficacy and effectiveness of primary medical care for children. In *Report of the Harvard Child Health Project: Vol. 2. Children's medical care needs and treatment* (pp. 71-76). Cambridge, MA: Ballinger.
- Sullivan, M. A., & O'Leary, S. G. (1990). Maintenance following reward and cost token programs. *Behavior Therapy*, 21, 139-149.
- Telch, C. F., Agras, W. S., Rossiter, E. M., Wilfley, D., & Kenardy, K. (1990). Group cognitive-behavioral treatment for the nonpurging bulimic: An initial evaluation. *Journal of Consulting and Clinical Psychology*, 58, 629-635.
- Turner, J. A., Clancy, S., McQuade, K. J., & Cardenas, D. D. (1990). Effectiveness of behavioral therapy for chronic low back pain: A component analysis. *Journal of Consulting and Clinical Psychology*, 58, 573-579.
- Vallis, T. M., Shaw, B. F., & Dobson, K. S. (1986). The cognitive therapy scale: Psychometric properties. *Journal of Consulting and Clinical Psychology*, 54, 381-385.
- Vermilyea, B. B., Barlow, D. H., & O'Brien, G. T. (1984). The importance of assessing treatment integrity: An example in the anxiety disorders. *Journal of Behavioral Assessment*, 6, 1-11.
- Vlegener, B. J., Perri, M. G., Nezu, A. M., Rcnjilian, D. A., McKelvey, W. F., & Schein, R. L. (1990). Effects of an intermittent, low-fat, low-calorie diet in the behavioral treatment of obesity. *Behavior Therapy*, 21, 499-509.
- Webster-Stratton, C., & Hammond, M. (1990). Predictors of treatment outcome in parent training for families with conduct problem children. *Behavior Therapy*, 21, 319-337.
- Weissman, M. M., Rounsaville, B. J., & Chevron, E. (1982). Training psychotherapists

Treatment Implementation Model

29

- to participate in psychotherapy outcome studies. *American Journal of Psychiatry*, 139, 1442-1446.
- Welsz, J. R., Walter, B. R., Weiss, B., Fernandez, G. A., & Mikow, V. A. (1990). Arrests among emotionally disturbed violent and assaultive individuals following minimal versus lengthy intervention through North Carolina's Willie M program. *Journal of Consulting and Clinical Psychology*, 58, 720-728.
- Wells, A. (1990). Panic disorder in association with relaxation induced anxiety: An attentional training approach to treatment. *Behavior Therapy*, 21, 273-280.
- Williams, K. E., & Chambless, D. (1990). The relationship between therapist characteristics and outcome of in vivo exposure treatment for agoraphobia. *Behavior Therapy*, 21, 111-116.
- Wurtele, S. K. (1990). Teaching personal safety skills to four-year-old children: A behavioral approach. *Behavior Therapy*, 21, 25-32.
- Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology*, 49, 156-167.