Treatment Strength and Integrity: Models and Methods

David S. Cordray and Georgine M. Pion

In 1979, Sechrest, West, Phillips, Redner, and Yeaton edited Volume 4 of the Evaluation Studies Review Annual. During the Annual's relatively short history, its various editors opted to either reprint a collection of exemplary evaluation studies or organize their respective volume around important methodological and technological themes underlying the conceptualization and execution of evaluation studies. Lee Sechrest and his fellow editors chose the latter course of editorial action, highlighting the importance of incorporating treatment strength and integrity into evaluation endeavors as their major theme. Labeling it as one of the "neglected problems in evaluation research," they argued that evaluation researchers had typically shortchanged the role of the treatment in their studies. In brief, treatment strength referred to the a priori intensity of the planned intervention, and treatment integrity referred to the fidelity with which the treatment was actually delivered. Building on earlier work (Sechrest & Redner, 1979), this theme was reiterated and expanded on by Sechrest and other collaborators (Scott & Sechrest, 1989; Sechrest, 1982; Sechrest & Yeaton, 1981; Yeaton & Sechrest, 1981a, 1981b).

Taken together, the ideas in these papers make up an important part of the intellectual foundation underlying many conceptual, methodological, and technical advances in evaluation research. This chapter recounts Sechrest et al.'s (1979) original notions of treatment strength and integrity. More recent work that was inspired, directly and indirectly, by their thinking also is summarized. Overall, their call for paying more attention to adequately assessing treatment integrity did not fall on deaf ears; new implementation measures have been developed, and an increasing number of intervention efficacy and effectiveness studies now include evidence on treatment integrity. The same conclusion is not justified for treatment strength as it was conceptualized in Sechrest et al.'s (1979) chapter. In reading subsequent papers citing their works, the two constructs were not often distinguished, being discussed as a single construct—strength and integrity. Many conceptual and methodological developments in evaluation and intervention research have emerged (see Lipsey & Cordray, 2000; Shadish, Cook, & Campbell, 2002) over the past 2 decades, making it is possible to restore the separate status of their original
two constructs. Now, a stronger argument for the a priori consideration of treatment strength is possible. Before engaging in those discussions, we summarize the main ideas raised by Sechrest and his colleagues in several important papers.

Treatment Strength and Integrity *a la* Sechrest and Colleagues

Sechrest et al.'s (1979) interest in treatment strength and integrity is clearly understandable if we place their ideas in a historical context. When their edition of the *Annual* was published, program evaluation research was still in its infancy, having begun, in earnest, in the mid-1970s. By 1979, these efforts had produced a succession of "no-difference" findings as reported in the literature. These disappointing results, along with the implication that "nothing works," prompted an industrywide, critical investigation of the circumstances surrounding the conduct of evaluation efforts.

Sechrest and his colleagues made at least two important contributions to the debate about the reasons for prior treatment failures. First, they made several key observations about the nature of treatments: (a) They are delivered in real settings and are rarely standardized; (b) they often are multifaceted, composed of multiple components or stages; (c) they are sometimes delivered by poorly trained, unmotivated, or resistant program staff; and (d) their delivery can be heavily influenced by events in the real world, many of which are not under the provider's or researcher's control. Each can engender potential problems with treatment delivery. Under such "messy" circumstances, Sechrest et al. (1979) concluded that "the failure of an actual treatment to produce a significant effect may tell us nothing about the potential effect had the treatment been correctly implemented" (p. 16).

Their second observation is perhaps better classified as a nonobservation. In classic "Sechrestian" style, they asked, "What is missing within these evaluation studies?" Their answer identified the absence of any compelling rationale, model, or theory underlying the programs that were tested. To make this point crystal clear, they grounded their concern about questionable treatment–outcome linkages by using a real evaluation study as an example. They asked,

> On what *a priori theoretical basis* could [these] researchers have expected that an hour or two a week of counseling by poorly trained correctional officers would have any detectable effect on the response [subsequent criminal activity] of the offenders to parole over a period of three years? (Sechrest et al., 1979, p. 23, italics added)

In other words, how could such a seemingly weak intervention produce an effect in a behavior that is notoriously difficult to change? The emphasis on treatment strength was to avoid developing and testing interventions that had no obvious chance for success.

To formalize their concerns, Sechrest and his colleagues offered two somewhat different definitions of treatment strength and integrity. Initially, they
defined treatment strength as "the intensity with which the researcher intends that the treatment be delivered" (Sechrest et al., 1979, p. 16, italics added). Integrity of treatment was defined as "the fidelity with which the treatment is actually delivered" (p. 16, italics added). Two years later, Yeaton and Sechrest (1981a) made the connection between the theory underlying the treatment and fidelity by referring to integrity as "the degree to which treatment is delivered as intended" (p. 160, italics added).

**Treatment Strength: A Multifaceted Concern**

There are at least four benefits to an increased concern about determining treatment strength that can be deduced from the discussions provided by Sechrest and his colleagues: (a) specifying the amount or "magnitude" of the treatment to be delivered, (b) understanding the coherence of the intervention or treatment, (c) enhancing the validity of inference that can be derived from studies of treatment effectiveness, and (d) optimizing the use of the results to guide interventions or clinical practices.

**The Magnitude of Treatment**

Treatments can differ in quantifiable ways. Sechrest et al. (1979) offered several dimensions of treatments that imply the possibility of differential levels of strength. In particular, they noted that treatments can vary with regard to the number of sessions that are to be provided, the duration over which they are to be provided, and the intended intensity (units per time interval) of the scheduled sessions.

Yeaton and Sechrest (1981a) stated that "strong treatments contain large amounts in pure form of those ingredients leading to change" (p. 156). Not to be confused with a general notion that more is always better, optimal treatment strength is viewed as the preferable target. Optimality depends on the detail with which the problem is specified, the magnitude of the problem to be alleviated, the differential responsiveness of individuals (e.g., Treatment × Participant interactions), and political realities.

**The Theoretical Coherence of Treatments**

Inspired by research and practice in medicine, Sechrest et al. (1979) pointed out that optimal treatments depend on knowledge about the conditions under which treatments (notably, drugs) produce desired outcomes. That is, medical practitioners need to know which drugs "work" on which medical conditions, the mechanisms that are at work, and the appropriate dosage adjustments in the presence of other attributes (e.g., weight, age, gender, and comorbid conditions of the patient). A similar network of knowledge is required for nonmedical interventions. Sechrest et al. argued that the treatment plan needs to be well grounded in a theory that links the type of intervention that is proposed, the specific population that is involved, and the anticipated outcomes. Strong
treatments have an acceptable theoretical rationale, treatment plan, and specification of the mechanisms or processes that are expected to produce the desired outcomes. Their definition stresses the a priori consideration of treatment strength, emphasizing its role in the initial conceptualization of the intervention.

Assessing the a priori strength of a treatment can be accomplished by asking experts to rate: (a) the likelihood that the treatment, as described, will produce the desired changes in the outcome; (b) its clarity; and (c) the extent to which it is focused on the intended problem. Strength estimates also can be developed in reference to the strongest possible treatment (e.g., a substantial fraction of the available time) and the attributes of standard treatments or norms about customary care. The prior experiences of potential clients in standard treatments can serve as a basis for rating the strength of an innovative form of treatment. Sechrest et al. (1979) also hinted at the potential role of meta-analytic methods in parametric studies.

Improving Causal Inference

Another “Sechrestian” feature of the published discussions concerning treatment strength and integrity is an emphasis on enhancing the validity of causal inferences about the effects of interventions. As such, it is not surprising that the role of strength and integrity of treatment is discussed in terms of Cook and Campbell’s (1979) definition of construct validity (of the cause)—the proper understanding of the true meaning of the treatment. Again, placing these discussions in historical context, construct validity had previously received little or no attention relative to the concerns expressed over internal validity (the proper attribution of an observed effect). Applying Cook and Campbell’s threats to construct validity, inappropriate inferences about the treatment effects could be due to confounding variables, nonspecific effects (e.g., expectancy, placebo effects), inadequate theoretical formulation of the treatment, and inadequate description of the treatment and control conditions. To this list, they added “inadequacy of the planned strength of the treatment and the integrity with which it is delivered” (p. 18).

Improving Clinical Practice

Research and evaluation are undertaken for a variety of reasons. Throughout the collection of papers issued by Sechrest and his colleagues runs the assumption that research results should serve as an important basis for improving practitioners’ delivery of social and behavioral treatments. In large measure, this concern is a natural extension of their emphasis on construct validity. The careful explication of treatment rationales, specification of the conditions necessary for their delivery, and description of the mechanisms that link treat-

---

1A year is equal to 8,760 hours; any treatment that approached that dose per year could be considered a strong treatment.
TREATMENT STRENGTH AND INTEGRITY

ment components and processes to outcomes (intended and unintended) serve as important bases for the development of practice guidelines. Improving clinical practice will be seriously hampered if we do not direct explicit attention to (a) treatment parameters that are plausibly linked to treatments; (b) the relative effects produced and outcomes achieved by treatments (e.g., improvement rates and achieved levels of functioning); and (c) the conditions under which effects and outcomes are produced as well as their differential effects on subgroups of clients.

An obvious question concerns the extent to which their concerns registered with the scientific and practitioner communities and prompted efforts to strengthen the evidence base on the nature and effectiveness of interventions. Recognizing that the discussion which follows falls seriously short of a comprehensive review, we nevertheless attempt to summarize some of the progress that has been made. Where possible, we also highlight issues that have not attracted the level of attention that they warrant.

Models for Linking the Strength of Causes With Their Effects

In the series of papers just described, Sechrest and his colleagues focused mainly on explicating the critical and, in their view, neglected aspects of the purported cause (treatment) within studies of treatment effectiveness. In most of their papers, Sechrest and his colleagues left the effect undefined. Because little attention had been directed at treatment strength, discussions of methods for measuring strength were suggestive, and following the highly original framework of field research captured in the book *Unobtrusive Methods* (Webb, Campbell, Schwartz, & Sechrest, 1966), the methods were creative. The past 2 decades of research and development have witnessed a surge of interest in the role of theories in intervention research (e.g., Chen, 1990; Chen & Rossi, 1992; Lipsey, 1993; Yin, 1997) and, in turn, interest in defining treatments, mechanisms by which treatments operate, and their linkages to outcomes. With these advances came the possibility of more explicit consideration of the original ideas developed by Sechrest and his colleagues about treatment strength and integrity.

Across a broad spectrum of intervention research (e.g., clinical psychology, psychosocial rehabilitation, substance abuse, and homelessness), causal analysis and evaluation methods reveal at least three models of how causes (treatments) can be linked to their expected effects. The first model stems from the literature on human judgment processes. We refer to it as the *cause–effect congruity model* because it focuses explicitly on a priori conceptions of the correspondence between the size (or strength) of the cause and the size (or magnitude) of the resulting effect. Second, the *counterfactual model* is derived from the experimental research and evaluation traditions. Here, the effect is defined as a relative, average difference between groups. As a consequence, the cause (treatment) has to be defined as a relative difference between treatment conditions. Moreover, the methodological underpinnings of the counterfactual model can provide the soundest basis for causal attribution. The third approach to linking treatment strength to outcomes is the *dose–response model*. This
model examines the form of the relationship between differing strengths of treatment and levels of outcomes.

The Cause–Effect Congruity Model

In their paper titled “Judging Probable Cause,” Einhorn and Hogarth (1988) provided a set of rules that, they argued, people use to make judgments of causality. In addition to the usual cues-to-causality used in causal research (i.e., the appropriate temporal order or X→Y, covariation of the cause and effect, and the absence of rival explanations for the observed X and Y relationship), Einhorn and Hogarth introduced the idea of cause–effect congruence as another important cue-to-causality. As depicted in Figure 6.1, two treatments are scaled according to their treatment strength (e.g., through expert ratings). Congruence occurs when strong treatments (e.g., rated at the 75th percentile of all relevant treatments) produce large effects (mean1 = 75, where 100 is a maximum level of performance) and small causes (weak treatments, e.g., a treatment rated in the 25th percentile of all relevant treatments) produce, on average, small effects (mean1 = 25). These different scenarios are designated as Paths A and B, respectively, in Figure 6.1. In practice, routine a priori assessment of treatment strength should minimize the implementation of a weak treatment, pressuring program developers to bolster treatment models in ways that strengthen the treatment and increasing the likelihood of success.

Incongruity occurs when it is thought that a small cause can produce a big effect or the reverse occurs. Incongruence motivates a search for additional mechanisms to explain the apparent mismatch between treatment and outcome strength (Cordray, 1986, 2000). As seen in Figure 6.1, Path C requires the specification of mechanisms that transform a weak treatment into a strong

---

1Consistent with the simple view of cause–effect congruence, the hypothetical examples shown in the figures assume a direct correspondence between the size of the effect and the size of the cause. This, of course, is too simplistic inasmuch as there are no constraints on this correspondence.
effect. At the planning stage, Path C probably reflects the “wishful thinking” bias in intervention research implied by Sechrest et al. (1979). However, with a priori acknowledgment that the planned intervention is weak, specification of the mechanisms needed to link a weak treatment to a large outcome and assessment of the plausibility of those linkages might avert further attacks of major social problems with treatments guided by wishful thinking.\(^3\)

Path D requires an explanation of the factors that “dampened” the influence of an initially strong treatment. In the language of Sechrest and his colleagues, one plausible explanation is the lack of treatment integrity—in reality, the presumed strong treatment degraded into a weak one that produced small (or null) treatment effects.

**Judging A Priori Treatment Strength.** Other aspects of Einhorn and Hogarth’s (1986) notions of ordinary human judgments of causality comport nicely with ideas described by Sechrest and his colleagues. In particular, in Einhorn and Hogarth’s terms, causal relations are customarily viewed within a specifiable complex *causal field*. The causal field described by Sechrest and his colleagues recognizes that treatment strength is to be judged in the context of the specific problem to be resolved, the presence of contextual factors, and attributes of the agents (e.g., people) to be affected. However, given the weak theoretical and empirical foundations for most social programs (Lipsey, 1993), we may have to accept a priori judgments of treatment strength that are cruder than implied earlier in this chapter.

**Modeling Treatment Cohesiveness.** The specification of a cohesive model that links the treatment to mechanisms of change and then the linking of these mechanisms to outcomes for a subclass of clients amounts to articulating a causal field of action. Although a normal part of native causal inference (Einhorn & Hogarth, 1986) and a necessary part of intervention research (Sechrest et al., 1979), deriving and judging the cohesiveness of an intervention is challenging, to say the least. Fortunately, within evaluation research and in causal modeling, a host of “methodological tools” and heuristics have been developed to aid in the specification of these models of the causal field.

In particular, explication of what activities and services constitute a program theory is often aided by the creation, refinement, and use of program logic models (Brekke, 1987; Brekke & Test, 1992; Cordray & Pion, 1993; Julian, 1997; Yin, 1997) and program templates (Scheirer, 1996) that depict the program activities in relation to each other and to the expected outcomes for service recipients. A handful of formal measures of strength, based on formal and informal program theories, have been reported in the literature (e.g., Brandt, Kirsch, Lewis, & Casey, 2004; Carroll et al., 1998; McGrew, Bond, Dietzen, & Salyers, 1994; Orwin, Sonnelfeld, Cordray, Pion, & Perl, 1998).

\(^{3}\)Note that, if substantively interesting mechanisms had been known a priori, the treatment would have been rated as a strong treatment under Sechrest et al’s scheme. As research accumulates, once-weak treatments could be better understood as strong ones.
Judging Strength or Cohesiveness. Program articulation (e.g., through a logic model), however, represents half the task. Sechrest et al. (1979) required a second step, an explicit judgment of cohesiveness or strength. We know of no instance in which logic models have been systematically reviewed for their implied strength. However, explanatory models in meta-analysis (see Cook et al., 1992) and in other forms of model-driven synthesis (Cordray & Fischer, 1995) hold some promise. For example, the Prospective Evaluation Synthesis, developed within the Program Evaluation and Methodology Division of the U.S. General Accounting Office (1990), has been used to determine whether the models underlying congressional legislation are likely to work if implemented. After conceptual and operational models are developed for a proposed piece of legislation, evaluation literature is consulted and synthesized to determine whether there is empirical support for each “causal” link in the models.

Direct Analyses of A Priori Treatment Strength. Sechrest et al. (1979) were clear about the distinction between treatment strength and treatment integrity. Integrity is to be assessed in light of the parameters specified by a priori judgments or assessments of the planned strength of the treatment. As such, treatment strength is awarded priority over treatment integrity.

However, a search of abstracts catalogued by PsycINFO produced only 14 documents that contained the words treatment strength in either the titles or abstracts. Of this group, only a handful possessed any relevance to the ideas articulated by Sechrest and his colleagues. After reviewing the available literature on treatment strength in physical medicine (e.g., rehabilitation from stroke), Keith (1997) concluded that there was a paucity of information about the nature and theory of these treatments and argued that an accumulation of research data on dose specificity, treatment effectiveness, and the mechanisms underlying effective treatments not only was feasible but also would greatly benefit the field in such ways as the development of practice guidelines. Two other articles involved empirical efforts to examine strength. Northup, Fisher, Kahang, Harrell, and Kurtz (1997) attempted to empirically ascertain the levels of two labor-intensive procedures (differential reinforcement of an alternative behavior and brief time-outs) that were necessary to achieve the desired reductions in severe behavior problems in 3 individuals with severe developmental disabilities. Braden, McClone, and Pennington (1999) examined the relationship between the amount of time spent in a self-help class designed to teach coping skills, problem solving, and cognitive reframing and outcomes among 313 individuals with systemic lupus erythematosus.

The Counterfactual Model

The counterfactual model (Holland, 1986; Rosenbaum, 2002; Rubin, 1974) begins with the specification of the purest form of causal inference: the effect of the cause on the same individual. As such, the effect of a cause on Person i

---

4 Proponents of the counterfactual model consider causes to be only those variables for which an individual could experience all levels (e.g., treatment = 1; no treatment = 0) of the variable.
can be simply represented as the difference between his or her score on an outcome (\(Y\)) under the treatment (represented as Tx) condition (\(Y^T\)) and the outcome under the counterfactual (non-Tx, or c) condition (\(Y^c\)). Under this model of causality, Holland (1986) and Rubin (1974) argued that the treatment effect—the causal parameter of interest—would be expressed as \(T = (Y^T) - (Y^c)\). This definition of the causal parameter requires an individual to be exposed simultaneously to both conditions. For obvious reasons, such a requirement cannot be met. To circumvent this fundamental problem of causal inference, Holland (1986) described a series of approximations to causal analysis.

**Defining the Effect.** Figure 6.2 provides a schematic overview of the linkage between the each level of the causal variable and the effects. Starting with the right-hand side of the figure, rather than observing the responses of an individual to both the treatment and counterfactual conditions (\(Y\) and \(Y^c\), respectively), the average responses of a group of individuals who experienced the treatment are compared with the responses of another group of individuals who did not experience the treatment. If groups are composed at random, Holland’s (1986) first approximation to causal analysis is the classic randomized experiment (Boruch, 1997).

As shown in Figure 6.2, the casual effect of treatment is actually a relative effect, which can be estimated as \(RE = E(Y^T) - E(Y^c)\), or the difference between the expected values for each group. In practical terms, this translates into the differences between group averages. In the simple example illustrated in the figure, it is expected that the \(T^A\) will produce an average outcome of 75 units of well-being and \(T^c\) will produce an average outcome of 25 units. The relative
effect of T⁰ is an increase of 50 units of well-being.⁴ For this to be an adequate approximation, a number of conditions and statistical assumptions must hold. In any given comparative causal study, these assumptions may be important or trivial, depending on the specific details of the research design that is used (Rosenbaum, 2002).

**Defining the Cause.** If the effect in a cause–effect relationship is defined as the relative, average difference between the means of the groups (Tx and C) on the outcome (Y), the cause also needs to be regarded as the difference between the treatment conditions (e.g., number of sessions, strength as assessed by cohesiveness). As shown in Figure 6.2, to produce a relative effect \( RE = E(Y^T) - E(Y^C) \), there must be a relative difference between groups on the essential causal variables (Cordray & Pion, 1993; Mackie, 1974).⁵

**Relative Strength.** Differences with respect to a causal variable reflect the relative strength (or RS = \([T^0 - T^x]\)) of the contrast. This can be expressed in both qualitative and quantitative terms. Qualitatively, the focus is on the presence/absence of causal elements or the alternative configuration of causal elements. Quantitative scales emphasize differences in the levels of causal variables within the groups. In Figure 6.2, T⁰ and T⁴ are 75 and 25 treatment strength units, respectively, with the relative strength being 50 treatment strength units. Orwin et al. (1998) used the number of program elements that were possible and received to index the strength of each innovation in a multisite evaluation of programs for homeless substance abusers.

**Treatment Differentiation: An Added Complexity.** Assessing the relative strength of treatment conditions is comparatively simple in the dose–response models developed by Howard, Kopta, Krause, and Orlinsky (1986). As the treatment conditions become more complex, so does the assessment of treatment differentiation. Although Waltz, Addis, Koerner, and Jacobson (1993) cast their discussion of treatment differentiation in terms of assessing therapist adherence and competence, their ideas are quite general. Their scheme for treatment differentiation includes the assessment of items that reflect (a) treatment components and therapist behaviors that are unique to the treatment modality and essential to it, (b) treatment components and therapist behaviors that are essential to the treatment but not unique to it, (c) treatment compo-

---

⁴Alternatively, the relative effect can be indexed as a statistic defined by Cohen (1988): the effect size. An effect size is a descriptive statistic that expresses the difference between the means of the treatment \( (M_T) \) and counterfactual condition \( (M_C) \) on an outcome in standard deviation units, or \( ES = (M_T - M_C) / SD_{pooled} \), where the pooled standard deviation is the weighted average of the standard deviations within each condition. Cohen posited that effect sizes of .20, .50, and .80 can be interpreted as representing small, medium, and large effects, respectively. The ES index is particularly useful as a means of establishing a statistical association (covariation) when the study sample sizes are small and, as a consequence, statistical precision (or power) is low.

⁵For the purposes of statistical analysis, each causal condition is represented by a group-assignment variable \( (G) \). Each causal condition is assigned a value (e.g., if \( G = 1 \), if \( G = 0 \) on a group-assignment variable to test whether group means on \( Y \) differ after exposure to the conditions.
ments and therapist behaviors that are compatible with the treatment model but not unique or needed, and (d) components and behaviors that are not supposed to be provided (proscribed).

Following the scheme proposed by Waltz et al. (1993), Carroll et al. (2000) developed and validated a 55-item measure (the Yale Adherence and Competence Scale) that is designed to be used by most behavioral programs that target drug addicts. Similarly, the Dartmouth Assertive Community Treatment Scale is designed to obtain program fidelity (integrity) information on programs that use the assertive community treatment (ACT) model for people with serious mental illness. Administered by interviewers, this 28-item instrument not only measures implementation of ACT (characterized by a team approach with shared caseloads, frequent staff meetings, intensive community-based services, and assistance with daily living skills) but also discriminates it from other types of case management programs and has shown to be related to certain client outcomes (McGrew et al., 1994; McHugo, Drake, Teague, & Xie, 1999; Salyers et al., 2003; Teague, Bond, & Drake, 1998).

The Dose–Response Models

One of the most intuitively appealing models of the relationship between the magnitudes of the treatment and the outcome is the dose–response model. Although it has a well-established place in psychotherapy research (Howard et al., 1986), it has been less investigated in other areas or has been found to yield no strong evidence for a relationship (e.g., National Institute of Child Health and Human Development Early Child Care Research Network, 2003).

Originating in the biological sciences to investigate the potency of stimuli with living participants, this research focuses on examining the impact of different doses of a stimulus on a specific response variable (Hansen, Lambert, & Forman, 2002). Unlike the contiguity model, which emphasizes the a priori consideration of strength, the dose–response model seeks to answer the question, "How much treatment is enough?" A prototypical pattern for the relationship is shown in Figure 6.3, in which treatment strength, as measured by treatment contact (or exposure) hours, is linked to treatment outcomes as a negatively accelerated curve. In our hypothetical example, 10 contact hours corresponds to a success rate of P = 0.20 (20% of clients report no symptoms), 30 hours is associated with a success rate of P = 0.55, and 60 sessions corresponds to a success rate of P = 0.85. Not surprisingly, this example mirrors empirical results in psychotherapy (e.g., Howard et al., 1986; Kadara, Lambert, & Andrews, 1996; Lambert, Hansen, & Finch, 2001). As intuitively appealing as the dose–response relationship appears, it is not without some difficulties. In particular, it relies on results for clients with different levels of exposure to the treatment, raising questions about the internal validity of the claim describing the functional form of the relationship. However, the logic of the dose–response model is not restricted to naturally occurring variation in treatment receipt. For example, Barkham et al. (1996) used the counterfactual model by randomly assigning individuals to receive different doses (8 or 16 sessions) of two types of time-limited psychotherapy. The encouraging news,
at least to our untrained eyes, is that over the range of treatment strength (up to 16 sessions), their results are not inconsistent with the negatively accelerated functions previously reported, and their model has the additional benefit of examining treatment integrity (therapist adherence) to the two therapy models being tested.

The dose–response model appears to have an advantage over the counterfactual model when data are available on a broad range of treatment strengths. Specifically, exclusive reliance on the counterfactual model, with its emphasis on relative strength and relative effects, could yield results that are less useful to practitioners than is desired (Yeaton & Sechrest, 1981b). In some extreme instances, very potent treatments could be ruled ineffective when their success rates are very high. In Figure 6.4, we array the hypothetical results of three studies that investigated the effects of a treatment model. Study 1 compared 10 versus 20 contact hours of the treatment, with success rates of 0.21 and 0.40, respectively (effect size = 0.42). Study 2 contrasted 30 versus 40 contact hours, exhibiting success rates of 0.50 and 0.70 (effect size = .41). By conventional standards, both studies revealed moderate relative effects. However, interpreting just the effect size ignores valuable information about the effectiveness of treatments. Specifically, although the effect sizes are roughly equivalent, without taking into account the level of success for each treatment, we would overlook the fact that the success rates of both conditions in Study 2 greatly exceed those in Study 1. In the extreme, a study in which both conditions involved substantial doses of the treatment (Study 3) might produce an effect size that is neither statistically nor practically meaningful, yet the success rates of both noticeably exceed those in all conditions of Studies 1 and 2.
TREATMENT STRENGTH AND INTEGRITY

Figure 6.4. Relative effects and treatment outcomes for three hypothetical studies. T = treatment; C = counterfactual condition; ES = effect size.

Treatment Integrity

Unlike treatment strength, as defined by Sechrest and his colleagues, the literature on treatment integrity is voluminous. Concern about treatment integrity—the extent to which the treatment is carried out according to its model, plan, or theoretical specifications—dates back to the 1950s in psychotherapy (Bond, Evans, Salyers, Williams, & Kim, 2000; Moncher & Prinz, 1991). Indeed, there is little doubt that concern over integrity could not be traced back to the earliest moments of civilization. For example, major German beer makers proudly advertise that their product is brewed in accordance with the German Purity Law of 1516! In this section, we focus on (a) a proposed connection between treatment strength and integrity, (b) the relationship of treatment integrity to similar constructs, (c) the prevalence assessment of treatment integrity in effectiveness studies, and (d) what is known about factors influencing the integrity with which treatments are delivered and received.

Bridging Treatment Strength and Integrity: Achieved Relative Treatment Strength

According to Sechrest et al. (1979), the relationship between treatment strength and treatment integrity is reasonably straightforward. Whereas treatment strength is the a priori magnitude of the treatments ($T^d$ and $T^c$), integrity refers to departures from their intended strengths. Departures can result from
numerous factors, including contextual, organizational, staffing, and participant variables.

Although it is theoretically possible that treatment strength could exceed original expectations on implementation in the field, the more likely scenario is that it is eroded by these departures. However, a variety of processes described by Shadish et al. (2002) can operate to transform the counterfactual treatment to more closely resemble the target treatment condition. If elements of the treatment condition are also provided to participants in the counterfactual group (e.g., compensatory equalization, treatment diffusion) or through other mechanisms (e.g., compensatory rivalry, crossing over of participants), the counterfactual condition, as realized in practice, could resemble the theoretical form (T') less and the target treatment condition (T') more than was intended.

Accordingly, relative strength becomes achieved relative strength at the study's conclusion because of factors that influence the integrity of both T'x and T'. To capture the distinction between relative strength and achieved relative strength, Figure 6.5 adds T'x and T'c to represent the achieved strength of each group in relation to T'x and T'c, respectively. Achieved relative strength (ARS) can be specified as ARS = [T'x and T'c]. In our hypothetical example diagrammed in Figure 6.5, ARS is 20 strength units (60 – 40), rather than the 50 strength units that were expected a priori. (See Figure 6.3.) The relative effect then shrinks, on average, to 20 well-being units.

The factors discussed in this section are thought to account for, in part, the transformation of T'x → T'x and the conversion of T'c → T'c or the treatments as planned (T's) versus the treatments as experienced (t's). Whereas relative treatment strength refers to the differences between T'x and T'c, the net effect of problems associated with a loss of treatment integrity is the achieved relative treatment strength, or T'x - T'c. We suspect that part of the reason that treatment
integrity has received so much more attention in the literature than has treatment strength is because, at the end of the study, it is the achieved relative treatment strength that matters most in explaining the presence or magnitude of treatment effects. Lipsy (1990) provides guidance on how to maximize the statistical power of an intervention study.

The loss of statistical power from the reduction in treatment strength represents a main concern of Sechrest et al. (1979). We might speculate that one reason for their repeated advice to first test the strongest possible treatment was to maximize the achieved relative strength between conditions. Boruch and Gomez (1977) provided a truly elegant analysis of the consequences resulting from the loss of treatment integrity for statistical power.

Comparable Concepts

Naturally, during its long history in the sciences, the notion of treatment integrity has been captured under a number of different labels. Our quick literature search (in PsycINFO) uncovered nearly 5,000 references to treatment integrity and its related concepts. Some of these alternative labels include treatment fidelity, compliance, and implementation. For example, Moncher and Prinz (1991) defined treatment fidelity in terms of the experimental tradition of a “manipulation check.” Compliance refers to the extent to which clients accept treatment services, and implementation is a broad term associated with the installation of an intervention (Lipsey & Cordray, 2000). Other terms focus on specialized aspects of the treatment implementation process. These include the adherence of practitioners to the treatment protocol and competence. Adherence generally refers to whether a practitioner uses treatment approaches prescribed in the treatment manual and avoids the use of approaches that are proscribed (Waltz et al., 1993). Adherence, then, is similar to the notion of treatment purity described by Scott and Sechrest (1989). Competence refers to the skill of the practitioner in delivering the treatment.

Elements of Integrity Assessments

Several excellent reviews of the elements of treatment integrity are available (e.g., Carroll et al., 2000; Waltz et al., 1993). In particular, Carroll et al. (2000) identified the “technology model of psychotherapy research” as characterizing the state-of-the-art approach to assessing aspects of treatment integrity. According to Carroll et al. (2000), the technology model

requires the specification of behavioral therapies in terms of their “dose” (the frequency and number of sessions), their active and inert ingredients (clarification of the unique and common elements of the therapy), the conditions under which they are administered, and assessment of whether the treatment was adequately delivered to all patients (compliance). (p. 225)

Consistent with suggestions made by Sechrest et al. (1979), Carroll et al. (2000) suggested that certain study features have become “virtual requirements” in
clinical trials. Efforts to promote treatment integrity include (a) the specification of treatments in manuals, (b) the training and supervision of implementers, and (c) the monitoring of treatment delivery.

**Integrity Assessment in Effectiveness Studies**

Historical reviews of treatment fidelity in drug abuse prevention (Dusenbury, Brannigan, Falco, & Hansen, 2003), psychiatric rehabilitation (Bond et al., 2000), and other helping professions (Moncher & Prinz, 1991) point to the slow but steady growth of empirical assessment of treatment integrity. Several reviews of the extent to which treatment integrity has become a focal feature within outcome and effectiveness research have used Carroll et al.'s (2000) three dimensions to track methodological progress in various fields (e.g., clinical psychology, behavior therapy, marital and family therapy, parent training, drug prevention, alcohol treatment). For example, Moncher and Prinz (1991) examined 359 treatment outcome studies in clinical psychology, behavior therapy, psychiatry, and marital and family therapy appearing between 1980 and 1988. In defining assessment of fidelity, they focused on the use of procedures to promote fidelity (manuals and supervision) and whether aspects of treatment delivery were verified (adherence checks). Over time, the use of manuals increased from 28% (1980–1982) to 39% (1986–1988). Similarly, the use of supervision to promote fidelity increased from 17% to 34%. Adherence checks were conducted more frequently in 1986 through 1988 (31.7%) than in the earliest period of their review (13%, in 1980–1982). Although progress had been made, 55% of the studies ignored issues of fidelity entirely. Moreover, only about 13% of the studies incorporated all three fidelity assessments.


These two reviews confirm the claims made by Sechrest and his colleagues. About the time their first papers were published, less than 20% of studies attended to issues of treatment integrity. In the interim, progress has been made, but there is still a long way to go before treatment integrity becomes a routine part of the portfolio of research practices in studies of treatment effectiveness.

**Factors Affecting Treatment Integrity or Relative Treatment Strength**

Cataloguing the potential influences on treatment integrity is greatly facilitated by the fact that factors are often nested or hierarchically organized. Behaviors of clients are influenced by the behavior of intervention staff; intervention staff can be influenced by policies and practices of their organization; and organizations can be influenced by other organizations (at the same level or operation) or at higher levels in the structure (e.g., rules, regulations, and laws of local, state, and federal governments). The cumulative influence of integrity problems emanating at higher levels of this hierarchy can have dra-
matic effects on services received by groups of clients. If treatment integrity is breached at the clinic level (by policy or lack of support for component X), the actions of all therapists will be constrained and the services will not be delivered and received by all clients. Similarly, if clients refuse to participate, the strength of the treatment law, regardless of whether the staff are willing and able to provide services, will be diluted. As such, there are multiple ways to adversely affect treatment integrity.

**Participant Compliance and Engagement.** The literature on treatment integrity focuses a great deal of attention on aspects of the treatment delivery process that the researcher has some hope of controlling through training, supervision, or treatment reengineering (e.g., Duan, Braslow, Weiss, & Wells, 2001). The behavior of participants plays an equally important role in determining the overall achieved relative strength of an intervention or treatment. In particular, their compliance with and engagement in the treatment protocol can dramatically affect how much treatment is received. Stecher et al. (1994) reported that 40% of clients assigned to (residential or nonresidential) treatment failed to become engaged (in the treatment) even for 1 day. Less than 10% of clients graduated from the treatment conditions. In another example, in assessing the discriminability conditions and integrity of Project MATCH treatments, Carroll et al. (1998) found that clients completed between 60% and 80% of the potential outpatient and aftercare treatment sessions. Although there were no between-condition differences in integrity, the noncompletion rates suggest that, on average, the conditions were not as strong as they might have been had all participants completed their full package of care.

**Organizational Factors.** Implementation of interventions is not always a smooth process within established or new organizations. Consequently, the strength of the treatment can vary over time. It may take a new or revamped program literally years to be up and running at full strength if, indeed, that ever occurs. Moreover, mature programs are vulnerable to periods when they deliver services incompletely, inconsistently, or not at all to a noticeable proportion of their target clients. For example, Stecher et al. (1994) compared residential and nonresidential treatments for dually diagnosed homeless adults. Both interventions involved two phases; the first had eight goals (e.g., client engagement, retention, assessment, and treatment planning) and an intended duration of 3 months. On the basis of interviews with treatment administrators and staff, it was found that less than 20% of the goals were accomplished for the nonresidential program during its first 9 months of operation. However, implementation had reached 75% by the 15th month and then remained stable. Even though the residential treatment had been operational for 4 years prior to the study, it was not until the 18th month that all eight program goals had been achieved. Orwin et al. (1998) showed that treatment strength and relative strength ebbs and flows over the course of a program’s time line. It is not unreasonable to believe that different client cohorts attained differential levels of outcomes. The dose–response model for linking outcomes to differential experiences within programs may provide a useful approach to capitalizing on this natural variation within programs.
The idea of investigating dose–response relationships is not confined to individuals or individuals within programs. The “dosages” provided by different sites with the same treatment model also may differ. For example, McGrew et al. (1994) measured the fidelity (the achieved strength) of 18 assertive community treatment programs and found variability in fidelity that correlated with a program-level client outcome. Similar achieved strength and outcome relationships have been found by Becker, Smith, Tanzman, Drake, and Tremblay (2001), and Yeaton (1994).

Outside Services. Because some community-based interventions rely on services outside the direct control of a treatment, assessing achieved strength requires the inclusion of these outside (or wraparound) services. An observational measure has been developed to measure the key elements of the wraparound approach for children with serious emotional disturbance and their families within a system of care (Nordness & Epstein, 2003). Wraparound services can also reduce the achieved relative strength of treatment conditions by enhancing the amount of treatment provided to participants in the control condition. Carroll et al. (1998) found that 28% to 38% of outpatient clients had involvement in at least one significant outside treatment session.

The Potential Influence of Changes in the Policy Context. Experimental assessments of interventions can take years to plan and execute. At the same time, changes in reimbursement policies and shifts in priorities of administrations can increase or decrease the availability of treatment resources that can affect the strength of the target treatment, the counterfactual treatment, or both. McHugo et al. (1999) reported that state-based changes in reimbursement for community outreach created incentives for traditional case managers to increase outreach (enhancing the strength of the control condition). Federal Medicaid changes for psychiatric hospitalization altered the payment structure, reducing the strength of treatment for both groups because the cap on the length of stay was reduced.

Changes in Technology. If the public sector (policy context) can interject its influence on the integrity or strength of treatments, so can the private sector. Again, McHugo et al. (1999) noted that clozapine emerged at the same time that their interventions were being tested, creating a competitor to their treatment theory. Shifts in economic circumstances can certainly affect the potential effects of interventions that rely on work or have employment as an outcome.

The foregoing discussion of treatment integrity was intended as a “snapshot” of progress and issues that have emerged since the time Sechrest et al. (1979) called for greater attention to the nature and strength of interventions. Given the volume of material that has appeared, our summary has attempted to highlight the contextual, organizational, and individual (both clients and implementers) factors that can and do exert an influence on treatments, as realized in the field.
Status of Treatment Strength and Integrity

Sechrest and his colleagues framed intervention research from two perspectives: (a) At the conceptual level, they stressed the a priori magnitude and theoretical cohesiveness of the treatment, referring to this as treatment strength; and (b) at the operational level, they defined treatment integrity as the extent to which the theoretically grounded treatment was implemented as planned. In the two decades since they issued these definitions, the field has witnessed increasing attention being directed at dimensions of treatment integrity. The question implied by the definition of treatment strength—"How much treatment is enough and how should it be optimally organized?"—has received less direct attention. Indirectly, however, many of the ideas underlying treatment strength have been incorporated into research practices through the development of theory-driven evaluations, logic models, program templates, and treatment manuals.

In its most intuitive form, the idea of treatment strength is quite appealing and understandable. In the context of models of treatment effectiveness research, the notion of relative treatment strength probably fits better. Furthermore, because the net influence of factors that dampen treatment integrity represents the achieved relative treatment strength in a comparative study, the distinction between strength and integrity may have been lost. We continue to believe that the a priori consideration of the strength of each treatment condition remains a fundamental question in intervention research.

Although nearly 5,000 references related to treatment integrity were uncovered in preparing this chapter, the most up-to-date reviews of effectiveness studies still suggest that the measurement of strength or integrity is not yet universal. In particular, assessing treatment integrity is dependent on having adequate measurement tools. Unless one simply relies on client participation data (e.g., length of stay, number of therapy sessions, or class attendance), the development of such measures is labor intensive and often the result of a lengthy program of research. This is particularly true for interventions that are complex or multifaceted. Given the number of dimensions that are to be assessed, data collection is labor intensive, often involving interviewers, observers, or multiple sets of respondents (e.g., clients, practitioners, and parents).

Much of the impressive work done to date on features of treatment integrity is the result of a program of research conducted on treatments with well-articulated theories, manuals, training materials, established measures, and so on. The majority of programs (interventions) in local organizations are not nearly as sophisticated. For these researchers, the "high-end" methods reported in this chapter are likely to be of little consequence.

Returning to the issues raised by Sechrest and his colleagues, there are three fundamental questions pertaining to treatment strength and integrity: (a) Is there a compelling enough rationale for the intervention to produce the desired effects? (b) Is the intervention sufficiently different from what would have been received had it not been installed? and (c) Did participants in the treatment condition get more services or treatment than those in the counter-
factual condition? At a minimum, answering the first two questions involves critical thinking, not an elaborate assessment package. Finding simple indicators of treatment receipt (e.g., number of treatment sessions attended) may be enough to answer the third question. To the extent that expected outcomes are confirmed, important rival explanations are ruled out, and the achieved relative strength is adequate (the cause is “ruled in”), the particular intervention would become a candidate for a longer and more intensive program of research. Studies need to be done “well enough” to avoid overlooking an intervention that has promise.

Sechrest and his collaborators, by framing the questions concerning strength and integrity of treatments the way they did and so early in the history of evaluation research, have contributed greatly to the search for effective ways of discovering how to ameliorate adverse social conditions.

References


