

The Meanings and Measurement of Clinical Significance

Alan E. Kazdin
Yale University

The previous articles in this special section make the case for the importance of evaluating the clinical significance of therapeutic change, present key measures and innovative ways in which they are applied, and more generally provide important guidelines for evaluating therapeutic change. Fundamental issues raised by the concept of clinical significance and the methods discussed in the previous articles serve as the basis of the present comments. Salient among these issues are ambiguities regarding the meaning of current measures of clinical significance, the importance of relating assessment of clinical significance to the goals of therapy, and evaluation of the construct(s) that clinical significance reflects. Research directions that are discussed include developing a typology of therapy goals, evaluating cutoff scores and thresholds for clinical significance, and attending to social as well as clinical impact of treatment.

Clinical significance refers to the practical or applied value or importance of the effect of an intervention—that is, whether the intervention makes a real (e.g., genuine, palpable, practical, noticeable) difference in everyday life to the clients or to others with whom the clients interact. The assessment of clinical significance represents an important advance in the evaluation of intervention effects, including treatment but extending to prevention, education, and rehabilitation as well. Apart from reliability of change or group differences (e.g., statistical significance) and the magnitude of experimental effects (e.g., effect size or correlation), the importance of the change and the impact on client functioning add critical dimensions. Treatments that produce reliable effects may be quite different in their impact on client functioning, and clinical significance brings this issue to light.

The methods of evaluating clinical significance have advanced as well, in no small part by the authors of the previous articles in this series (e.g., Jacobson & Revenstorf, 1988; Kendall & Grove, 1988). In the present series, the authors elaborate on the measures of clinical significance; provide advances in their use, application, and computation; and identify issues pertinent to treatment evaluation more generally. They raise several fundamental issues that serve as the basis of the present comments. Specifically, in the present article, I examine the meaning and interpretation of measures of clinical significance, the importance of relating assessment of clinical significance to the goals of therapy, and the construct(s) that clinical significance reflects. Recommended directions for research are also highlighted and include developing a typology of therapy goals, evaluating cutoff scores and thresholds for clinical significance, and attending to social as well as clinical impact of treatment.

I am extremely grateful for research support from the Leon Lowenstein Foundation, the William T. Grant Foundation, and the National Institute of Mental Health (Grant MH59029) during the period in which this article was completed.

Correspondence concerning this article should be addressed to Alan E. Kazdin, Department of Psychology, Yale University, P.O. Box 208205, New Haven, Connecticut 06520-8205.

Meanings of Clinical Significance

There are key issues that influence or determine whether a change is clinically significant. The issues reflect consideration of the question “What do mental health professionals mean when they refer to change as clinically significant?” The answers to which the question is directed are not, of course, the operational definitions in use but rather the concepts and constructs these definitions are designed to represent.

Amount or Degree of Change

The amount or degree of change is the most striking characteristic of the meaning of clinical significance. Measures reviewed in the previous articles indicate that a rather large, reliable change in symptoms and return to normative levels are primary indexes of clinical significance. Although a large change of the type referred to in the previous articles is important, I believe a more paradoxical claim can be made about the relation of the amount of change and clinical significance. Specifically, I suggest a clinically significant change can occur when there is a large change in symptoms, a medium change in symptoms, and no change in symptoms. The suggestion that any amount of change might be clinically significant is not sophistry but rather conveys that clinical significance can and does mean many things, and these vary as a function of the type of problems and the goals of treatment.

Consider three situations and how the clinical significance of change might be demonstrated. First, consider the situation in which the client comes to treatment with many symptoms (e.g., of depression) and the goal is to reduce or eliminate these symptoms. After treatment, the client's symptoms have been reduced substantially. On standardized measures of depression, and indeed on broader measures of psychopathology, this might be reflected in a statistically reliable change and symptom scores that fall within the normative range. More will be said later about the meaning of a reduction of symptoms and entry into the normative range. At this point, it is important to begin with the notion that clinical significance can mean a large change in symptoms, as thoroughly detailed previously (Jacobson, Roberts, Berns, & McGlinchey, 1999; Kendall, Marrs-Garcia, Nath, & Sheldrick, 1999).

Second, consider the same client but a slightly different outcome. Here the client improves, but at the end of treatment the

client's behavior has not changed enough for it to fall within the normative range. From the researcher's standpoint, the criterion of clinical significance may not be met. However, on a priori and perhaps even commonsense grounds, the change may be important and potentially clinically significant (i.e., in keeping with the definition of making a difference and having a practical value). After all, from the standpoint of symptoms, one can be a little better or a lot better (e.g., fewer or less severe symptoms) without being all better or just like most people (e.g., no symptoms, normative range of symptoms, or recovered). If one is a little better or a lot better, that is important to identify for research and clinical purposes. Sometimes a little means a lot. (I mention the other situation later when "sometimes a lot" may mean "very little.")

For example, a review of psychotherapy for depression suggests that treated cases change but at the end of treatment clients are still more depressed than are normative samples (Robinson, Berman, & Neimeyer, 1990). For severely depressed and suicidal patients, perhaps those who are hospitalized, an improvement might be sufficient to return them to everyday functioning, even though their depressive symptoms are hardly near normative levels (cf. Tingey, Lambert, Burlingame, & Hansen, 1996). For less severely depressed patients, a small change in symptoms also may keep them out of the normative range. Could it be that such individuals have made a clinically important change, even though they do not fall within the normative range? The clients themselves might judge the changes to be quite significant, if clinical researchers introduce the client's perspective on the matter, as raised by previous articles (Foster & Mash, 1999; Gladis, Gosch, Dishuk, & Crits-Cristoph, 1999). The perspective of the person judging clinical significance is relevant and raises a broader issue discussed later. Nevertheless, it is conceivable that a little change goes a long way, could make a great difference to the client (i.e., be clinically significant), and affect his or her functioning in everyday life.

Finally, consider a different situation that arises in psychotherapy in which symptom change is not at issue. Several circumstances may make symptom change not relevant or not the main objective of therapy. It might be that the symptoms are deteriorating and an effective treatment stops or postpones the deterioration. In another situation, perhaps the symptoms cannot be changed very much or at all (e.g., Tourette's syndrome or self-mutilation). The absence of effective treatments or the failure of ordinarily effective treatments that have been applied may lead us to consider that the symptoms are not likely to change. Alternatively, the impetus for seeking therapy may be a personality or character trait that one's dear friends and relatives find annoying. Many of these characteristics are stable and lifelong and might not be expected to change or to change very much. For each of these situations, researchers and clinicians may not be able to do very much in terms of changing symptoms, but they may do a lot in terms of helping people cope with symptoms or improve the quality of life (Gladis et al., 1999).

Helping people cope is relevant to a wide range of issues brought to treatment, as reflected in coping with a personal disability (e.g., disfigurement or loss of mobility); with emergent and emotionally wrenching challenges (e.g., care of a child or spouse with an acute trauma or chronic disability); or with one's past (e.g., guilt, remorse in relation to a parent, or abuse by a parent), present (e.g., diagnosis of a terminal disease or loss of a relative), or future (e.g., *angst* over a personal crossroads or an impending major life event). For purposes of discussion, one might say that the symp-

toms of the client are not the problem. Rather, the goals of therapy include coping with the situation, altering one's views, and taking action to manage the situation.

The three situations converge to make one point; namely, it is conceivable that therapeutic change can be important (i.e., clinically significant) when symptoms change a lot; when they change a little; and when they do not change at all, but the client is better able to cope with them. The determination of clinical significance in these situations is not arbitrary and does not challenge existing measures. Rather, the illustrations convey that clinical significance depends on the problems that are brought to treatment and the goals of treatment, a point underscored by Foster and Mash (1999). The challenge is developing a method of connecting outcome measures and clinical significance on these measures to the goals of treatment.

Key Constructs

A marked change in symptoms could readily signal a clinically significant effect, but few would say that clinical significance is restricted to changes in symptoms. What are the constructs that underlie clinical significance, or what are the defining dimensions? Each of the articles in this series takes up the matter, and several constructs are noted, including symptom change, meeting role demands, functioning in everyday life, quality of life, and subjective judgments. To date, evaluation of clinical significance in treatment outcome research has emphasized symptom reduction and has used one of the procedures developed and discussed in the first two articles in this series (Jacobson et al., 1999; Kendall et al., 1999). In research, clients are usually recruited on the basis of meeting inclusion criteria for a particular level of symptoms, and hence symptom reduction may be especially relevant.

Symptoms are important, but it is interesting to consider their role in treatment referral and treatment more generally. The number of people with symptoms in everyday life (and who are not in treatment) is probably quite high. Researchers know that approximately 18% to 20% of children, adolescents, and adults in everyday life meet criteria for a psychiatric disorder in a given year (e.g., Burke, Burke, Regier, & Rae, 1990; United States Congress, Office of Technology Assessment, 1991). These refer to *diagnoses* (sets of multiple symptoms). The rates of people having one or two symptoms or having subsyndromal disorders (sets of symptoms that fall below meeting diagnostic criteria) necessarily must be higher. Most people with symptoms and disorders are not referred to or receive psychotherapy. Of course, just because individuals do not come to treatment does not mean they are functioning well. Seeking treatments depends on a number of factors, including the nature of the clinical problem, the availability of resources, and cultural views about seeking treatment. Yet it is quite feasible that large numbers of individuals with disorders or sets of symptoms are managing or functioning adaptively, even if not optimally. Level of symptoms may not be the basis of receiving or evaluating treatment or the primary determinant of functioning adaptively or well.

Impairment may be much more critical than symptoms for entering treatment. Impairment includes difficulties in meeting role demands, interacting with others, and being restricted by what one can do in settings, situations, and activities in which one is involved. Impairment is related to, but readily distinguishable from, symptoms and disorders (Sanford, Offord, Boyle, Peace, &

Racine, 1992). Moreover, impairment is related to seeking treatment. In the case of child treatment, for example, impairment more than symptoms predicts the likelihood of being referred for treatment (Bird et al., 1990). Among adolescents, level of impairment at the end of treatment predicts the likelihood of relapse (Lewinsohn, Seeley, Hibbard, Rohde, & Sack, 1996). In adult therapy, many individuals who do come for therapy do not meet criteria for diagnoses, at least when assessed through standardized methods (e.g., Howard, Lueger, & Kolden, 1997). Quite possibly, functioning in everyday life, apart from or in combination with symptoms, is the basis for seeking treatment.

Clearly, symptoms and impairment can be related, and often both are core diagnostic features of various disorders (e.g., substance abuse and schizophrenia, to mention two; American Psychiatric Association, 1994). However, symptoms and impairment may not invariably be related or, in any given case, necessarily highly related. This is worth mentioning because treatment might reduce symptoms (e.g., to normative levels) and not necessarily affect impairment in important (clinically significant) ways or fail to reduce symptoms to a clinically significant degree but improve functioning in daily life.

Symptoms and symptom changes are important as indexes of clinical significance, as the previous articles convey. At the same time, there are reasons to be cautious because of the impetus for and conditions related to seeking treatment, the goals of treatment, and the possibility that important therapeutic changes may be unrelated to symptom change. It is meaningful to ask what are the key constructs or dimensions along which clinical significance ought to be evaluated. There are likely to be many constructs and dimensions. Quality of life is a candidate or critical component. A detailed analysis of quality of life nicely illustrates the complexities in conceptualization and assessment of just one of these constructs that may be central to clinical significance (Gladis et al., 1999). Impairment, and no doubt other constructs, might be proposed as well.

Perspectives and Convergence of Measures

In psychotherapy research, there is a long-standing recognition that evaluation of treatment effects entails many different perspectives, including those of the client, those in contact with the client (e.g., spouse, parents, or coworkers), mental health professionals, and society at large (e.g., Kazdin & Wilson, 1978; Strupp & Hadley, 1977). Clinical significance invariably includes a frame of reference or perspective. It is quite appropriate for many treatment goals to ask, "To whom is the treatment effect clinically significant?"

Emphasis on symptom change may reflect the perspective of the investigator. The outcomes regarded as clinically significant are based on what researchers have decided as reasonable definitions and conventions. Symptom change may not reflect what is actually important to the client from his or her perspective. Moreover, a large symptom change may not be reflected in other indexes of practical or applied importance.

The perspective of the client has not been well attended to in the evaluation of clinical significance.¹ Indeed, the client has been largely excluded from the process of defining a clinically significant change. Does the client, at the end of treatment, consider the change to be very important or one that has had palpable impact on his or her life? Of course, there are many cases in which one might

not want the opinion of the client (e.g., a young child with autism, an adolescent with conduct disorder, an adult with borderline personality disorder). Also, as noted by others in this series (Foster & Mash, 1999; Gladis et al., 1999), client opinions (e.g., global ratings or judgments) can be influenced by a variety of factors and biases that could indicate significant change when, in fact, these changes are not reflected in other domains. Indeed, client satisfaction with treatment is not invariably related to changes in symptoms (Lunnen & Ogles, 1998; Pekarik & Wolff, 1996). Thus, assessment and interpretation of the client's perspective raise their own challenges and, hence, cannot be considered as the singular or unambiguous criterion for whether or not treatment has had genuine impact on functioning. Even so, in outpatient treatment for adults, for example, it would seem that the client's perspective is absolutely critical.

It is worth distinguishing for a moment *actual* change and *perceived* change.² As an example of actual change, consider that the client has improved substantially in symptoms at the end of treatment. The actual changes in functioning, as reflected on objective and standardized tests, are obviously important (Jacobson et al., 1999; Kendall et al., 1999). Perceived change on the part of the client or those with whom the client interacts are critical as well (Foster & Mash, 1999; Gladis et al., 1999). The difference between actual and perceived characteristics are readily evident and recognized as important in everyday life as, for example, reflected in *being* competent and *feeling* (perceiving oneself as) competent, in *being* in control and *feeling* in control, and in *being* attractive and *feeling* attractive. In the context of therapy, actual change or level of symptoms at the end of treatment (e.g., demonstrated on standardized measures) is distinguishable from perceived change (e.g., views about how much one has changed along the same dimension as the standardized measures). Actual and perceived change may be correlated. Whatever the correlation is, there may be no relation in a particular individual, and the relation might well be altered with therapy (e.g., because one changes and the other does not or the changes are in different directions). That is, a client may retain his or her symptoms, social ineptness, and brutish demeanor but feel or perceive himself or herself as much better, as noticeably happier, and as having a better quality of life.

To simplify (and dichotomize) for purposes of presentation, envision a 2×2 matrix in which the rows are actual changes (clinically significant vs. not clinically significant) in the clients and the columns are perceived changes (clinically significant vs. not clinically significant). Among the four cells that combine these, the cases in which there are discrepancies are perhaps especially interesting (e.g., the data show that there is a clinically significant change but no change or modest change in client perceptions, and vice versa). The 2×2 matrix is simple because only two perspectives and a single construct (symptom change) are

¹ By "perspective of the client," I do not mean merely the use of self-report measures. Changes on a self-report measure of symptom change do not address the points noted here; to wit, do the changes reflect everyday functioning beyond the symptom measure?

² I recognize that making a distinction between actual and perceived change is equivalent to epistemological *hara-kiri*. On psychological grounds, the distinction requires care as well. Actual and perceived changes are likely to be related in dynamic and reciprocal ways so that change in either one is very likely to prompt a sequence in which many other changes occur.

included. Even so, the illustration raises the issue of perspective and correspondence of measures of clinical significance, and both of these relate to the goals of treatment.

Much of psychotherapy research focuses on cognitively based treatments. These treatments underscore the importance of beliefs, attributions, and thought processes in relation to disorder or bases of therapeutic change. At the same time, there has been less appreciation of cognitive processes as a focus of treatment; changes in such processes are often an end in themselves rather than a means of reducing symptoms or changing disorders. How individuals view themselves, the world, and others are not only critical because they influence depression, but they are also critical by themselves because they are related to the misery that many people bring to treatment, whether or not they are depressed. There are many reasons to be wary of ratings and global judgments, which have been eschewed as major outcome criteria as a way of improving precision in outcome assessment (Kazdin, 1998). Consequently, it would be a mistake, I believe, to turn backward and to reduce outcome assessment to a few ratings of "how was that for you" on a 5-point scale ranging from 1 (*this was fun, but I'm no better*) to 5 (*what an amazing [clinically significant] change*). That said, the client's views of the benefits of treatment are also critical to ensure that the benefits are demonstrated in some other way (Foster & Mash, 1999; Wolf, 1978). Moreover, it is not difficult to validate client views by demonstrating their empirical connections to standardized measures (e.g., Kazdin & Wassell, 1998). Similarly, client views of change could be validated against measures of impact in everyday life.

In statistical evaluation, one recognizes that a result may be statistically significant ($p < .05$), even though the null hypothesis (H_0) is true. That is, the results show that there is a difference, even though there really is no difference in the world, a circumstance referred to as a Type I error. One also recognizes that the results of a study may not be statistically significant, even though H_0 is false (i.e., there really is a difference in the world), a circumstance called a Type II error. No doubt there are conceptually equivalent errors in relation to clinical significance, *mutatis mutandis*. These errors may be evident with a single index of clinical significance, such as symptom scores falling in the normative range. For a given client whose scores fall within the normative range, there may be no real change or a change that is large in everyday life, which the symptom score is assumed to reflect. That is, there is a clinical Type I error in which one finds a change and entry into normative range on the hypothetical "how was that for you" measure, but the symptoms these scores are designed to reflect in everyday life may not have changed for this individual or changed as much as the data suggest. The equivalent of Type I and Type II errors may also be evident when one compares different perspectives (client or mental health practitioner) and when one perspective reflects change and the other does not. To call these "errors" is of course questionable, but Type I and Type II errors convey the concept of discrepancies among criteria and measures and the extent to which conclusions based on one measure of clinical significance correspond to those based on a criterion or another measure.

Assessment of Clinical Significance

Interpretation of Current Indexes

The meaning of current measures of clinical significance is not entirely clear, in part because there has been little validation of the

measures. The reason for the paucity of validity studies may be due to the fact that the measures of clinical significance are not really new or different measures; rather, measures of clinical significance are ways of using other measures, many of which are often well validated. Thus, clinical significance might be inferred by a change on the Minnesota Multiphasic Personality Inventory, the Beck Depression Inventory, or the Child Behavior Checklist. There seems to be no need to validate these measures anew because of the enormous amount of background research, including data on normative samples. Nevertheless, there remain problems in the use of clinical significance indexes even when they are based on well-validated measures.

It might seem obvious that a client whose symptoms are outside the normal range before treatment and within that range after treatment has made a clinically significant change. Assume, for the moment, that the standardization and normative data for the measure have firmly established some normative range, leaving aside the important issues related to deciding the range and cutoffs noted previously (Kendall et al., 1999). At the end of treatment, what does a score in the normal range mean? First, the normative data from standardization samples are rarely based on the scores of individuals tested on two separate occasions. In contrast, when used to evaluate treatment, the measures are typically based on repeated assessment (pre- and posttreatment). The repetition of the measures combined with the context in which the measures are completed (treatment evaluation) makes the score at posttreatment for the clients not necessarily comparable with the data obtained in community samples. Simply stated, the assessment conditions of the normative sample and the client sample at posttreatment are quite different.

Second, identical scores within the normative range from someone in a community sample and someone referred for treatment who has improved may not have the same meaning or correlates. For example, adolescents who met criteria for major depression before treatment may show a clinically significant change insofar as they achieve a cutoff that places them in the nonpsychiatric disorder range once treatment is completed. However, adolescents who no longer meet diagnostic criteria, but who once did, remain different from those who never met these criteria in terms of current impairment and long-term functioning (Gotlib, Lewinsohn, & Seeley, 1995; Lewinsohn et al., 1994).

In general, scores from community and clinical samples that fall within the same range (at the end of treatment) do not necessarily have the same meaning (e.g., concurrent and predictive validity). The quick reply to this concern is that in a treatment study, the use of a no-treatment control handles these ambiguities because clients in this group have repeated assessments and hence provide a basis of comparison. A control group is not relevant here. What is relevant is whether these clients (who fall within the normal range or who made a large change) function well or show palpable effects of treatment. It seems to me that even on the basis of scores on standardized and well-validated measures, one cannot tell.

Interpretation of measures of clinical significance depend on the extent to which they relate to other criteria that reflect an impact on a client's functioning in everyday life or perceived functioning, depending, again, on the configuration of constructs that compose clinical significance. Currently, there are operational definitions of clinical significance, many of which have been refined in remarkable ways. However, the operations to measure clinical significance ought not to be confused with the constructs they operation-

alize. Refinements in how these indexes are computed, statistical niceties (such as controlling for regression to the mean), and tinkering with cutoff scores are all methodologically interesting and perhaps even important, but they do not address the overarching question of clinical impact. The question for any measure or index of clinical significance is the extent to which the measure in fact reflects a change that does have an impact on the individual's functioning in everyday life or a change that makes a difference. Validation is needed to attest to the fact that the measure relates to other indexes of everyday functioning. Stated another way, clinical significance is not being measured because researchers call the measures clinically significant or adopt them for convention. The measures must relate to the construct of interest. Measures of clinical significance require supporting evidence to establish that they actually do reflect important, practical, worthwhile, and genuine changes in functioning in everyday life.

The Criterion Issue or Problem

It is easy to state that current measures ought to be validated, but there is no clear criterion against which to validate the measures. Indeed, there are likely to be multiple criteria, goals of therapy, and perspectives, as raised earlier. Validation steps might begin by comparing treated patients who do show a clinically significant change with those who do not, as reflected on measures suggested in the previous articles. One could show that those who have made a clinically significant change (e.g., fall into the normative range on symptoms) have higher mean scores on measures of marital adjustment, adaptive functioning, quality of life, and other such outcomes when compared with those who have not made a clinically significant change. Yet this does not establish that those identified as showing a clinically significant effect are doing better in their everyday lives.

Among the alternatives for selection of criteria against which measures could be validated would be to identify the clients who state that they have made an important, worthwhile, and genuine change over the course of treatment. What are the predictors and correlates of these statements? A similar case might be made by asking others (i.e., persons with whom the clients interact) whether differences in treatment are clear and important; this, too, could be a criterion to help develop the construct or latent variable (clinically significant change) and determine how various measures relate to that construct.

There may be no single criterion that can be used to validate existing measures of clinical significance. No doubt some treatment effects and clinical foci may be more easily validated than others. For example, therapy as applied to health domains may provide criteria that are more readily assessed. Treatment of obesity or cigarette smoking, as two examples, might be able to connect status at posttreatment (e.g., percentage overweight or number of cigarettes smoked) to other outcomes that can serve as validation for deciding if a clinically significant change or improvement has occurred. Falling within a range of 10% above normal body weight, as opposed to 50% above normal body weight at pretreatment, may greatly reduce the risk for all sorts of diseases and be used as a reasonable criterion for clinical significance. Here the use of normative data has validation evidence in its behalf because the data (10% above body weight) can be related to all sorts of other criteria (e.g., morbidity and mortality). Of course, just because health measures are available does not invariably

provide a rock-solid criterion for clinically significant change, but the grounding is better if one can relate level of functioning to the likelihood (risk) of other outcomes.

In general, much more attention ought to be given to the criteria that are used to define clinical significance. Measures in current use warrant validation in relation to those criteria. Without such validation, it is unclear what these measures assess beyond the descriptive statement of an individual's score on that measure. Stated another way, one can say that the results of treatment produced clinically significant changes, but one must bear in mind that these changes may not have any impact on client functioning in everyday life, unless the indexes have been related to such other measures.

General Comments

A standard assessment recommendation is to convey the need for multiple measures. This is intended to refer to multiple measures of a given construct because no single measure can capture all of the components and each individual measure has a method component (e.g., type of measure or reactivity of assessment) that can contribute to the score and its interpretation. These concerns are relevant to the assessment of clinical significance. However, there are prior concerns that warrant assessment consideration. First, there are multiple constructs and meanings of clinical significance in light of the various clinical problems and goals of therapy. These warrant elaboration at the conceptual level. Second, can one translate these meanings into criteria to validate measures or indexes of clinical significance? What index can be used to reliably and validly reflect the likelihood that someone has made an important change? Also, one ought to ask the same questions about the indexes that are used to reflect the likelihood that the client (or others affected by the client) perceives that there has been an important change.

These questions are fundamental to assessment and evaluation of clinical significance. However, they raise broad issues about the multiple purposes and goals of treatment and the feasible, realistic, and ideal outcomes that may result. The implications are broad because they have bearing on such weighty topics as identifying empirically supported treatments. One may wish to judge treatments on the extent to which they change symptoms, but the results could be quite different if other criteria were used, such as impairment, quality of life, or impact on others (see Kazdin, in press).

Research Directions: Briefly Noted

Classification of Goals and Problems

There are many reasons for seeking treatment, and no doubt these can vary widely over the course of childhood, adolescence, and adulthood. Clinically significant therapeutic change is probably quite different in meaning as a function of the different foci of treatment. Perhaps even when the focus is similar (e.g., anxiety and depression), the meaning and measure of clinical impact may vary over the course of development. A key question for research is whether the problems, goals, or foci of treatment can be categorized in some way to alert psychologists to the most relevant means of assessing clinical significance.

Perhaps a typology of treatment goals or treatment foci can be developed. This is not a classification of clinical problems (such as disorders) but a typology at a higher level of abstraction to address the main goals of treatment. As an illustration, one might say that treatment goals include (a) reducing symptoms; (b) improving interpersonal relations and role functioning; (c) enhancing self-esteem and confidence; (d) enhancing the capacity to cope with or reconcile a particular situation, crisis, or problem; and (e) clarifying or addressing issues related to a past, current, or impending situation. These are not listed to propose a complete typology but rather to convey the broader issue that a typology of goals may be instructive. Needless to say, treatment may have many goals, goals are not independent, and goals can and do change.

The purpose of a typology would be to call forth those types of clinical significance that are most relevant. Making a large change in symptoms and falling into the normative range could be the primary or exclusive index of clinical significance, but it is not necessarily relevant when the primary goals of treatment are Goals b through e above. Also, for some therapy, the outcome is not as critical as the process (i.e., working on issues, clarifying meaning, soliciting someone else's perspective, and having a friend). The ride, as it were, may be as or more important than arriving at a destination. This may not be the focus of therapy in outcome research, but it addresses situations in which clients search for meaning and do so through psychotherapy. There is no reason to emphasize this latter focus here except to convey how the goals of therapy can vary. Perhaps researchers and clinicians would profit from a way of identifying the primary domains of clinical significance to which they ought to attend and from a set of measures to operationalize these. Current measures of clinical significance do not begin with the view that there are multiple goals of therapy and that clinical significance is defined in relation to those goals. Measures also do not begin with the clients' views of what an important change would denote if treatment were helpful or wildly successful.

Perhaps insufficient appreciation of client views about the goals of treatment also contributes to some of the discrepancies in treatment outcomes more broadly. For example, among clients who drop out of treatment "prematurely" and against therapeutic advice, a rather significant proportion has definitely improved (Kazdin & Wassell, 1998). It is likely that the goals of treatment of the clients were achieved, even though the goals of the therapist may not have been. Lack of correspondence between client and therapist or other measures is not a methodological "problem" but rather a substantive and conceptual issue about the goals of treatment and the criteria for evaluating impact.

Cutoff Scores and Thresholds of Clinical Significance

There may be use in referring to clinical significance categorically (e.g., in or out of the normative range, recovered or not recovered). This may be particularly useful or clear when symptoms are eliminated (e.g., no more panic attacks, tics, or encopresis). More often than not, the changes are on a continuum. Any cutoff point to determine whether the change is clinically significant will raise the same issue. (For example, some individuals who are considered to have changed to a clinically significant degree did not change on the criterion; others who did not change to a clinically significant degree on the measure did change on the criterion.) One can investigate the classification of individuals at

the end of treatment to determine which cutoff is the best at capturing those cases that made a clinically significant change. Using some cutoff based on normative data or a degree of change from pre- to posttreatment may not necessarily capture that.

Research is needed that evaluates alternative cutoff points and their utility in defining a clinically significant change. Central to this research is development of a criterion (or set of criteria) on which to judge the extent to which a particular cutoff in fact identifies individuals who have changed in marked ways, as discussed previously. On the measure of clinical significance (e.g., normative data), any particular cutoff score (or range) is likely to identify *true positives* (those who are correctly identified as having made a clinically significant change on the measure, such as falling within the normative range, and who show an impact on the criterion measure from everyday life), *false positives* (those who are considered to have made a clinically change on the measure but not on the criterion measure), *true negatives* (those correctly identified as not having made a clinically significant change on the measure or on the criterion), and *false negatives* (those who did not show a clinically significant change on the measure when, in fact, treatment clearly had an important impact on the criterion). The cutoff point to maximize correct identification of cases that make a clinically significant change can be determined empirically.³

Insofar as clinical significance includes client report, there is a related matter of the threshold that individuals have for saying, perceiving, and believing that an important (clinically significant) change was made in treatment. It is readily conceivable that two clients coming to treatment with the same set of symptoms and who change equally (and fall within the normal range) will view their change quite differently. One might see the change as clinically important, and the other might see the change as nugatory. This is why client perception (e.g., the person seeking treatment) may be critical in many applications of therapy.

The point about thresholds extends beyond the perception of the client. It is quite possible that a given change in symptoms or functioning for two clients may in fact have a different impact on their lives. A reduction in marital conflict that is medium to large for two couples may be quite sufficient to improve and preserve the marriage for one couple but not enough to achieve these ends for the other couple. The reason might be driven by variation between the families in other factors, including characteristics of the parents (e.g., parent history of divorce or psychopathology) and family (e.g., socioeconomic disadvantage or strains of child rearing), and contextual influences (e.g., availability of social support or adequate day care), to mention a few. The larger point is that the value, significance, and impact of a therapeutic change of a given magnitude may vary considerably. Theory and research that elaborate the impact of diverse influences on clinically significant

³ Tests of sensitivity and specificity are used to delineate different cutpoints to identify cases. In the present context, *sensitivity* is the probability of showing a clinically significant change on an outcome measure (e.g., falling within the normative range) among those individuals who have made a clinically significant change in everyday life (e.g., on another criterion). *Specificity* is the probability of *not* showing a clinically significant change on an outcome measure (e.g., not falling within the normative range) among those individuals who have *not* made a clinically significant change in everyday life. Methods for evaluating sensitivity and specificity, elaborated elsewhere (see Kraemer, 1992), are quite relevant to the evaluation and validation of measures of clinical significance.

change could have rather important implications for therapy more generally.

Social Significance

Clinical significance focuses on the importance or applied value of the change in everyday life. Although the emphasis has been on the individual client in therapy, the thrust of the concept draws attention to a broader question. Is there an equivalent domain of interest that focuses on the importance of the impact of treatment for society at large? The notion of social validity encompasses evaluation of treatment goals, procedures, and outcomes, as elaborated by Foster and Mash (1999). Social significance as noted here is restricted to the outcome focus; that is, to what extent does the intervention produce outcomes that are important to or have an impact on society?

One usually considers impact on society to apply to large-scale interventions or programs (e.g., prevention, education, legislation, or social and public health policy). Psychotherapy is conducted on a relatively small scale, and hence having a significant impact in this way (e.g., reducing the proportion of people with a disorder at a given point in time [prevalence]) may be unlikely. Of course, large-scale impact ought not to be ruled out as new technologies deliver therapeutic interventions through computers, the Internet, and television (see Marks, Shaw, & Parkin, 1998; Newman, Consoli, & Taylor, 1997).

Rather than large-scale impact, social significance can also refer to changes on measures that are important to society. Clearly one example might be cost, particularly, cost savings or cost benefit (Yates, 1995). However, there are also other measures, including rates of arrest, truancy, driving while intoxicated, illness, hospitalization, and death. Does an intervention have an impact on measures for which there is social interest? Sometimes the measures and criteria for clinical and social significance may be the same or at least very similar. For example, reduction of antisocial behavior can be relevant to the individual (e.g., measures of symptoms) and to society (e.g., arrest rates). Often the measures and indexes will not be the same, and clinical and social significance address different dimensions. For example, individuals who abuse alcohol may stop drinking (clinically significant change) but not show any reductions in driving accidents (in light of the abuse of another substance that impairs driving).

The research agenda needed to clarify and elaborate the meaning of clinical significance is sufficiently long. Adding a line of research on social significance is hardly needed. However, I am not suggesting anything new. The practical, applied, and real-world impact of interventions has always been a concern in therapy. The societal, rather than the clinical, aspect has come to the fore in light of issues related to managed care, reimbursement, and accountability. As already mentioned, measures of cost, which are included in many outcome studies, reflect this concern. It is not a leap to evaluate the impact of treatment on measures of interest to society as an extension of clinical significance.

Discussion and Perspective

The measures of clinical significance have evolved and include many refinements developed and documented by the authors in this series. As measures have evolved, there has been an expansion of the meaning of the concept. Of course, there is no fixed or single

meaning of clinical significance. For example, within this series, concepts encompassed by clinical significance include functioning within the normative range after treatment, the persuasiveness of treatment effects, quality of life, and social importance of the change, to mention a few.

Clinical significance has been intended to reflect the extent to which treatment makes a difference, one of practical or applied value in everyday life. A number of measures of clinical significance are available, but researchers do not have a clear idea of the meaning of results that are clinically significant (i.e., beyond meeting the criteria for the operational definition). Also, it is still quite possible that multiple clients meet the operational definitions of clinically significant change but, in fact, are not functioning much better, do not feel better, or are not seen as improved by significant others.

Some of the early studies (e.g., during 1970s) using normative data to evaluate treatment outcome directly assessed how individuals were functioning at home, at school, and in the community (see Kazdin, 1977, for a review). When clients returned to within normative ranges at the end of treatment, the effects were persuasive because functioning reflected performance in everyday settings. Direct assessment in daily life does not resolve all ambiguities because the measure may not relate to functioning outside of the assessment context (e.g., situation, setting, or conditions of assessment). Yet such observations appeared closer to the definition of clinical significance as a change that has an impact on daily functioning. Any other type of measure can be just as or indeed more useful, valid, and meaningful, as long as evidence is provided showing that the measure or cutoff used to define clinical significance relates to other criteria that more directly reflect real-world impact.

There are enormous research opportunities for the assessment and evaluation of clinical significance. It would be valuable to begin with a definition and conceptual view of clinical significance and to derive and validate measures or indexes in keeping with that conceptualization. What is, in fact, a clinically significant change as defined by clients or consumers of treatment (e.g., through focus group or qualitative research)? A conceptual view is needed to organize the many facets (e.g., subjective evaluation and reports of others) and to permit tests that go beyond merely correlating measures with each other. Perhaps research ought to begin with the idea that clinical significance is multidimensional. The dimensions and their interrelations can be identified. The purpose in identifying multiple dimensions would be to evaluate clinical significance somewhat differently from current practices. Once key dimensions were identified and operationalized, researchers might use the data to obtain a profile of individual functioning (i.e., the client's score on each dimension that defines clinical significance). Alternatively, it may be that the goals of treatment will prompt the relevant dimensions or prioritize the dimensions along which clinically significant change ought to be measured. Clearly, important questions wait to be addressed in relation to the meaning and measurement of clinical significance.

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Bird, H. R., Yager, T. J., Staghezza, B., Gould, M. S., Canino, G., & Rubio-Stipec, M. (1990). Impairment in the epidemiological measure-

- ment of psychopathology in the community. *Journal of the American Academy of Child and Adolescent Psychiatry*, 29, 796–803.
- Burke, K. C., Burke, J. D., Regier, D. A., & Rae, D. S. (1990). Age at onset of selected mental disorders in five community populations. *Archives of General Psychiatry*, 47, 511–518.
- Foster, S. L., & Mash, E. J. (1999). Assessing social validity in clinical treatment research: Issues and procedures. *Journal of Consulting and Clinical Psychology*, 67, 308–319.
- Gladis, M. M., Gosch, E. A., Dishuk, N. M., & Crits-Cristoph, P. (1999). Quality of life: Expanding the scope of clinical significance. *Journal of Consulting and Clinical Psychology*, 67, 320–331.
- Gotlib, I. H., Lewinsohn, P. M., & Seeley, J. R. (1995). Symptoms versus a diagnosis of depression: Differences in psychosocial functioning. *Journal of Consulting and Clinical Psychology*, 63, 90–100.
- Howard, K. I., Lueger, R. J., & Kolden, G. G. (1997). Measuring progress and outcome in the treatment of affective disorders. In H. H. Strupp, L. M. Horowitz, & M. J. Lambert (Eds.), *Measuring patient changes in mood, anxiety, and personality disorders: Toward a core battery* (pp. 263–282). Washington, DC: American Psychological Association.
- Jacobson, N. S., & Revenstorf, D. (1988). Statistics for assessing the clinical significance of psychotherapy techniques: Issues, problems, and new developments. *Behavioral Assessment*, 10, 133–145.
- Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, 67, 300–307.
- Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification*, 1, 427–452.
- Kazdin, A. E. (1998). *Research design in clinical psychology* (3rd ed.). Needham Heights, MA: Allyn & Bacon.
- Kazdin, A. E. (in press). *Psychotherapy for children and adolescents: Directions for research and practice*. New York: Oxford University Press.
- Kazdin, A. E., & Wassell, G. (1998). Treatment completion and therapeutic change among children referred for outpatient therapy. *Professional Psychology: Research and Practice*, 29, 332–340.
- Kazdin, A. E., & Wilson, G. T. (1978). Criteria for evaluating psychotherapy. *Archives of General Psychiatry*, 35, 407–416.
- Kendall, P. C., & Grove, W. M. (1988). Normative comparisons in therapy outcome. *Behavioral Assessment*, 10, 147–158.
- Kendall, P. C., Marrs-Garcia, A., Nath, S. R., & Sheldrick, R. C. (1999). Normative comparisons for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology*, 67, 285–299.
- Kraemer, H. C. (1992). *Evaluating medical tests: Objective and quantitative guidelines*. Newbury Park, CA: Sage.
- Lewinsohn, P. M., Roberts, R. E., Seeley, J. R., Rohde, P., Gotlib, I. H., & Hops, H. (1994). Adolescent psychopathology: II. Psychosocial risk factors for depression. *Journal of Abnormal Psychology*, 103, 302–315.
- Lewinsohn, P. M., Seeley, J. R., Hibbard, J., Rohde, P., & Sack, W. H. (1996). Cross-sectional and prospective relationships between physical morbidities and depression in older adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35, 1120–1129.
- Lunnen, K. M., & Ogles, B. M. (1998). A multiperspective, multivariable evaluation of reliable change. *Journal of Consulting and Clinical Psychology*, 66, 400–410.
- Marks, I., Shaw, S., & Parkin, R. (1998). Computer-aided treatments of mental health problems. *Clinical Psychology: Science and Practice*, 5, 151–170.
- Newman, M. G., Consoli, A., & Taylor, C. B. (1997). Computers in assessment and cognitive behavioral treatment of clinical disorders: Anxiety as a case in point. *Behavior Therapy*, 28, 211–235.
- Pekarik, G., & Wolff, C. B. (1996). Relationship of satisfaction to symptom change, follow-up adjustment, and clinical significance. *Professional Psychology: Research and Practice*, 27, 202–208.
- Robinson, L. A., Berman, J. S., & Neimeyer, R. A. (1990). Psychotherapy for the treatment of depression: A comprehensive review of controlled outcome research. *Psychological Bulletin*, 108, 30–49.
- Sanford, M. N., Offord, D. R., Boyle, M. H., Peace, A., & Racine, Y. A. (1992). Ontario Child Health Study: Social and school impairments in children aged 6–16 years. *Journal of the American Academy of Child and Adolescent Psychiatry*, 31, 60–67.
- Strupp, H. H., & Hadley, S. W. (1977). A tripartite model of mental health and therapeutic outcomes. *American Psychologist*, 32, 187–196.
- Tingey, R. C., Lambert, M. J., Burlingame, G. M., & Hansen, N. B. (1996). Assessing clinical significance: Proposed extensions to method. *Psychotherapy Research*, 6, 109–123.
- United States Congress, Office of Technology Assessment. (1991). *Adolescent health* (Publication No. OTA-H-468). Washington, DC: U.S. Government Printing Office.
- Wolf, M. M. (1978). Social validity: The case of subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11, 203–214.
- Yates, B. T. (1995). Cost-effectiveness analysis, cost-benefit analysis, and beyond: Evolving models for the scientist–manager–practitioner. *Clinical Psychology: Science and Practice*, 2, 385–398.

Received October 5, 1998

Revision received November 16, 1998

Accepted November 16, 1998 ■