
Negative Effects From Psychological Treatments

A Perspective

David H. Barlow
Boston University

The author offers a 40-year perspective on the observation and study of negative effects from psychotherapy or psychological treatments. This perspective is placed in the context of the enormous progress in refining methodologies for psychotherapy research over that period of time, resulting in the clear demonstration of positive effects from psychological treatments for many disorders and problems. The study of negative effects—whether due to techniques, client variables, therapist variables, or some combination of these—has not been accorded the same degree of attention. Indeed, methodologies suitable for ascertaining positive effects often obscure negative effects in the absence of specific strategies for explicating these outcomes. Greater emphasis on more individual idiographic approaches to studying the effects of psychological interventions would seem necessary if psychologists are to avoid harming their patients and if they are to better understand the causes of negative or iatrogenic effects from their treatment efforts. This would be best carried out in the context of a strong collaboration among frontline clinicians and clinical scientists.

Keywords: psychotherapy, psychotherapy research, negative effects from therapy, idiographic research

I took my first psychotherapy course in 1965. Although the deepening shadows of time have obliterated most of the content of those early lectures, one admonition remains crystal clear. The instructor, with impeccable academic credentials and extensive experience in psychotherapy, announced that we would begin our course of study with what was then called client-centered therapy. The reason? With this approach, there would be less chance that we would actually harm our clients as we began the process of becoming psychotherapists.¹ Another mentor in that era, a psychologist in a respected child clinical center, recounted an anecdote of riding the elevator with a child from the reception area to a treatment room. On the way, the elevator stopped at an intermediate floor where he was joined by the parents of the child and their therapist. All said “hello.” After the session, the psychologist was castigated by the supervising psychiatrist for not timing his ride better and for the “irreparable harm” caused to therapeutic relationships by the blurring of professional roles when the family and the child inadvertently viewed each other with their therapists. Influential books during this period also underscored the grave harm that could occur during therapy

(e.g., Stuart, 1970). Being awakened to the possibility that one could inflict dire harm on patients during each visit to the consulting room (or even on the way to it) was an ever present source of anxiety during those early years for many of us. However, this anxiety sparked interest in the variety of ways that both benefit and harm during therapy might occur.

To take one example, one of the clear proscriptions communicated to all therapists in those early years was to avoid provoking anything more than mild anxiety in patients, and the advice had firm theoretical grounding at the time in both psychoanalytic and behavioral theorizing. From the psychoanalytic point of view, the dangers of experiencing intense conflict-driven emotion and the role of defense mechanisms in preventing this experience were already widely accepted (Fenichel, 1945). From the behavioral point of view, I had the good fortune in 1966 to study with Joseph Wolpe, who developed systematic desensitization to treat anxiety and fear. However, systematic desensitization was designed to work very gradually up a hierarchy of anxiety and fear on the premise that individuals with fears, phobias, and anxiety could tolerate only incremental increases in these emotions. According to Wolpe (1958), more intense experiences might result in the Pavlovian construct of *transmarginal inhibition*, or a state of complete shutdown of the organism. Similar but less dramatic concerns focused on further sensitizing the individual through excessive stimulation (Groves & Thompson, 1970). Thus, from a theoretical point of view, psychoanalytic and behavioral approaches concurred on the dangers of experiencing intense emotion without providing any guidelines on how much emotion was too much. As a consequence, therapists were very cautious indeed. It wasn't until the late 1960s or early 1970s that experimentation with more intensive therapist-guided in vivo expo-

My gratitude goes to Allen Bergin for his comments on an earlier version of this article, to Anke Ehlers for assembling and reanalyzing some data, and to Ben Emmert-Aronson for his research and organizational efforts.

Correspondence concerning this article should be addressed to David H. Barlow, Center for Anxiety and Related Disorders, Boston University, 648 Beacon Street, 6th Floor, Boston, MA 02215. E-mail: dhbarlow@bu.edu

¹ Ironically, Bergin (1963) provided some data indicating that client-centered therapy was the one therapeutic approach with evidence to suggest that it might cause deterioration in some clients.



David H. Barlow

sure-based procedures in patients with severe phobias began to demonstrate that these assumptions were incorrect (Agras, Leitenberg, & Barlow, 1968; Marks, Boulougouris, & Marset, 1971).

Despite the notion that potential danger or harm was a possible outcome of each and every session of psychotherapy, and the certainty with which this was conveyed in supervision, results from some of the first research studies of psychotherapy at that time seemingly revealed quite the opposite. That is, therapy had relatively little effect, either positive or negative, when results from treatment groups and comparison groups not receiving therapy were examined (Bergin, 1966; Bergin & Strupp, 1972). Classic early studies, such as the Cambridge–Somerville Youth Study (Powers & Witmer, 1951), which took decades to complete, arrived at this finding, as did other early efforts involving large numbers of patients treated in approximations of randomized controlled clinical trials (e.g., Barron & Leary 1955). More process-based research conducted on large numbers of outpatients to the point where outcomes were examined came to similar conclusions. In this era, Eysenck (1952, 1965) published his famously controversial thesis based on data from crude actuarial tables, which proposed that outcomes from psychotherapy across a heterogeneous group of patients were no better than rates of spontaneous improvement without psychotherapeutic intervention over varying periods of time. Although this conclusion was outrageous to many and served to fly in the face of clinical experience, it had enormous impact because it was difficult to rebut with the dearth of evidence available.

Thus, psychologists were faced in those early years with a paradox. On the one hand, potential sources of harm in therapy were highlighted, and supervisory sessions fo-

cused as much on these threats as on other more positive process issues associated with the potential for change. On the other hand, psychotherapy research of the day, such as it was, could not substantiate either these fears or the assumption that psychotherapy had any effect whatsoever.

Bergin's Deterioration Effect

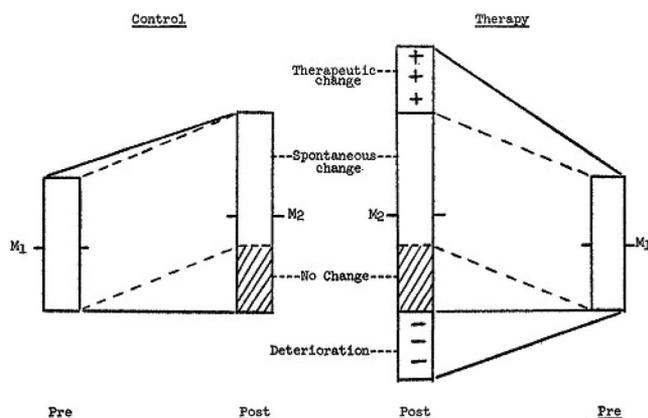
This state of affairs began to change in 1966 with the publication of Allen Bergin's seminal article, "Some Implications of Psychotherapy Research for Therapeutic Practice," in the *Journal of Abnormal Psychology*. What Bergin concluded, on the basis of a further analysis of some preliminary data first published in Bergin (1963), was that "psychotherapy may cause people to become better or worse adjusted than comparable people who do not receive such treatment" (Bergin, 1966, p. 235).

In fact, Bergin (1966) carefully reviewed seven studies in which no differences were apparent between treated and untreated groups but where a closer examination of the data revealed a much wider dispersion in change scores in the treatment groups compared with the comparison groups. Confirming to some degree Eysenck's (1952, 1965) notorious conclusions, Bergin noted, "Typically, control subjects improve somewhat with the varying amounts of change clustering around the mean. On the other hand, experimental subjects are typically dispersed all the way from marked improvement to marked deterioration" (p. 235). He further observed that because the length of therapy in these studies lasted from several months to several years, it was unlikely that any deterioration observed could have been due to temporary regression that occurred on occasion during treatment. This led Bergin to propose a schematic of the *deterioration effect* as presented in Figure 1. One can see in this figure similar average change from pretreatment to posttreatment but more people showing either greater improvement or greater deterioration with therapy compared to the control group. Again, his finding was that some people in therapy did improve substantially, but this was counterbalanced to some extent by those who showed more substantial deterioration. The fact that these data indicated psychotherapy can make some people considerably better off than comparable untreated patients was the first objective evidence against Eysenck's assertion that all changes associated with psychotherapy were due to spontaneous remission. As Bergin noted, "Consistently replicated, this is a direct and unambiguous refutation of the oft cited Eysenckian position" (p. 237).

From a historical perspective, this was a very important conclusion from the point of view of both science and policy. However, the emphasis on the conclusion that some people improved obfuscated the equally interesting finding that some people also deteriorated. Unfortunately, there was no way to go back and ascertain just who deteriorated and why. Of course, specific statistical tests of the significance of differential variance dispersion were not available, but it is noteworthy that this result was found in seven unrelated studies.

Bergin continued to explore this finding over the ensuing years, updating these results periodically in the iconic

Figure 1
The Deterioration Effect: Schematic Representation of Pre- and Posttest Distributions of Criterion Scores in Psychotherapy-Outcome Studies



Note. Plus signs indicate greater improvement, whereas minus signs indicate greater deterioration. M_1 = pretest mean criterion score; M_2 = posttest mean criterion score. From "Some Implications of Psychotherapy for Therapeutic Practice," by A. Bergin, 1966, *Journal of Abnormal Psychology*, 71, p. 238. Copyright 1966 by the American Psychological Association.

Handbook of Psychotherapy of Behavioral Change (e.g., Bergin & Garfield, 1971; Garfield & Bergin, 1978). In 1971, Bergin reported 23 additional studies, for a total of 30, showing that in a proportion of patients, some deterioration occurred that exceeded results from comparable control groups, where a bit of deterioration also occurred. After reporting these results, Bergin (1971) observed,

In recent years I have received numerous communications from both therapists and patients who have provided rich detail regarding the process of therapist-caused deterioration. I have found some of these examples most disturbing, perhaps because I have been too naïve regarding the way life really is. Apparently there are many areas of error and malpractice that are regularly covered up by practitioners in every field. It seems to be an all too common procedure to ignore these incidents, no matter how serious the consequences may be for the patients involved. Indeed, I hope that one of our suicide centers might do a careful study of the possibility of therapist-precipitated suicides. In general, deterioration of various kinds is much too common to be ignored. (p. 250)

There is little reason to believe that this state of affairs has changed over the decades.

Advances in Psychotherapy Research

Bergin (1966), in the process of articulating his influential argument, also noted the substantial deficits in extant studies of psychotherapy at that time and, in so doing, began to pave the way for the marked improvement in psychotherapy research methods to unfold in the coming decades. He

noted, for example, that experimental and control groups were often not well matched, with differences in initial severity on various measures being a common finding. He also noted that individuals assigned to control groups were often subject to substantial nonexperimental influences, including therapeutic interventions of various sorts occurring outside the context of the clinical trial. He suggested the need to carefully ascertain whether these groups were indeed acting as controls and/or to directly measure the effects of nonexperimental influences that might affect outcomes. He also presented some preliminary data showing that training was an important variable if therapists were indeed to deliver the treatment as intended, contributing to what is now referred to as *treatment integrity* of the intervention under study (Hayes, Barlow, & Nelson-Gray, 1999). This issue arose in some earlier studies where therapists had little or no training, and it was unclear just what they were doing (e.g., Powers & Witmer 1951).

In addition to these critiques of existing studies, Bergin and Strupp (1972) made proactive recommendations for the future conduct of psychotherapy research, recommendations that were to have substantial impact. One of the observations focused on the marked individual differences among patients in these studies, particularly patients with emotional or behavioral disorders. They suggested that attempts to apply broad-based and ill-defined treatments such as *psychotherapy* to a heterogeneous group of clients only vaguely described with labels such as *neurosis* would be hard pressed to answer basic questions on the effectiveness (or ineffectiveness) of a specific treatment for a specific individual. This heterogeneous approach also characterized early meta-analyses in that era (Smith & Glass, 1977). Thus, Bergin and Strupp's review suggested that asking "Is psychotherapy effective?" was probably the wrong question. Bergin and Strupp cited Gordon Paul (1967), who suggested that psychotherapy researchers must start defining their interventions more precisely and must ask, "What specific treatment is effective with a specific type of client under what circumstances?" (p. 112).

In addition, Bergin and Strupp (1972) suggested that a more valid tool for looking at the effects of psychotherapy and delineating possible harmful outcomes would involve a more intensive study of the individual. "Among researchers as well as statisticians there is a growing disaffection from traditional experimental designs and statistical procedures which are held inappropriate to the subject matter under study" (Bergin & Strupp, 1972, p. 440). In fact, they recommended the individual experimental case study as one of the primary strategies that would move the field of psychotherapy research forward because changes of clinical significance could be directly observed in the individual under study (followed by replication on additional individuals). In such a way, changes could be clearly and functionally related to specific therapeutic procedures. These ideas contributed to the development of single case experimental designs for studying behavior change (Barlow, Nock, & Hersen, 2009; Hersen & Barlow, 1976). These designs, then and now, play an important role not only in delineating the positive effects of therapy, but also in

observing more readily any deleterious effects that may emerge, thus complementing efforts to extract information on individuals from the response of a group in a clinical trial (Kazdin, 2003).

This emphasis on individual change of clinical and practical importance, as well as the possibility of deterioration in some individuals, also contributed to a revision of the ways in which data from large between-groups experimental designs (clinical trials) were analyzed (Kazdin, 2003). Specifically, over the ensuing decades, psychotherapy researchers began to move away from exclusive reliance on the overall average group response on measures of change and began highlighting the extent of change (effect sizes and confidence intervals), whether the change was clinically significant, and the number or percentage of individuals who actually achieved some kind of satisfactory response (with a passing nod to those who did not do well; Jacobson & Truax, 1991). Data analytic techniques also became more sophisticated, powerful, and valid, with a move away from comparison of means among groups to multivariate random effects procedures, such as latent growth curve and multilevel modeling, which evaluate the extent, patterns, and predictors of individual differences in change (e.g., Brown, 2007).

Another important development was a much greater delineation and definition of the actual psychotherapeutic procedures undergoing evaluation. The shortcomings of early studies in this regard are best exemplified in the classic Cambridge–Somerville Youth Study, where the independent variable was defined as instructing 10 therapists with no formal training to do whatever they thought best over a minimum of five sessions per year for up to five years with predelinquent boys. Equally important was a greater specification of the psychopathological processes most often targeted for change. Over the ensuing decades, the very nature of psychopathology in its various manifestations became increasingly well understood and defined through research in this area. This led to the appearance of nosological conventions through which psychotherapy researchers could begin to reliably agree on what was being treated and how to measure change (Barlow, 1991). Investigators increasingly made use of this information to assess both the process and outcomes of interventions (e.g., Elkin et al., 1989). Thus, by the 1980s, the field was specifying and operationalizing psychotherapeutic procedures as well as associated therapist, client, and relationship factors. Researchers were specifying and measuring the targets of treatments in the form of identifiable psychopathology and were doing so in a way that allowed individual differences in response to be highlighted. By the 1990s, publications of large clinical trials, some begun 10 years prior to publication, rapidly grew in number.

It seemed at the time (to me at least) that the stage was set for an informed and intensive study of not only positive effects but also negative effects, in other words, a thorough-going analysis at a more individual level of who might experience adverse effects for one reason or another and why. In reality, these trials had enormous impact because they established causal relationships for the effects

of specific psychological procedures and interventions, and during this era, the efficacy of psychotherapy and psychological treatments was firmly established (Kazdin & Weiss, 2003; Nathan & Gorman, 1998, 2007; Roth & Fonagy, 1996, 2004). However, results from these trials continued to draw mostly nomothetic conclusions about average responding in a treated group, the percentage meeting a clinically significant threshold of response, or similar results that obscured whether harmful effects occurred in some individuals because of treatment or for other reasons.

Clinical Practice Guidelines and Negative Effects

In the meantime, the greatly increased interest over the ensuing decades of health care policymakers in health care interventions, including psychological treatments (Barlow, 2004), because of increasing amounts of public monies allocated to pay for health care, was not without consequences. Various organizations, including government agencies that were paying for these services, began to suggest standards of care based, sometimes loosely, on extant research evidence in order to improve the overall quality of health care practices. These standards were variously called *best practices*, *best care algorithms*, or *clinical practice guidelines*. Some of these early guidelines, typically those emanating from some managed care organizations, were woefully lacking in even the rudiments of evaluating and articulating the evidence base (Barlow, 2004). To address this issue, the American Psychological Association (APA) in 1995 developed a template, entitled *Template for Developing Guidelines: Interventions for Mental Disorders and Psychosocial Aspects of Physical Disorders*, to guide the optimal construction of clinical practice guidelines. This template, which was updated in 2002 (APA, 2002), detailed a hierarchy of evidence based on existing methodologies. The APA undertook this task in view of the methodological expertise available within the association, but also because of its experience, decades previously, with a similar issue. At that time, psychological tests were proliferating without a template or set of scientific criteria on which to base the development of those tests. In response, the APA in 1966, in cooperation with the American Educational Research Association and the National Council on Measurement in Education, developed the first version of the well-known *Standards for Educational and Psychological Tests and Manuals* (APA, 1966). The purpose was to delineate the scientific criteria of reliability and validity that any psychological test would have to meet to be credible and useful. This document, revised several times since (American Educational Research Association, APA, & National Council on Measurement in Education, 1999), has become the standard in the field and is widely used by professionals, policymakers, and courts to determine the adequacy of various psychological tests (Hayes et al., 1999).

Although most of the methodologies in the hierarchy of evidence specified in the guidelines template focused on evaluating the positive effects of certain therapies, there

was also a brief allusion to potentially harmful effects. Specifically, Criterion 5.0 notes that clinical practice guidelines “should specify the outcomes the intervention is intended to produce and the evidence should be provided for each outcome” (APA, 2002, p. 1055). Point 8 under this criterion is headed “Iatrogenic negative effects or side effects of treatments” and states, “Thorough outcome evaluation not only considers potential benefits but also examines possible side effects or negative outcomes associated with treatment” (p. 1055). Furthermore, when considering feasibility of treatments, Criterion 11.0 specifies, “Guidelines should explicitly note and evaluate possible adverse effects of interventions as well as their benefits” (p. 1057). The difficulty was that these recommendations did not clearly conceptualize what kind of evidence might be sufficient to demonstrate these negative effects. Should evidence of this kind depend on randomized controlled trials, one or more unfortunate case reports, or other methods? And what characterizes a negative outcome? Need it be deterioration during or after treatment or, perhaps, improving less than individuals in an untreated comparison group? Is the crucial period of observation during active treatment, or is it also some (specific) period after conclusion of treatment? The answers are likely to be different for different disorders or problem areas.

For example, major depressive episodes are characterized by a highly fluctuating course (Judd, 1997). Therefore, a convention among experts working in the area of mood disorders is to discuss outcomes of treatment in the context of the 5 Rs (Hollon, Thase, & Markowitz, 2002). Thus, a *response* refers to a significant reduction in symptom severity (typically 50%), but *remission* is a more complete response reflecting a return to normal. Full *recovery*, on the other hand, is assumed not to occur until a significant period of time after remission, typically at least six months. A return of depressive symptoms prior to that time is considered a *relapse* of the original episode, but after full recovery is reached, a return of depression would be considered a *recurrence*, or new episode.

When ascertaining negative effects from treatment, one might look for slower response, less remission or recovery, higher rates of relapse or recurrence, or some combination of these. However, the best reference condition would be an untreated group because major depressive episodes ultimately run their course and remit on their own, only to recur later. Obviously, this specific comparison would no longer be possible ethically, because therapists have effective evidenced-based treatments for depression that should be implemented without delay. This begs the question of where researchers are likely to find such evidence for negative effects.

Sources of Evidence on Negative Effects

Lilienfeld (2007), in an influential article that reignited interest in negative effects from psychological treatments, suggested two main sources of evidence that he judged to be relevant. The first is the occasional randomized con-

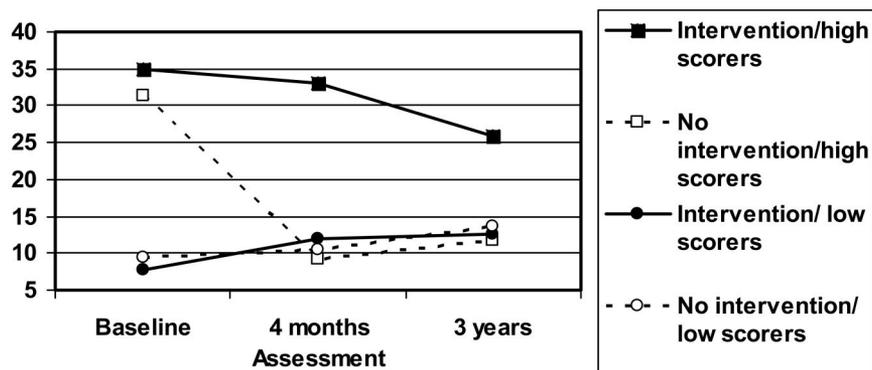
trolled clinical trial that (inadvertently) demonstrates the opposite of the hypothesized positive outcome. That is, a psychological treatment under investigation actually makes individuals worse at some point following the intervention as compared to a control group. He noted that some statisticians have referred to findings that run counter to the hypothesized effect as *Type III errors* (Leventhal & Huynh, 1996). One example of a finding in this category is the effect of critical incident stress debriefing (CISD) for individuals who have recently experienced trauma. Most studies and meta-analyses have found no overall effect from this intervention when compared to the absence of intervention (Litz, Gray, Bryant, & Adler, 2002), but several randomized controlled trials have found significantly higher symptomatology at some later times among treated groups compared to matched untreated control groups (Mayou, Ehlers, & Hobbs, 2000; McNally, Bryant, & Ehlers, 2003). A similar pattern has been reported for certain specific group therapy programs for conduct disordered and substance-abusing adolescents (Rhule, 2005).

Looking more closely at Mayou et al.'s (2000) widely noted study on CISD for victims of motor vehicle accidents, it is interesting for the purposes of this article to examine the reasoning that led to this important analysis (A. Ehlers, personal communication, December 8, 2008). After finding no overall effect for CISD, the investigators thought that the reason for null effects might have been that the analysis lumped together two groups of participants: those who seemingly needed intervention because they had high scores on the Impact of Events Scale (Horowitz, Wilner, & Alvarez, 1979) following the accident and were at risk for posttraumatic stress disorder and those with low scores on this scale who might not need intervention. No differences as a function of whether they were treated were evident for individuals with low scores, but those with high scores were considerably worse off after four months as well as at a three-year follow-up after receiving CISD (see Figure 2). Lumping outcomes from these victims together obscured this important effect. Of course, randomized controlled trials designed to show a positive effect that actually end up reporting a negative effect, as opposed to simply no effect, are likely to be few and far between. One reason is that results such as this would be inadvertent and unexpected. In Mayou et al.'s study, the investigators quite deliberately adopted a more idiographic approach by analyzing, on the basis of a priori empirical findings, those most likely to be impacted by the intervention one way or the other. Also, given the paucity of data on effective treatments in this area and the fact that many individuals receive no intervention, an untreated comparison group was readily available to make this analysis possible.

More fine-grained analysis in the form of dismantling studies of multicomponent psychological treatments have also yielded important conclusions on potentially negative effects of some procedures. For example, the discovery that breathing retraining and applied relaxation may actually detract from the effects of exposure-based procedures for individuals with panic disorder with agoraphobia, along with a new focus in treatment on the importance of aware-

Figure 2

Impact of Event Scores for High and Low Scores in Intervention and No-Intervention Groups at Baseline Assessment and Four-Month and Three-Year Follow-Ups



Note. From "Psychological Debriefing for Road Traffic Accident Victims: Three-Year Follow-Up of a Randomised Controlled Trial," by R. A. Mayou, A. Ehlers, and M. Hobbs, 2000, *British Journal of Psychiatry*, 176, p. 592. Copyright 2000 by the Royal College of Psychiatrists.

ness and acceptance of negative emotions, has led to a de-emphasis or elimination of breathing retraining and relaxation as coping procedures during exposure exercises (as opposed to utilizing these procedures outside of exposure exercises to reduce tension etc.; Craske & Barlow, 2008). In this example, a negative effect was noted by examining outcomes with and without breathing retraining or relaxation.

A second type of evidence would be case study reports of the occasional dramatic negative effects immediately following an intervention. Perhaps the best known example would be the several deaths resulting from rebirthing techniques in oppositional children (Mercer, Sarnier, & Rosa 2003). However, traumatic negative effects, such as deaths or severe physical injuries resulting directly from psychological treatments, are also likely to be rare occurrences. When these effects do occur, the consequences must be clearly and unambiguously connected to the intervention as was the situation in the case studies of rebirthing techniques in which children were inadvertently smothered as part of the procedure. Evidence less concrete than this in only one or two cases would not rise above the level of clinical anecdote.

Although Lilienfeld's (2007) initial attempt at a categorization of *potentially harmful therapies* on the basis of these two types of evidence has heuristic value, it was clearly meant as little more than a rough approximation of how we can usefully approach the issue of ascertaining the nature and causes of negative effects from psychological treatments. The more noteworthy objective was to spark renewed interest in and deeper consideration of this topic, and his effort has obviously been successful in this regard.

The Need for an Idiographic and Nomothetic Balance in Ascertaining Effects of Treatments

In summary, psychologists are in an age when health care policymakers and the public at large have accepted that psychological interventions can have beneficial effects and should be included in the health care system (Barlow, 2004; Nathan & Gorman, 2007). We also have the beginnings of some evidence that some interventions might be harmful, as indicated by deterioration in functioning and/or the dashed hopes and expectations that often come with failed efforts. Most therapists would agree that it is crucial to be concerned about and sensitive to harmful effects, however small. If we are to move forward for the greater public good and our own edification, however, it is important to delineate the variety of ways in which we can ascertain the totality of both positive and negative effects of our interventions and substantially increase our sensitivity to these effects. Perhaps it is time to re-examine Bergin and Strupp's (1972) admonitions from over 30 years ago and attend to the responding of each and every individual to avoid burying potentially important negative effects in the group average of clinical trials, whether those negative effects are due to unrelated life events, untoward therapeutic influences, or the direct effect of a given psychological treatment interacting with individual client variables. To do this, in my opinion, will require emphasis on a more idiographic approach in methods and data analysis and a close collaboration among practitioners and clinical scientists (Barlow & Nock, 2009).

In this respect, the major differences between the idiographic and nomothetic traditions are in approaches to

intersubject variability and the generality of findings. Because variability is often considerable among clients responding to treatment, the task of any psychologist is to discover functional relations between treatment and outcome over and above the welter of environmental and biological variables influencing the client at any given time. A nomothetic approach makes an implicit assumption that much of this variability, including occasional deterioration, is intrinsic to the client or due to uncontrollable external events, and this approach uses sophisticated data analytic procedures to look for reliable effects over and above this variability. Significant effects are then assumed to be more or less generalizable on the basis of the number of individuals included in the experimental group and the representativeness of the population of such individuals (i.e., the use of random sampling). Clearly, a renewed emphasis on identifying crucial individual differences among individual clients on the basis of good empirical or theoretical reasons and conducting sensitive moderator or mediator analyses is increasingly important as exemplified in Mayou et al.'s (2000) study. However, it is also very important that these analyses do not stray far from the individual data so that clinicians can generalize from these data sets to the individuals that they serve. As Sidman (1960) pointed out a number of years ago in discussing approaches to variability and generality of findings,

Tracking down sources of variability is then a primary technique for establishing generality. Generality and variability are basically antithetical concepts. If there are major undiscovered sources of variability in a given set of data, any attempt to achieve subject or principle generality is likely to fail. Every time we discover and achieve control of a factor that contributes to variability, we increase the likelihood that our data will be reproducible with new subjects and different situations. Experience has taught us that precision of control leads to more extensive generalization of data. (p. 152)

Single-case experimental designs are particularly well suited to identifying intersubject and intrasubject variability and immediately tracking down sources of this variability in a fast-acting and flexible manner if the variability involves some deterioration in an individual client. In this sense, these methods, although capable of establishing cause-effect relationships, are closer to usual and customary procedures in the clinic (Barlow et al., 2009; Kazdin & Nock, 2003). To take one straightforward example, one individual suffering from major depression rather quickly reaches remission during treatments, but the second individual does not. The clinician, quite naturally, will hypothesize why and then adapt the treatment accordingly. These adaptations might involve the introduction of new metaphors to promote greater understanding of the nature of depression and the process of treatment, or perhaps a creative new procedure to promote more substantial behavioral activation (Dimidjian, Martell, Addis, & Herman-Dunn, 2008) by incorporating this procedure more individually into the client's routine. Assuming measures of progress are administered periodically, clinicians could use a variety of clinician-friendly procedures to systematically

evaluate these innovations in subsequent clients (Barlow et al., 2009).

The more important development is that therapists do not have to wait for the next clinical trial to look for deterioration in individuals. On the basis of the pioneering work of Lambert et al. (2003) among others, and taking advantage of policy changes requiring outcomes measures to track progress during the course of interventions, an enormous database will soon be available from clinicians to examine deterioration effects when they occur. In the context of practice research networks (Borkovec, Echemendia, Ragusea, & Ruiz, 2001), clinicians would rapidly become aware of lack of progress or even deterioration and could attempt to remediate this effect, a strategy that Lambert et al. have shown can be successful. More important, clinicians acting as local clinical scientists (Stricker & Trierweiler, 1995) could hypothesize potential mediators and moderators of these unsuccessful or negative outcomes. This information could then be either evaluated directly by clinicians using single-case experimental evaluative procedures, as noted earlier, or fed back to clinical research centers where these idiographic data could be analyzed more closely and hypotheses could be prospectively tested.

Conclusion

With the rapid dissemination of empirically supported psychological treatments across health care systems internationally (McHugh & Barlow, in press), it is time to focus attention in a more systematic manner on those unfortunate cases where harm might occur or benefit is conspicuously absent. We need to refine our methods to accomplish this important task and develop a consensus on how to best define and explicate negative effects. Psychologists are in a unique position to implement strategies with a more idiographic focus across the spectrum of mental health care to achieve these goals.

REFERENCES

- Agras, S., Leitenberg, H., & Barlow, D. H. (1968). Social reinforcement in the modification of agoraphobia. *Archives of General Psychiatry, 19*, 423–427.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Psychological Association. (1966). *Standards for educational and psychological tests and manuals*. Washington, DC: Author.
- American Psychological Association. (1995). *Template for developing guidelines: Interventions for mental disorders and psychosocial aspects of physical disorders*. Washington, DC: Author.
- American Psychological Association. (2002). Criteria for evaluating treatment guidelines. *American Psychologist, 57*, 1052–1059.
- Barlow, D. H. (1991). Introduction to the special issue on diagnosis, dimensions, and *DSM-IV*: The science of classification. *Journal of Abnormal Psychology, 100*, 243–244.
- Barlow, D. H. (2004). Psychological treatments. *American Psychologist, 59*, 869–878.
- Barlow, D. H., & Nock, M. K. (2009). Why can't we be more idiographic in our research? *Perspectives on Psychological Science, 4*, 19–21.
- Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). Boston, MA: Allyn & Bacon.

- Barron, F., & Leary, T. (1955). Changes in psychoneurotic patients with and without psychotherapy. *Journal of Consulting Psychology, 19*, 239–245.
- Bergin, A. E. (1963). The effects of psychotherapy: Negative results revisited. *Journal of Counseling Psychology, 10*, 224–250.
- Bergin, A. E. (1966). Some implications of psychotherapy for therapeutic practice. *Journal of Abnormal Psychology, 71*, 235–246.
- Bergin, A. E. (1971). The evaluation of therapeutic outcomes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change: An empirical analysis* (pp. 217–270). New York, NY: Wiley.
- Bergin, A. E., & Garfield, S. L. (Eds.). (1971). *Handbook of psychotherapy and behavior change: An empirical analysis*. New York, NY: Wiley.
- Bergin, A. E., & Strupp, H. H. (1972). *Changing frontiers in the science of psychotherapy*. Chicago, IL: Atherton.
- Borkovec, T. D., Echemendia, R. J., Ragusea, S. A., & Ruiz, M. (2001). The Pennsylvania Practice Research Network and future possibilities for clinically meaningful and scientifically rigorous psychotherapy effectiveness research. *Clinical Psychology: Science and Practice, 8*, 155–167.
- Brown, T. A. (2007). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Craske, M. G., & Barlow, D. H. (2008). Panic disorder and agoraphobia. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders* (4th ed., pp. 1–64). New York: Guilford Press.
- Dimidjian, S., Martel, C. R., Addis, M. E., & Herman-Dunn, R. (2008). Behavioral activation in the treatment of major depressive disorder. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders* (4th ed., pp. 328–364). New York: Guilford Press.
- Elkin, I., Shea, T., Watkins, J., Imber, S., Sotsky, S. M., Collins, T. F., et al. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry, 46*, 971–982.
- Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology, 16*, 319–324.
- Eysenck, H. J. (1965). The effects of psychotherapy. *International Journal of Psychiatry, 1*, 97–178.
- Fenichel, O. (1945). *The psychoanalytic theory of neurosis*. New York, NY: Norton.
- Garfield, S. L., & Bergin, A. E. (Eds.). (1978). *Handbook of psychotherapy and behavior change: An empirical analysis* (2nd ed.). New York, NY: Wiley.
- Groves, P. M., & Thompson, R. F. (1970). Habituation: A dual-process theory. *Psychological Review, 77*, 419–450.
- Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. O. (1999). *The scientist practitioner: Research and accountability in the age of managed care* (2nd ed.). Needham Heights, MA: Allyn & Bacon.
- Hersen, M., & Barlow, D. H. (1976). *Single case experimental designs: Strategies for studying behavior change*. New York, NY: Pergamon Press.
- Hollon, S. D., Thase, M. E., & Markowitz, J. C. (2002). Treatment and prevention of depression. *Psychological Science in the Public Interest, 3*(2), 39–77.
- Horowitz, H., Wilner, N., & Alvarez, W. (1979). Impact of Events Scale. A measure of subjective stress. *Psychological Medicine, 41*, 209–212.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.
- Judd, L. L. (1997). The clinical course of unipolar major depressive disorders. *Archives of General Psychiatry, 54*, 989–991.
- Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston, MA: Allyn & Bacon.
- Kazdin, A. E., & Nock, M. (2003). Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, 44*, 1116–1129.
- Kazdin, A. E., & Weiss, J. R. (Eds.). (2003). *Evidence-based psychotherapies for children and adolescents*. New York, NY: Guilford Press.
- Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielson, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science & Practice, 10*, 288–301.
- Leventhal, L., & Huynh, C. (1996). Directional decisions for two-tailed tests: Power, error rates and sample size. *Psychological Methods, 1*, 278–292.
- Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2*, 53–70.
- Litz, B., Gray, M., Bryant, R., & Adler, A. (2002). Early intervention for trauma: Current status and future directions. *Clinical Psychology: Science and Practice, 9*, 112–134.
- Marks, I., Boulougouris, J., & Marset, P. (1971). Flooding versus desensitization in the treatment of phobic patients: A crossover study. *British Journal of Psychiatry, 119*, 353–375.
- Mayou, R. A., Ehlers, A., & Hobbs, M. (2000). Psychological debriefing for road traffic accident victims: Three-year follow-up of a randomized controlled trial. *British Journal of Psychiatry, 176*, 589–593.
- McHugh, R. K., & Barlow, D. H. (in press). Dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist*.
- McNally, R. J., Bryant, R. A., & Ehlers, A. (2003). Does early psychological intervention promote recovery from posttraumatic stress? *Psychological Science in the Public Interest, 4*(2), 45–79.
- Mercer, J., Sarnet, L., & Rosa, L. (2003). *Attachment therapy on trial*. Westford, CT: Praeger.
- Nathan, P. E., & Gorman, J. M. (1998). *A guide to treatments that work*. New York, NY: Oxford University Press.
- Nathan, P. E., & Gorman, J. M. (2007). *A guide to treatments that work* (3rd ed.). New York, NY: Oxford University Press.
- Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology, 3*, 109–118.
- Powers, E., & Witmer, H. (1951). *An experiment in the prevention of delinquency: The Cambridge–Somerville Youth Study*. New York, NY: Columbia University Press.
- Rhule, D. (2005). Take care to do no harm: Harmful interventions for youth problem behavior. *Professional Psychology: Research and Practice, 36*, 618–625.
- Roth, A., & Fonagy, P. (1996). *What works for whom?* New York, NY: Guilford Press.
- Roth, A., & Fonagy, P. (2004). *What works for whom?* (2nd ed.). New York, NY: Guilford Press.
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York, NY: Basic Books.
- Smith, M., & Glass, G. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*, 752–760.
- Stricker, G., & Trierweiler, S. J. (1995). The local clinical scientist: A bridge between science and practice. *American Psychologist, 50*, 995–1002.
- Stuart, R. B. (1970). *Trick or treatment: How and when psychotherapy fails*. Champaign, IL: Research Press.
- Wolpe, J. (1958). *Psychotherapy by reciprocal inhibition*. Stanford, CA: Stanford University Press.