
Evaluation of Psychotherapy

Efficacy, Effectiveness, and Patient Progress

Kenneth I. Howard
Karla Moras
Peter L. Brill
Zoran Martinovich
Wolfgang Lutz

Northwestern University
University of Pennsylvania
Compass Information Services, Inc.
Northwestern University
Forschungsstelle für Psychotherapie, Stuttgart

Treatment-focused research is concerned with the establishment of the comparative efficacy and effectiveness of clinical interventions, aggregated over groups of patients. The authors introduce and illustrate a new paradigm—patient-focused research—that is concerned with the monitoring of an individual's progress over the course of treatment and the feedback of this information to the practitioner, supervisor, or case manager.

There are three fundamental questions that can be asked about any treatment (intervention): (a) Does it work under special, experimental conditions? (b) does it work in practice? and (c) is it working for this patient? In current scientific terminology, Question a is a treatment efficacy question that requires the use of standard, experimental methodology and is answered in terms of the mean or average response of patients, and Question b is a treatment effectiveness question that requires evaluation of a treatment in naturalistic clinical settings and the use of quasi-experimental procedures (e.g., Campbell & Stanley, 1966). Question b is also answered in terms of mean group response. The crucial, immediate question posed by the practitioner, however, is Question c. This question is answered idiographically, in terms of a particular patient's response to the treatment being provided by a particular clinician. The first two questions are treatment focused in that they seek answers about the general impact of particular interventions, whereas the third question is patient focused in that it seeks answers about the characteristics of particular cases.

Answers¹ to all three questions, of course, depend on what it means for an intervention to *work*—in other words, how the desired outcome of treatment is defined and assessed. Outcome usually entails some valued and changed state (e.g., better concentration and, hence, return of one's work functioning to desired levels). However, what is valued always depends on the perspective of the evaluator. At least six parties can be identified who have a vested interest in the evaluation of mental health treatments: patients, clients, clinicians, managers, sponsors, and researchers (Krause & Howard, 1976; Strupp & Hadley, 1977). *Patients* are the persons who directly receive the treatment. *Clients* are the persons or institutions

whose interests are intended to be served by the treatment—although the patient also is usually the client, there are situations in which the client is a parent, work supervisor, or institution (e.g., school or court). *Clinicians* are the persons who conduct the treatment. *Managers* are the persons who make decisions regarding the allocation of treatment resources (e.g., number of treatment sessions). *Sponsors* are the persons or institutions who pay for the treatment. Finally, *researchers* are persons who are concerned with the application of proper scientific methodology (measurement technology and standards of evidence) for assessing treatment effects. Each of the foregoing interested parties brings its own values and standards of evidence to bear when evaluating mental health treatments. Thus, it is important to specify the evaluative perspective or audience when treatment outcome questions are being asked.

Clinical scientists tend to focus on the first question, the efficacy question. In practice, the question is stated in a somewhat more specific way when it is used to guide an efficacy study. For example, "Does this new intervention produce better outcomes than does an already commonly used intervention or than a putatively inert, 'placebo,' control intervention?" The standard method for addressing this efficacy question is the randomized clinical trial, a method that uses procedures designed to optimize the internal validity of a study—in other

Kenneth I. Howard, Department of Psychology, Northwestern University; Karla Moras, Department of Psychology, University of Pennsylvania; Peter L. Brill, Compass Information Services, Inc., Philadelphia, PA; Zoran Martinovich, Department of Psychology, Northwestern University; Wolfgang Lutz, Forschungsstelle für Psychotherapie, Stuttgart, Germany.

Correspondence concerning this article should be addressed to Kenneth I. Howard, Department of Psychology, College of Arts and Sciences, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208-2710. Electronic mail may be sent via Internet to k-howard@nwu.edu.

¹ Of course, any answer is dependent on a prospective design—in other words, the availability of measures of the patient's status before, during, and after treatment. Assessments of the effects of a treatment that are based on impressionistic, global, retrospective accounts of patients bear little relationship to assessments of change that are based on "before and after" measures.

words, to increase our confidence that the cause of any observed mean differences is due to treatment conditions (Campbell & Stanley, 1966; Cook & Campbell, 1979). For example, randomization procedures are used to help ensure the pretreatment comparability of the treatment and control groups. Also, the treatment conditions are clearly specified (e.g., in a specific treatment manual), and these specifications are used to train study therapists. Furthermore, patients are included according to highly specific inclusion–exclusion criteria, and experimental conditions are closely monitored (e.g., controlling for dosage or assessing the integrity of the interventions that are actually delivered).

The *sine qua non* of this kind of experiment is the random assignment of patients to comparison groups, a procedure that is intended to ensure comparability of groups (but rarely does in any specific experiment; see Howard, Krause, & Lyons, 1993) before exposure to the experimental conditions. Because this experimental control is quite expensive, pilot data are usually collected to buttress the plausibility of embarking on a full clinical trial. Consequently, it is quite rare that a randomized experiment fails to conclude that the experimental treatment works. In a sense, this is as it should be, because a plausible (rationalized) treatment should work under some conditions (e.g., very large *N*, modification of inclusion–exclusion criteria, or increase in dosage).

Of course, randomization only works in the long run, so replication of results is essential. Moreover, there are many problems with the realization of a randomized clinical trial (e.g., Howard, Krause, & Lyons, 1993; Howard, Krause, & Orlinsky, 1986). As Seligman (1995) reminded us, the constraints of such a trial, including the necessary random assignment of patients to conditions, severely limit external validity (i.e., generalizability of findings to other patients, therapists, and settings).

The second question, “Does the new treatment work in practice?” is asked by mental health service researchers. Service researchers address what is referred to as the *effectiveness question* (e.g., see Weisz, Donenberg, Han, & Weiss, 1995): “Does this treatment produce beneficial results as it is administered in actual clinical settings (i.e., clinics and offices)?”² The preferred method for addressing this question is the systematic, naturalistic experiment, a method that emphasizes external validity and attempts to ensure the generalizability of findings to other clinicians, clinical settings, and patient groups. Assignment to comparison groups (e.g., successful cases vs. failure cases, or patients who attend many sessions vs. patients who attend few sessions) is not random, and, thus, comparison groups may differ (pretreatment) on many variables in addition to the (independent) variable that has been selected for study.³ Consequently, because of threats to internal validity, any observed results are subject to multiple interpretations (plausible alternative explanations). Such quasi-experiments require constructive replication to test such competing hypotheses.

To date, by far the least systematic research exists for the third question, the one of most immediate, day-

to-day concern to clinicians: “Is this patient’s condition responding to the treatment that is being applied?” From the practitioner’s perspective, the most salient issue is that the patient has sought amelioration of some (appropriate) malady, and it is the practitioner’s job to provide a treatment that will provide this amelioration. In this context, it is not sufficient for the practitioner to know that a particular treatment can work (efficacy) or does work (effectiveness) on average—in other words, the kind of conclusion that can be drawn from the results of main effects analyses of efficacy and effectiveness studies. The practitioner needs to know what treatment is likely to work for a particular individual and then whether the selected treatment is working for this patient. Thus, from the clinician’s evaluative perspective, one critically important task of research is to provide valid methods for systematically evaluating a patient’s condition in terms of the ongoing response of that condition over the course of treatment. Moreover, the clinician is interested in relevant feedback about the patient’s condition (assessment of progress during the course of treatment, not the assessment of outcome after the termination of treatment). This is the focus of our current work and model construction. However, before describing this work, we want to introduce its conceptual and empirical foundations.

The Dosage and Phase Models of Psychotherapy

The dosage model of psychotherapeutic effectiveness (Howard, Kopta, Krause, & Orlinsky, 1986) demonstrated a lawful linear relationship between the log of the number of sessions and the normalized probability of patient improvement. This log–normal relationship is quite common in psychology and reflects that more and more efforts (e.g., sessions, trials, and milligrams of medication) are needed to produce incremental changes in the desired response (cf. the Weber-Fechner law of just noticeable differences). Subsequent research has provided evidence of the differential responsiveness to psychotherapy of various symptoms and syndromes (e.g., Horowitz, Rosenberg, Baer, Ureño, & Villasenor, 1988; Howard, Lueger, Maling, & Martinovich, 1993; Kadera, Lambert, & An-

² There is no logical connection between showing that a treatment can work and showing that a treatment does work. That is, a treatment that cannot be shown to produce statistically significant mean group differences in a carefully conducted clinical trial may still be demonstrably beneficial as actually practiced. Similarly, a treatment that has been shown to be effective in a clinical trial may not be effective as practiced. The political function of the randomized clinical trial is to provide scientific warrant for practice (i.e., “We have conducted research that is supportive of, or consistent with, our treatment approach”).

³ The recent *Consumer Reports* (CR; 1995) study (see Seligman, 1995) was designed to ask an effectiveness question of the type, “Do psychotherapeutic interventions produce beneficial results as they were administered in clinics and offices?” The CR study focused on the patient’s perspective in the evaluation of psychotherapy and, more specifically, on retrospective reports from this perspective; hence, the study suffers from the limitations of any survey (e.g., unknown respondent bias). In actuality, the CR study addressed the question, “Do current or former patients report that therapy was helpful?” The resounding answer was “Yes.”

draws, 1996; Kopta, Howard, Lowry, & Beutler, 1994; Maling, Gurtman, & Howard, 1995; Pilkonis & Frank, 1988; Seligman, 1995; Simons, Gordon, Monroe, & Thase, 1995).

Why does the log-normal model fit? Could it be that the target of improvement differs across the course of treatment? Such speculation about the dosage model gave rise to the following three-phase conception of the change process that occurs in psychotherapy (see Howard, Lueger, et al., 1993).

Remoralization

Some patients are so beset by problems that they become demoralized and feel that they are at their "wits' end." This type of experience is pervasive and severely disrupts a person's ability to mobilize his or her coping resources. The person begins to feel frantic, hopeless, and desperate. Demoralization (Frank, 1973; Frank & Frank, 1991) tends to respond quickly to psychotherapy; remoralization is usually accomplished in a few sessions.

Remediation

A second phase of therapy is focused on remediation of the patient's symptoms, the symptoms that led that person to feel so upset and demoralized that he or she had to seek treatment. During this second phase, treatment is concerned with refocusing the patient's coping skills in a way that brings symptomatic relief. The attainment of symptomatic relief is more gradual and typically requires about 16 sessions (depending on the type of severity of these symptoms; cf. Kopta et al., 1994).

Rehabilitation

A third phase of treatment is probably what has traditionally been thought of as "psychotherapy" in that it is focused on unlearning troublesome, maladaptive, habitual behaviors and establishing new ways of dealing with various aspects of life (e.g., problematic relationship patterns, faulty work habits, and trouble-causing personal attitudes). During psychotherapy, the rehabilitation of life functioning is quite gradual, and the number of sessions required is dependent on the severity of disability and the particular area of problematic functioning (e.g., work, family, or self-management; cf. Maling et al., 1995).

To the extent that these three phases are distinct, they imply different treatment goals and, thus, the selection and assessment of different outcome variables to measure progress in each phase. This model also suggests that different interventions are appropriate for different phases of therapy and that certain tasks may have to be accomplished before others can be undertaken. For example, Howard, Lueger, et al. (1993) demonstrated that these three phases are probabilistically, sequentially, and causally dependent: remoralization → remediation → rehabilitation.

The outcome criteria for each of these phases are subjective well-being, symptoms, and life functioning, respectively (for a description of the relevant scales, see Howard, Brill, Lueger, O'Mahoney, & Grissom, 1995; Howard, Orlinsky, & Lueger, 1995; Sperry, Brill, Howard,

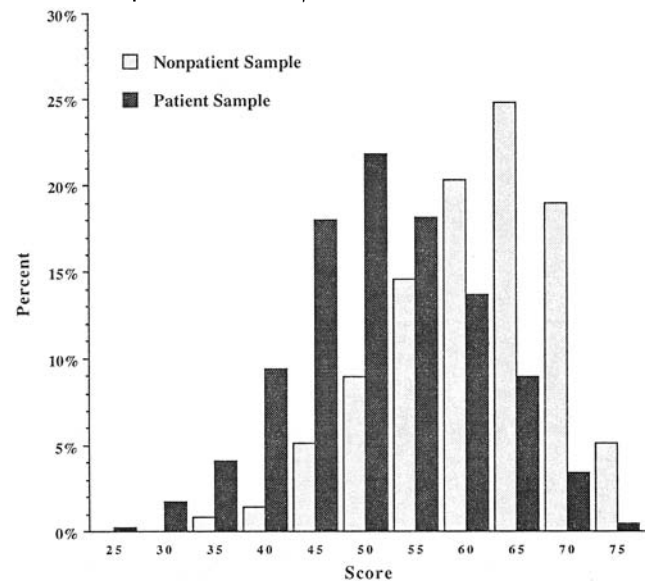
& Grissom, 1996). An overall treatment criterion, the Mental Health Index (MHI), consists of the sum of Subjective Well-Being, the Current Symptom total score, and the Current Life Functioning total score. The MHI has an internal consistency of .87 and a (three–four week) test–retest stability of .82. Figure 1 shows the MHI frequency distributions for a sample of 6,591 psychotherapy patients at the initiation of psychotherapy and for a sample of 493 nonpatients.⁴ The average MHI is clearly lower for psychotherapy patients than for nonpatients; people with MHI scores above 60 are much more likely to be nonpatients than patients (i.e., would be considered to be in the *normal* range; cf. Jacobson & Truax, 1991).

Patient Profiling and the Evaluation of Progress

Given a measure such as the MHI and a criterion for initiating or terminating treatment such as exceeding a predetermined medical necessity criterion (e.g., scoring below the *normal* range), it is relatively straightforward to plot the course of treatment for a patient. This simply requires periodic assessments of a patient's status on the selected outcome variables during treatment. Such information provides a description of a particular patient's progress. However, this descriptive information does not include a criterion against which the patient's progress can be evaluated. For example, it provides no information about how the patient

Figure 1

Frequency Distributions of Mental Health Index Scores for a Sample of 6,591 Patients at Initiation of Psychotherapy and a Sample of 493 Nonpatients



⁴ Mental Health Index scores in Figure 1 were converted to *T* scores ($M = 50, SD = 10$) that were based on psychotherapy patient norms at the initiation of treatment.

is doing compared with the rate and type of change that might be reasonably expected, given his or her clinical characteristics. In other words, every patient does not have the same expected outcome or expected course of treatment response, even to a well-standardized treatment for a well-specified clinical problem (i.e., even for a "validated" treatment). To accommodate this individuality, the patient profiling system that we have developed includes an estimated, expected course of treatment response (e.g., expected progress) for each patient that is based on his or her initial clinical characteristics.

To accomplish this individualized profiling, using a large sample of patients in outpatient psychotherapy, each patient's MHI was modeled as a log-linear function of session number. Following the dosage model, we assumed that the true response to treatment for each patient could be characterized by the relationship between the log of the number of sessions and MHI score. For each patient, this modeling yielded two parameters: intercept and slope. The intercept represented each patient's expected MHI at the first session; the slope represented the expected change in MHI per log of the session number. Given these estimates, it was then possible to search for predictors of these growth parameters by constructing models in which estimated intercepts or slopes were dependent variables. Hierarchical linear modeling (HLM) was particularly appropriate for this task (Bryk & Raudenbush, 1992).

Using HLM, we modeled the slope and intercept (dependent variables) for each patient as a linear function of 18 clinical characteristics (independent variables) of that patient at the initiation of treatment. These characteristics included such variables as severity of disturbance (assessed by the clinician and by the patient), chronicity of problems, pattern of presenting problems, and attitudes toward treatment (e.g., confidence that treatment will help). Thus, it was possible to generate an expected MHI score for any session of psychotherapy for each patient.⁵ These expected or estimated MHI scores could then be compared with the actual obtained scores for a particular patient in treatment.

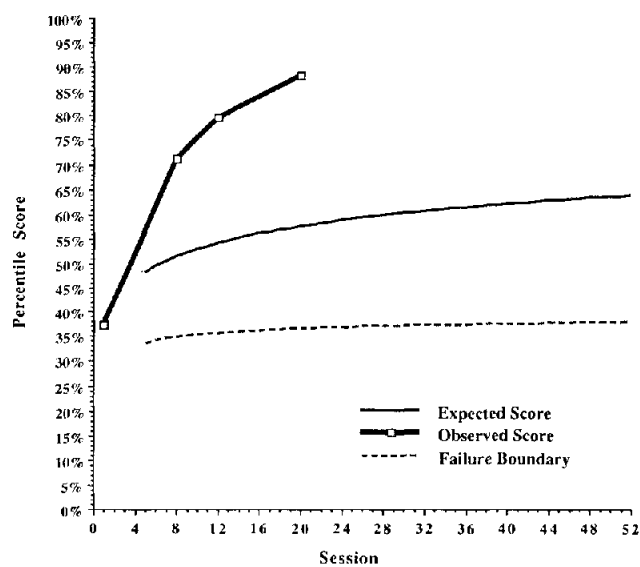
Using these estimates that were based on the patient's initial clinical characteristics, we created a graph that depicts the expected response to treatment for that patient. As periodic assessments of the patient become available during treatment, results can be entered on the same graph to provide a basis for evaluating whether the treatment benefits are occurring at a rate that could be expected for that patient.

Patient Profiling: Some Case Examples

Next, we present some examples of patient profiling. These cases were taken from a large ($N \sim 6,500$) data set of patients in outpatient psychotherapy in diverse settings across the country. Figure 2 depicts the course of therapy for Patient A, who was a 36-year-old, African American, remarried man. His clinical diagnosis was dysthymia. He presented with various family problems, including feeling overwhelmed at home, having financial problems, having a wife with failing health, and experi-

Figure 2

Course of Mental Health Index for Patient A



encing the recent death of a friend. During treatment, he worked on grief issues and on having more realistic expectations regarding his marriage. His wife was brought in for a few sessions, which seemed to help the therapeutic process. Inspection of the graph of this treatment indicates that Patient A made better progress than expected.

Figure 3 depicts Patient B, who began therapy in the average patient range in terms of MHI status (56th percentile). This patient was a 36-year-old, employed, married man diagnosed with an adjustment disorder. He sought treatment for a "marital relationship problem," with attendant job problems. His clinical characteristics predicted that therapy would be moderately effective at best. However, even the modest estimated progress and outcome were not achieved, and Patient B spent almost a year in an unproductive treatment.

Patient C (see Figure 4) was a 42-year-old, employed, married woman. She presented with significant family problems. This patient began therapy in the moderate range of severity according to her MHI but profited nicely (and as expected) from a year of treatment. The dip in her MHI status around Session 20 seemed attributable to the sudden death of her mother.

⁵ The model includes estimates of *fixed effects* (coefficients describing the relationship between the 18 intake characteristics and expected treatment response) and *conditional random effects* (estimates of slope and intercept variation not explained by the set of intake characteristics). The random effect estimates may be used to estimate residual variance away from predicted values given by fixed effect estimates. These residual deviations were a function of session number and were used to generate a 25th-percentile prediction bound below the expected values for any session and patient.

Conclusion

Despite the accumulation of evidence indicating that psychotherapeutic interventions are efficacious and effective for mental health problems, the provision of such treatments is faced with major challenges. Managed mental health care, coupled with concerns about growing costs, have challenged the field to document the comparative cost-effectiveness of different treatments in actual clinical practice. The difficulties of this challenge are many. First, the scientific methods for establishing efficacy and effectiveness often are not clearly linked to clinical application. Second, consensus is lacking about the outcome variables of interest, the relevant evaluative perspective(s), and the proper measurement technology. Third, little systematic work has been done on the application of research results to the treatment decisions and recommendations that practitioners must make.

The approach presented in this article—patient profiling—helps to address some of the foregoing challenges. The method is based on well-articulated and empirically supported guiding theories: the dosage and phase models of psychotherapy. The method can be used in conjunction with existing data from naturalistically conducted treatments to graph the expected course of progress for a patient. The resulting two sets of information (i.e., the patient's actual progress and the estimated progress for this patient that is based on the patient's clinical characteristics) can be used by a clinician, a managed care monitor, or both, to determine the appropriateness of the current treatment and the need for further treatment. Profiling information also can be used to prompt a clinical consultation for patients who are not progressing at expected rates.

Figure 3

Course of Mental Health Index for Patient B

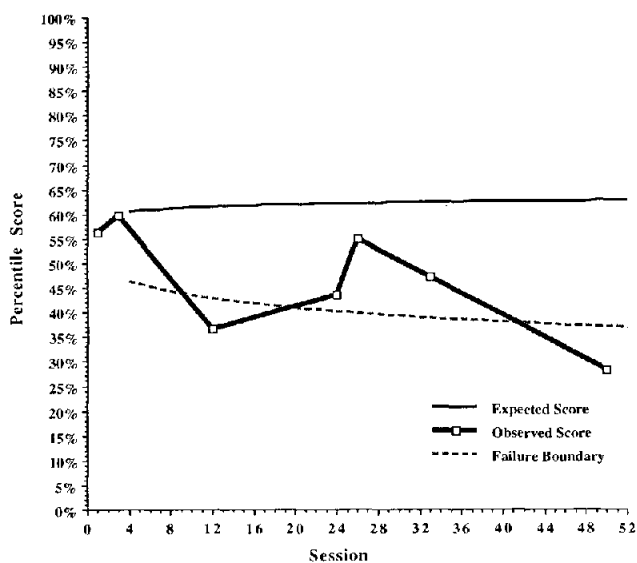
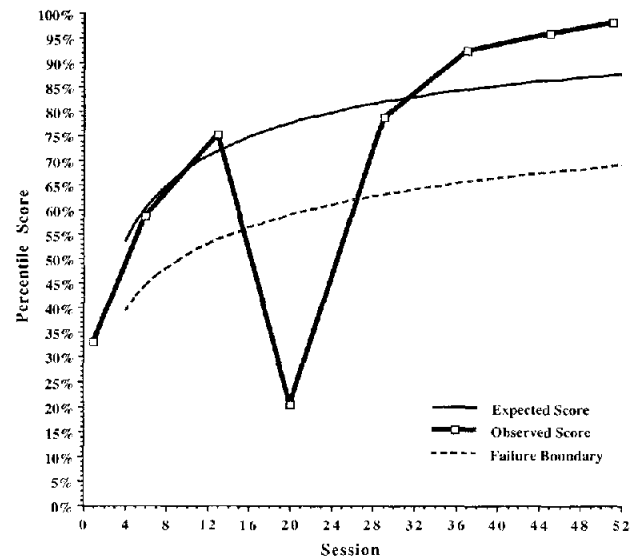


Figure 4

Course of Mental Health Index for Patient C



From a research perspective, the patient profiling method for modeling and monitoring treatment response allows us to pursue several goals:

1. We can evaluate the expected effectiveness of treatment.
2. We can group patients on the basis of their expected response to treatment and search for clinical consistencies within these groups.
3. We can study the characteristics of patients whose response to treatment deviates from expectation (e.g., examine faster responders, slower responders, nonresponders).
4. We can compare providers or provider groups on a case-mix adjusted basis (i.e., adjusting case loads for expected treatment responsiveness of the patients).
5. We can compare treatments in terms of dose-response relationships (the process of outcome) as well as in terms of final outcome.

In short, in contrast to information relevant to the average case that is provided through group comparisons, we can now provide information relevant to the particular case in treatment.

REFERENCES

- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Consumer Reports. (1995, November). Mental health: Does therapy help? 734-739.
- Cook, T. D., & Campbell, D. T. (Eds.). (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Frank, J. D. (1973). *Persuasion and healing*. Baltimore: Johns Hopkins University Press.

- Frank, J. D., & Frank, J. B. (1991). *Persuasion and healing: A comparative study of psychotherapy*. Baltimore: Johns Hopkins University Press.
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G., & Villasenor, V. S. (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*, 885-892.
- Howard, K. I., Brill, P. L., Lueger, R. J., O'Mahoney, M. T., & Grissom, G. R. (1995). *Integra outpatient tracking assessment*. Philadelphia: Compass Information Services, Inc.
- Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist, 41*, 159-164.
- Howard, K. I., Krause, M. S., & Lyons, J. S. (1993). When clinical trials fail: A guide to disaggregation. In L. S. Onken, J. D. Blaine, & J. J. Boren (Eds.), *Behavioral treatments for drug abuse and dependence* (NIDA Research Monograph No. 137, pp. 291-302). Washington, DC: National Institute for Drug Abuse.
- Howard, K. I., Krause, M. S., & Orlinsky, D. E. (1986). The attrition dilemma: Toward a new strategy for psychotherapy research. *Journal of Consulting and Clinical Psychology, 54*, 106-110.
- Howard, K. I., Lueger, R. J., Maling, M. S., & Martinovich, Z. (1993). A phase model of psychotherapy: Causal mediation of outcome. *Journal of Consulting and Clinical Psychology, 61*, 678-685.
- Howard, K. I., Orlinsky, D. E., & Lueger, R. J. (1995). The design of clinically relevant outcome research: Some considerations and an example. In M. Aveline & D. A. Shapiro (Eds.), *Research foundations for psychotherapy practice* (pp. 3-47). Sussex, England: Wiley.
- 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.
- Kadera, S. W., Lambert, M. J., & Andrews, A. A. (1996). How much therapy is really enough? A session-by-session analysis of the psychotherapy dose-effect relationship. *Journal of Psychotherapy Practice and Research, 5*, 132-151.
- Kopta, S. M., Howard, K. I., Lowry, J. L., & Beutler, L. E. (1994). Patterns of symptomatic recovery in time-unlimited psychotherapy. *Journal of Clinical and Consulting Psychology, 62*, 1009-1016.
- Krause, M. S., & Howard, K. I. (1976). Program evaluation in the public interest: A new research methodology. *Community Mental Health Journal, 12*, 291-300.
- Maling, M. S., Gurtman, M. B., & Howard, K. I. (1995). The response of interpersonal problems to varying doses of psychotherapy. *Psychotherapy Research, 5*, 63-75.
- Pilkonis, P. A., & Frank, E. (1988). Personality pathology in recurrent depression: Nature, prevalence, and relationship to treatment response. *American Journal of Psychiatry, 145*, 435-441.
- Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The *Consumer Reports* study. *American Psychologist, 50*, 965-974.
- Simons, A. D., Gordon, J. S., Monroe, S. M., & Thase, M. E. (1995). Toward an integration of psychologic, social, and biologic factors in depression: Effects on outcome and course of cognitive therapy. *Journal of Consulting and Clinical Psychology, 63*, 369-377.
- Sperry, L., Brill, P. L., Howard, K. I., & Grissom, G. R. (1996). *Treatment outcomes in psychotherapy and psychiatric interventions*. New York: Brunner/Mazel.
- Strupp, H. H., & Hadley, S. W. (1977). A tripartite model of mental health and therapeutic outcomes: With special reference to negative effects in psychotherapy. *American Psychologist, 32*, 187-196.
- Weisz, J., Donenberg, G., Han, S., & Weiss, B. (1995). Bridging the gap between the lab and the clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63*, 688-701.