
Efficacy and Effectiveness of Psychotherapy: Two Paradigms, One Science

Jim Mintz

University of California, Los Angeles

Robert E. Drake
Dartmouth University

Paul Crits-Christoph
University of Pennsylvania

Professor Martin Seligman (December 1995) added his voice to a growing chorus expressing the view that findings from controlled efficacy studies of psychotherapy need to be "transported" to more naturalistic studies of effectiveness in real-world service settings (e.g., Hoagwood, Hibbs, Brent, & Jensen, 1995). Describing and contrasting features of *efficacy* and *effectiveness* methods in psychotherapy research, Seligman asserted that "the efficacy study is the wrong method for empirically validating psychotherapy as it is actually done" (p. 966). His proposition that "the best features of these two methods can be combined into a more ideal method" (p. 965) echoes Clarke's (1995) belief that investigators from "both groups... should be encouraged to generate hybrid studies" (p. 724).

We agree with Seligman (1995) that effectiveness studies are needed to determine what treatment works in the field, and we agree that such studies are all too rare (e.g., Weisz, Donenberg, Weiss, & Han, 1995). We are concerned, however, lest Seligman's readers may draw from his arguments the erroneous conclusion that the transition from efficacy to effectiveness paradigms requires a relaxation of methodological rigor. To the contrary, as Clarke (1995) concluded, investigators must "broaden the sample and issues under study to address effectiveness concerns while still maintaining as much rigor and experimental control as possible to eliminate or minimize competing explanatory hypotheses" (p. 724).

Seligman (1995) described the design of the "ideal study" (p. 924) for determining the effectiveness of psychotherapy and medications as an uncontrolled, prospective survey of a large sample of treatment seekers, who are assessed before and after treatment with a comprehensive self-assessment and supplemented with "blind diagnostic workups" (p. 974). Cook and

Campbell (1979), certainly no enemies of quasi-experimental designs, concluded that

we should usually not expect hard-headed causal inferences from the simple before-after design when it is used by itself... Our hope is that persons considering the use of this design will hesitate before resorting to it and will decide to incorporate into it some of the design adjuncts we shall mention later. (p. 103)

Those design adjuncts are control conditions. Ever since Eysenck's (1952) legendary assault on conventional psychotherapy outcome wisdom, the need for appropriate controls has been generally accepted even by leading quasi-experimental methodologists. Research has repeatedly demonstrated that people assessed at a low point tend to improve, that most mental health problems are fluctuating or time limited, and that most patients are satisfied with almost any treatment short of abuse (because of rationalization and many other psychological variables). Seligman (1995) wrote:

Because there are no control groups, the *CR* [Consumer Reports, 1995] study cannot tell us directly whether talking to sympathetic friends or merely letting time pass would have produced just as much improvement as treatment by a mental health professional. (p. 972)

A study design that leaves this question unanswered can hardly be called "ideal."

It is misleading to claim that the uncontrolled survey design of *CR*'s reader satisfaction poll is an exemplar of effectiveness research. True effectiveness research is as methodologically rigorous as efficacy trials, is sensitive to the same threats to internal validity, and typically includes many of the same fundamental methodological features (for the same reasons). To be sure, effectiveness studies often incorporate many of the elements that Seligman (1995) identified as missing from efficacy studies, such as use of typical practitioners and settings, variable amounts of treatment (often with ample patient choice), mixtures of treatments, inclusion of patients with multiple disorders, and measures of improvements other than symptom relief.

Effectiveness studies also tend to include several elements that Seligman (1995) neglected but that enhance knowledge about what works in the field, such as careful assessment of the characteristics of settings and providers; measures of the representativeness of samples; measures

of the actual treatments received, both type and amount; assessments of why treatments were not received; costs of treatment and, sometimes, other societal costs; and, most important, the use of controls, with quasi-experimental methods often used but with random assignment still the preferred standard. There are examples of effectiveness studies that incorporate all of these elements, including studies of medications (Essock et al., 1996), psychotherapies (e.g., Weisz et al., 1995), rehabilitation (Drake, McHugo, Becker, Anthony, & Clark, 1996), and other interventions.

In spite of their "artificiality," well-designed efficacy studies have made many significant contributions to our assessment of the value of real-life treatments, and a number of important methodological features that Seligman (1995) characterized as "niceties" (a word derived from the Middle French time period for *foolishness*) of the efficacy study deserve careful consideration in any thoughtful psychotherapy research plan. These include randomization, specification of treatment procedures, methods for assuring fidelity of treatment delivery, blind ratings, rigorous control groups, inclusion-exclusion criteria, and operationally defined outcomes. These niceties are essential scientific components if one hopes to draw any strong causal inference, including the attribution of treatment effects to the treatment *per se* and not to a variety of other variables.

In what we hope is a flourish of rhetorical excess, Seligman (1995) speculated that randomization "may turn out to be... a major flaw of the efficacy method... Random assignment, the prettiest of the methodological niceties in efficacy studies, may turn out to be worse than useless" (p. 974). We are willing to bet otherwise. Seligman is undoubtedly right that "appropriately assigning individuals to the right treatment, the right drug, and the right sequence of techniques... may be crucial to getting better" (p. 974), but that is no argument against randomization. If we knew what those "right" decisions were, research would not be necessary.

Randomization is a cornerstone of strong causal inference. It is by no means a panacea (e.g., attrition may threaten equivalence), and, in some situations (fewer than one might think), randomization is neither feasible nor ethical. Nevertheless,

the interests of internal validity are paramount when the cost of being wrong about a causal inference is high—e.g., when, because of experimental results, an ineffective policy could be implemented widescale or an effective one reduced in scope... The desirability of randomized experiments in field settings is less in question

than their feasibility.... We strongly suspect that a debased randomized experiment will still be superior to its feasible quasi-experimental and nonexperimental alternatives for inferring cause. (Cook & Campbell, 1979, p. 385)

Another of Seligman's (1995) criticisms of the efficacy method is that it tends to use highly "filtered" (p. 970) samples of "single-disordered" (p. 970) patients. He lauded CR for studying treatment benefits in samples of individuals with varied (and unspecified) diagnoses, including multiple problems and what Seligman termed *subclinical* states that may not meet any conventional diagnostic criteria at all. Seasoned veterans of psychotherapy research will recognize that studies of mixed, diagnostically unspecified samples were quite typical of psychotherapy research about 30 years ago. Such designs were largely replaced during the last several decades by studies aimed at clarifying the value of specific treatments for specific, objectively diagnosed problems (including, at times, comorbid conditions). Studies that use diagnostically unspecified samples simply cannot yield conclusions regarding any particular condition.

Consumers are clearly interested in knowing what works for specific problems. That is undoubtedly why CR (1995) included a section in their article specifying which treatments have been found to work best for various disorders. It is a somewhat disconcerting jolt to turn from Seligman's (1995) methodological prescription to CR's own diagnostically specific conclusions regarding "What Works Best?" (p. 787). On a one-page insert, CR summarized "the right treatments for your troubles" (p. 787). What are they? Let us focus on the psychotherapies. Under depression, CR discussed only cognitive therapy and interpersonal therapy; under anxiety—relaxation and cognitive-behavioral therapy; under panic—combined drugs and cognitive therapy; phobias?—systematic desensitization and flooding. "In agoraphobia," CR reported, "behavior therapy is best combined with an antidepressant drug" (p. 787). None of these specific recommendations could, or should, have been drawn from the kind of database CR obtained. Nor were they. All of them were obviously based directly on the results of randomized, controlled clinical trials of specific, well-defined treatments in diagnostically homogeneous samples.

As for treatment manuals, the standardization of psychotherapy that they typically impose reduces outcome variation that is due to therapist differences, making

it easier to draw valid inferences about treatment differences (Crits-Christoph & Mintz, 1991). That would be a poor trade-off of external for internal validity, however, if Seligman (1995) were right when he suggested that manuals may limit psychotherapy benefits by restricting the flexibility of therapists. The evidence is to the contrary. A recent meta-analysis (Anderson & Lambert, 1995) found that studies that used treatment manuals had larger effect sizes when compared with studies that did not use treatment manuals.

It is probably true, as the CR (1995) study found, that most people will be more satisfied with more of a service—not only mental health care but also education, financial advice, legal services, massage therapy, and many other services. Services have a cost, however, and effectiveness research must, therefore, attend to costs as well as benefits. If patients pay for health care themselves, they make decisions regarding how much care they wish to purchase and how much they can afford in relation to expected benefits. It is well-known, for example, that patients choose to purchase less mental health care as the copayments increase (Keeler, Wells, Manning, Rumpel, & Hanley, 1986). On the other hand, if society pays for mental health care, society must decide how much mental health care should be made available in relation to competing demands for other types of health care, as well as other services. This is a complex choice. Society must determine how much of what types of mental health care for what conditions should be covered by insurance in the same manner as general medical care. This is precisely the debate before the Congress at this point in time.

"We have met the enemy," Pogo once observed, "and he is us." Knowledge is built on the scientific method, in psychotherapy as well as other areas of health care. This is the fundamental difference between science and pseudoscience. If the field of psychotherapy research were to adopt Seligman's (1995) recommendation uncritically and accept uncontrolled, self-selected consumer surveys—or anything less than the most rigorous scientific standards—as a model for determining the benefits of treatments, it would be taking a giant step away from the rest of medical research and its staggeringly impressive accomplishments. Such a move would aim psychotherapy practice directly into Carl Sagan's (1995) prescientific "demon-haunted world," and, ultimately, toward the loss of scientific credibility.

REFERENCES

- Anderson, E., & Lambert, M. (1995). Short-term dynamically oriented psychotherapy: A review and meta-analysis. *Clinical Psychology Review, 15*, 503-514.
- Clarke, G. N. (1995). Improving the transition from basic efficacy research to effectiveness studies: Methodological issues and procedures. *Journal of Consulting and Clinical Psychology, 63*, 718-725.
- Consumer Reports*. (1995, November). Mental health: Does therapy help? 734-739.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation. Design & analysis issues for field settings*. Chicago: Rand McNally.
- Crits-Christoph, P., & Mintz, J. (1991). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *Journal of Consulting and Clinical Psychology, 59*, 20-26.
- Drake, R. E., McHugo, G. J., Becker, D. R., Anthony, W. A., & Clark, R. E. (1996). The New Hampshire study of supported employment for people with severe mental illness. *Journal of Consulting and Clinical Psychology, 64*, 391-399.
- Essock, S. M., Hargreaves, W. A., Dohm, F. A., Goethe, J. N., Carver, L., & Hipsman, L. (1996). Clozapine eligibility among state hospital patients. *Schizophrenia Bulletin, 22*, 15-25.
- Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology, 16*, 319-324.
- Hoagwood, K., Hibbs, E., Brent, D., & Jensen, P. (1995). Introduction to the special section: Efficacy and effectiveness in studies of child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63*, 683-687.
- Keeler, E. B., Wells, K. B., Manning, W. G., Rumpel, J. D., & Hanley, J. M. (1986). *The demand for episodes of mental health services*. (Report No. R-3432-NIMH). Santa Monica, CA: Rand Corporation.
- Sagan, C. (1995). *The demon-haunted world: Science as a candle in the dark*. New York: Random House.
- Seligman, M. (1995). The effectiveness of psychotherapy. The *Consumer Reports* study. *American Psychologist, 50*, 965-974.
- Weisz, J. R., Donenberg, G. R., Weiss, B., & Han, S. S. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63*, 688-701.

We thank Mark Kotkin of *Consumer Reports*. He generously provided us with copies of CR's research instruments; shared his candid assessment of the aims, the strengths, and the limitations of CR's survey; and reminded us with his unflagging scientific curiosity that we are all in this together.

Correspondence concerning this comment should be addressed to Jim Mintz, Methodology and Statistical Services Unit, 11301 Wilshire Boulevard (116AR), University of California, Los Angeles, CA 90073. Electronic mail may be sent via Internet to jmintz@ucla.edu.