MAKING SCIENCE MATTER IN CLINICAL PRACTICE: REDEFINING PSYCHOTHERAPY

Larry E. Beutler

Pacific Graduate School of Psychology/

Stanford University College of Medicine

Key Words:

Empirically Supported Treatment

Systematic Treatment Selection

Principles of Change

Prescriptive Treatment

Methodology

Psychotherapy

Footnote 1: This paper was originally presented under a different title as an invited address at the 2008 American Psychological Association meetings in Boston, Massachusetts.

Footnote 2; Correspondence on this article should be sent to the author, Pacific Graduate School of Psychology, 405 Broadway St.. Redwood City, CA 94063

Abstract

Evidence suggests that the well-known chasm that exists between science and practice is maintained less by the intransience of practitioners than by the failure of scientists to: 1) provide a workable model of how to integrate science to clinicians and 2) to recognize the pervasiveness within their own ranks of several myths about the strength and viability of research findings and methods that provide the basis for identifying empirically supported treatments (ESTs). A rational weighing of the status of current evidence behooves scientists to take another, more careful look at why empirically supported treatments (ESTs) have failed to distinguish themselves from other treatments and to use this information in framing a broader approach to psychotherapy research.

MAKING SCIENCE MATTER IN CLINICAL PRACTICE: REDEFINING PSYCHOTHERAPY

Originally, this paper was prepared and presented as an invited address for the 2008 American Psychological Association meeting in Boston. It was originally to be addressed to clinicians under the title, "Why Science Matters to Clinicians, Even if They Don't Know It". As I was reviewing the research literature in order to develop this latter theme, however, it rapidly became apparent that there was considerably less evidence than I had supposed to suggest that science mattered or should matter to clinicians. As I tried to summarize an arsenal of research findings, I concluded that under the original title, I could make only three relatively weak points: 1) Psychotherapy works better than no treatment at all and about as well as or better than most other treatments for most problems; 2) Many things that are done in the name of psychotherapy do not work and in fact can be harmful; and 3) both the therapist and the patient are important to the outcome of psychotherapy. There was not a lot of earth shaking news there.

As I struggled unsuccessfully with how to make these obvious points meaningful and interesting, it occurred to me that the problem was that I was speaking to the wrong audience. My intended audience of clinicians already knew these things. Contemporary scientific findings, I realized, had much more to say to scientists than to clinicians about what would advance our understanding and optimization of clinical effects. More specifically, I became convinced that scientists were using methods and defining psychotherapy and research-informed practice in ways that hindered clinicians from being optimally effective. And, they didn't seem to realize it. Since I believed (and still

do) that scientific methods offer the best hope of finding optimal and effective ways to intervene with behavioral health problems, I was forced to re-think what science has given us and then to identify the disconnects between scientific assertions and scientific evidence..

I began by revising my title—"Making Science Matter in Clinical Practice: Redefining Psychothrapy"—in order to capture the new direction I was compelled to take. With my revised title, the objectives of this paper have become clearer: 1) to establish the importance of a broad research armamentarium and agenda to replace the narrow one that has increasingly shackled clinical science in the past thirty years; 2) to propose a research-informed definition of "psychotherapy" that fits clinical practice and use it to replace the narrow view of psychotherapy that has been used by the Empirically Supported Treatments (ESTs) movement; and 3) to stimulate a dialogue about the nature of "Research-Informed" practices that (I hope) will result in a more practice-friendly role of research than is currently used in the contemporary EST movement.

To achieve these objectives, I will first review, briefly the history of a continuing struggle that has characterized those who seek to "know"—i.e., to identify and gain knowledge. Simply put, this struggle to know, through which all civilization has and is going, is stimulated by two opposing models of "truth". One model defines what is "true" as that which is avowed and owned by a recognized authority. The other model emphasizes the concept of a truth that can only be seen darkly or probabilistically, one that is pursued and refined through a process of discovery. These disparate models constitute, if you will, a contrast between a top-down definition and a bottoms-up definition of what is true, and the escape from the devastation that has occurred by

reliance on the top-down model requires rethinking some of the research assumptions that we have applied to the problem of psychotherapy efficacy.

In order to stimulate a dialogue with those who care, I will illustrate the advantages of the premises and definitions that I propose. Specifically, I will describe what I think are some compelling findings arising from an interesting combination of research methods in my own and my colleagues' research laboratories and clinics. These findings illustrate the advantages of the broad definitions of practice and science that I will illustrate as a backdrop to these examples.

The Struggle Between Truth via "Authoriyt/Faith" and truth via "Science/Discovery"

The Evolution of a Two World View of "Truth"

Until the 13th century, it was commonly believed that religious leaders had access to definitive and unerring answers to matters of "truth". Thus, faith in the infallibility of the Pope speaking from his papal throne was seen as providing a beacon to the common person that would keep one blanketed by truth. Indeed, such faith was often thought to be the one and only way to real knowledge, and lest one forget this fact, these religious authorities had at their disposal the means to revive your memory. These means were severe but many a memory was revived by threats of the martyr's fire, beheadings, and the like. However, for hundreds of years, there was little need for these reminders and the infallibility of the Pope was almost without dispute. It was not until the advent and persistence of the black plague that relatively large numbers of rebellious individuals—largely scientists and progressive artists and philosophers—began looking for a cure for this disease through science rather than religion. In one sense, it is not too far fetched to suggest that the Black Plague opened the door both to the enlightenment of science and to

the atrocities of the Inquisition. For the first necessitated the second. It was the black plague that provoked questions about the validity of religious answers and that, in turn, caused the church to rise to denounce, punish, and even kill all those heretics and "witches" who sought answers from those in academic rather than priestly robes.

For nearly 600 years, from the 13th to the 19th centuries, the Inquisition was launched and advanced by the church to stop science and scientists from their investigations, to denounce them and to punish them for the audacity and heresy of seeking truth through discovery and experiment rather than from Papal authority. It was during this time that the world became locked between two world views of how truth is to be known. Scientists, great thinkers—like Copernicus (1473-1543) who introduced the heliocentric view of the heavens, and Galileo (1564-1642) who first espoused and then was made to denounce the Copernican view—were tortured and eventually became martyrs for preaching a scientific approach to truth. It was not until 1808 that the torture to which these and others were subjected was legally outlawed and the Inquisition was deprived of its major tool of enforcement. It appeared, for the moment, that science and objectivism had won.

Contemporary Expressions of the Struggle Between Two World Views

The legal sanction against torture as an enforcement tool served neither as an enduring antidote that would advance the scientific method nor as a final guard to stop the use of torture to enforce authoritarian views of truth. Many of the major political and social issues of our day remind us that this two world view struggle has not died. Daily, we see efforts made to define truth by pitting objective evidence and scientific discovery, on one hand, against appeal to religious or theoretical authorities, on the other. We see

continuing examples of governments and organizations who use fear of God's punishment and eternal death to persuade non-believers to give up their sexual and lifestyle practices; we still see rigid pronouncements of religious theory being substituted for scientific findings in schools and government. And, as it has for centuries, the two-world view is invoked in arguments for and against stem cell research; creationism versus evolution; when life begins and abortion rights, and the list goes on. Everywhere, we are forced to choose which source of truth we accept.

While learned societies, like our own, typically have come down in favor of objectivism and scientific discovery in these debates, within our profession there are still strong pulls that threaten to draw us back to appeals based on faith in authority and to ignore systematic evidence. The draw between appeals to faith and appeals to science, within the field of psychology, are most directly seen on two fronts. At the most obvious level, we can observe the two-world view in the struggle to identify and promote an acceptance among practitioners, treatments that are "empirically supported" or "Research Informed". This struggle frequently pits the subjective experiences and theoretical adherent of some practitioners against those who seek to discover truth through evidence rather than theoretical allegiance. The field remains pulled to methods whose only support is someone's good intentions and a theory. At least some of these clinical methods have been associated with great harm (Beutler, 2000; Lillifield, 2007).

But not all that pulls us away from evidence is as obvious as the draw between theory and science. At a more nuanced and insidious level, the two world view has permeated clinical science itself. It is played out in a contest between two sides, both of whom claim the mantel of science. In a replay of the Nicean and Crusader wars, the

argument resolves into a battle to find the one, best scientific method—i.e., a "gold standard" by which to define and apply truth in clinical practice. Some science-devotees have become consumed, not with the question being addressed nor the appropriateness of the method used, but whether the finding was based on one *particular method of science*. The most obvious, but not the only example of this transmigration of science to methodolotry is seen in the reverence given to Randomized Clinical Trials (RCT) as the sole means of defining ESTs. This view contrasts with the less absolutist view that the *goldness* of the standard used depends on the suitability of procedure to the questions and variables being studied. Thus, we have seen tensions escalate and occasionally explode as EST scholars argue over the imperative value of basing conclusion on RCT research designs, regardless of all other considerations. On each side of such arguments we find those who seek pronouncements of an unerring "truth", a gold standard that can be fixed by their consistency with the pronouncements either of a theory-based authority or a statistical edict. These top down criteria replace a thorough, thoughtful, and systematic probability statement.

Myths About Empirically Supported Psychotherapy

The contemporary struggle between the sides of this two-world view of "knowing" have led many scientists to over value certain kinds of scientific findings and to ignore others. Let me illustrate the truth of this assertion with a thought experiment. Regardless of whether you consider yourself more of a scientist than a practitioner or vice versa, answer the following three questions *as you believe the preponderance of scientific studies has found.*

1. (True or False) Psychotherapy would be more effective if everyone practiced an “Empirically-Supported Treatment”.
2. (True of False) Cognitive and Cognitive-Behavioral Therapy are more effective than relational and insight-oriented forms of psychotherapy.
3. True of False) The Relationship between the patient and therapist determines most of the meaningful outcomes that can be attributed to psychotherapy.

Now, think to yourself about the source of your conclusions. Can you remember a research report or study—at least the authors or the year and journal in which it appeared? Or, does your answer depend on what you have heard some “expert” say at a meeting? Or, does it just fit with your own theory of what “should” be true, rationally?

In fact, *the best scientific evidence available indicates that all three of these assertions are myths*—they are false or so nearly so that it makes little difference. This conclusion is based on a compilation of best evidence derived from several different sources. It reflects a strong preponderance of the evidence and is reflected in many replicated meta-analytic studies that have addressed these questions. Meta-analysis is simply a way of combining results across studies and as such, allows us to determine what trends and effects are present when the potential of errors that exist in individual studies are averaged out. Such analyses average the effects found in all the available studies that meet defined criteria and report them as that proportion of a normal curve that separates the compared treatments. This comparison, or *Effect Size* (ES) is reported as the statistic, *d*. The trends that will be summarized in the following paragraphs are replicable and commonly known, but their magnitudes and meanings are both ignored and surprising.

1. Myth #1: Psychotherapy would be more effective if everyone practiced an "Empirically-Supported Treatment".

There are many, scientists and scholars who would probably accept this myth as being true, on its face. For evidence, they would probably cite two bodies of research. One of these would be drawn from the many studies and meta-analyses that have demonstrated that psychotherapy is an effective means of treatment when it is compared to the outcomes of those who receive no treatment, a delayed treatment, or a placebo treatment (e.g., Beutler, Malik, et al, 2003; Shapiro & Shapiro, 1982; Smith, Glass, & Miller, 1980; Wampold, 2001). Typically, such comparisons earn ES estimates ranging from $d = .6$ to $d > 1.0$ and average around $d = .80$), testifying to psychotherapy's effectiveness.

The other body of literature cited would be the many task force reports that identify the treatments that have met the various criteria that have been used to identify ESTs. Almost universally, such criteria require treatments to have been shown to be effective in two or more Randomized Clinical Trial studies (e.g., see reviews of these task force findings in Chambless & Ollendick, 2001 and in Beutler, Malik, et al, 2003). Over 150 models and brands of psychotherapy have met criteria established by one or another task force sufficiently to be identified as being Empirically Supported.

Unfortunately, both of these bodies of literature are frequently but incorrectly interpreted to support the value of empirically supported treatments over the usual therapy that is used. Neither of these bodies of literature have directly compared therapies that qualify as ESTs with Therapy as Usual (TAU) in clinical settings. On the other hand, direct comparisons of ESTs and TAUs have been done.

In a meta-analytic comparison of 90 studies (Shadish, Matt, Navaro, & Phillips, 2000), all of which compared a research-based, EST treatment and a "treatment as usual" condition, revealed few differences in the benefits (Effect sizes averaged about zero). Naturalistically applied psychotherapy was as effective as using structured manuals and research-derived interventions. Both among clinically representative and non-representative samples of patients, the two kinds of treatment produced similar levels of improvement.

In a large mega-analysis, Lipsey and Wilson (1993) reviewed and combined the effects of 302 meta-analyses. Each of these meta-analyses were based on individual studies that compared a research-based form of psychotherapy with various naturalistic interventions, largely based on psychoeducational models of change. The authors found that highly structured, research treatments and the usual naturalistic treatments applied in uncontrolled, natural settings achieved equivalent results. As before, the Effect Sizes were near or at zero.

A concern with the foregoing study, however, was that it did not systematically differentiate between comparisons that used real clinical populations and those that used analogue or minimally distressed populations. Nor did this latter analysis distinguish among true psychotherapy conducted by psychotherapists and either educational interventions or those that used student therapists. A third meta-analysis corrected for the foregoing problems (Wampold, 2001). This latter analysis included studies that contrasted various treatments that were done in actual clinical settings, on actual patients, and that were conducted by experienced therapists. All identified treatments were compared with one another. The obtained effect size associated with these comparisons

was again zero—the structured and "proven" treatments did not differ from naturalistic treatments, unstructured ones, or usual treatments.

That is not to say that all studies obtained nonsignificant results. Indeed, a selective review of individual studies (e.g., Addis & Cardermil, 2007; Ollendick & King, 2006) will disclose some findings that favor one or another EST compared to a TAU condition (e.g.,Schulte, Kunzel, Pepping, & Schulte-Bahrenberg, 1992). However, it is equally easy to find individual studies that produce results favoring the TAU (e.g., Emmelkamp, Bouman, & Blaauw, 1994). The meta-analysis procedure derives an average effect across thousands of comparisons, and virtually every meta-analysis of this question has no systematic advantage favoring research-informed treatments over treatments as usual. The advantage of ESTs over a credible, treatment as usual, is virtually meaningless (see also, Beutler, Malik, et al, 2003; Duncan & Miller, 2006).

2. Myth #2: Cognitive and Cognitive-Behavioral Therapy are more effective than most relational or insight-oriented forms of psychotherapy.

While there are wide differences in ESs associated with different models of treatment (Beutler, Machado, & Neufeldt, 1994; Smith, Glass, & Miller, 1980; Shapiro & Shapiro, 1982), these differences typically disappear when the treatments are equally structured and delivered with equal skill. This lack of distinction favoring cognitive and non-cognitive therapies has been disputed, particularly as pertains to the treatment of patients with anxiety disorders (Chambless & Ollendick, 2001; Ollendick & King, 2006). Certainly, among both research-minded practitioners and research-oriented academic psychologists, it is commonly thought that cognitive and cognitive-behavioral therapies are more effective than other procedures, particularly psychodynamic and relationship

based ones (e.g., Beutler, Williams, Wakefield, & Entwistle, 1995). This impression probably arises from three sources: 1) a few highly cited studies that have found differences favoring these latter therapies, 2) an early meta-analysis that reported such differences, and 3) the general popularity of cognitive therapy in research studies and among research-oriented professionals.

In spite of the popularity, both of this belief and of cognitive therapies themselves, systematic and direct, head-to-head comparisons of cognitive and cognitive-behavioral models with psychodynamic therapy, interpersonal therapy, experiential therapy, and other models, have failed to yield reliable and strong differences. The first such comparison was conducted by Smith and Glass (1977) in a comprehensive analysis that included every counseling and psychotherapy available at the time. This first excursion into the use of meta-analysis yielded some differences in favor of cognitive and cognitive-behavioral therapy, finding that were repeated when the authors published their findings in a book (Smith, Glass, & Miller, 1980). However, in the latter report, the authors corrected their findings by adjusting for the variations in the reactivity of the outcome measures. They noted that under these circumstances, the differences disappeared. This led them and authors of later meta-analyses to accept what has come to be called, the Dodo bird verdict (Luborsky, Singer, & Luborsky, 1970). That is, all psychotherapies, structured or not and cognitive or not, produce essentially equivalent findings when tested against one another (e.g., Shapiro & Shapiro, 1982; Berman, Miller, & Masserman, 1985; Grissom, 1996; Wampold, 2001).

In perhaps the most carefully done meta-analytic assessment of the specific effects of cognitive therapies, Wampold, Mondin, Moody, Stich, Benson, and Ahn (1997)

clustered all treatments into classes (e.g., cognitive, cognitive-behavioral, psychodynamic, etc.) and then compared both each class of treatment with every other class and every treatment within a class with others within that same class. The best estimate of an overall, mean ES representing these comparisons was zero. Cognitive Therapy accounted for less than 1% of the variations in differences among the therapies when compared directly. An inspection of these effects broken down by specific disorders does not change the picture. A recent meta-analysis of long term, psychodynamic therapies compared to various other alternatives, in fact, reveals that for chronic problems and personality disorders, insight oriented therapies may be the treatments of choice (Leichsenring & Rabung, 2008). Collectively, it seems quite clear that there is virtually no reliable evidence to suggest that cognitive therapy outperforms other major forms of psychotherapy in the treatment of either anxiety disorders or depression.

3. Myth #3: The Relationship between the patient and therapist determines most of the meaningful outcomes that can be attributed to psychotherapy.

An alternative to the perspective that treatment methods determine most of psychotherapy outcomes is often advanced by scholars who disagree with the RCT research methods and devotion to theoretical models that are used in EST research. This alternative asserts that the principle effects of psychotherapy are derived from the quality of the interpersonal or working relationship that is developed between the patient and the therapist (Norcross, 2002). This is a persuasive argument, and like the belief that some treatments are better than others, has earned the devotion of a large group of scholars and even a larger number of clinical practitioners. And, like the evidence that psychotherapy is effective, there is a broad range of research that confirms the value of therapeutic

relationship as a contributor to therapeutic change (e.g., Norcross & Lambert, 2006; Duncan & Miller, 2000; Wampold, 2001; Westen, Novotny, & Thompson-Brenner. 2004).

A review of meta-analytic studies of the role of therapeutic alliance or relationship in outcomes confirms the presence of a consistent correlation between various aspects of the therapeutic relationship and the client's experience of benefit. While consistent, an inspection of these findings reveal that the magnitude of this relationship is relatively and surprisingly small. For methodological reasons, effect sizes in this literature are usually expressed as correlations, but we have translated them to *d* statistics to ease the cmparison with the preceding results.

Stevens, Hynan, and Allen, (2000) reported a mean effect size of *r*=.11 (*d* = .03). Somewhat larger relationships have been reported by others, however. Horvath and Symonds (1991), for example, reported effect sizes ranging from *r* = .21 (d = .11) to *r* =.26 (*d* = .17), while Martin, Garske, and Davis (2000) reported a mean *r* of .23 (*d* = .13). These ESs, while consistent, are small and of less predictive power than frequently attributed to them. They suggest that the therapeutic relationship is important but accounts for less than 7% of the variation among outcomes. This hardly qualifies as support for the truth of the relationship myth.

What Can We Conclude?

The two world view continues to dominate clinical psychology and especially dominates our views about psychotherapy. Strong beliefs, theoretical preferences, dictates of charismatic leaders, and our own good intentions are preferred over scientific data as sources of knowledge. The fact that these myths are held by scientists impedes

the advance of more striking scientific findings by binding them to using and advocating methods of science that do not lead to optimizing psychotherapy effects.

I do not offer these latter conclusions lightly. But, I have been forced to them as I inspect carefully what are the most consistent research findings available from a large body of research. This struggle to understand what scientific evidence can do to improve clinical practice has led me to a series articles of faith which may help the reader clarify some of the foregoing points.

Five Articles of Faith Related to Psychotherapy Research

From the forty years of experience that I have devoted to the task of seeking truth in clinical practice, as a scientist, as a practitioner, as a teacher, as a consultant, and as a psychotherapy patient, I have drawn five basic conclusions. I have framed these as "Articles of Faith" since I continue to encounter disagreement about them.

1. Evidence based on strong belief alone, only works for a short while.

Throughout the 100+ year history of psychotherapy, the principle test of the efficacy of any intervention has been the personal belief of the therapist, usually based on the testimony of a charismatic authority (Cummings & O'Donohue, 2008). When the opinion leaders of our field replace the search for discovery, we become little different than a religion and the result is a tower of Babel around what is true and what was really meant by the original Guru. The nexus of this conflict is in the tendency to attribute to leaders a degree of infallibility and to mark truth by the number of followers one has rather than by the veracity of the assumptions that one pronounces.

The faith given to these pronouncements, in the absence of more substantive and replicable evidence of their value has proven to be of limited lasting value. One thinks,

for example, of the many survivors of Hurricanes Katrina, Rita, and Ike who had faith that their houses would survive the oncoming storm; that the levees in New Orleans would hold; that they would be rescued; and that FEMA would come to their aid. Now, three years after Hurricane Katrina, hundreds of people with such faith are still dead, the levees are still points of danger, and thousands have yet to be able to rebuild their houses or relocate permanently.

In psychotherapy, interventions whose evidence relies on this top-down model of knowledge transmission, have enjoyed widespread appeal. Programs in *rebirthing therapy, reprogramming therapy, past lives therapy, recovered memory therapy,* and many others that are based on pronouncements of a leader, more than on scientific evidence, have come under fire when incidents occur to reveal their ability to harm others (Beutler, 2000; Lilienfeld, et al, 2003). In some cases, such as rebirthing therapy, the result has been death, in others, such as reprogramming therapy, it has been the psychological destruction of lives and families (Beutler, 2000; Public Broadcasting System, 1995). These incidents of harmful effects are often so striking that they become headlines of sufficient concern as to almost invariably reflect poorly on psychotherapists and psychotherapy.

For example, when rebirthing therapy was revealed in a newspaper and subsequent television stories, to have produced the death of a 10 year old girl, the effects were widespread and all of psychotherapy was harmed (Mercer, Sarner, & Rosa, 2003). Psychotherapists of all kinds were forced to share the shame. When the errors of judgment and the probable weaknesses of the method so strikingly are at discord with the hype and expectation, it does not take a research study to bring their practice to a close

and to turn the light on all of psychotherapy. Reliance on personal experience, and especially on the experience of others, no matter how strong one's good will and beneficent intentions, constitute weak support in the face of such public opinion.

2. *The common socially derived alternatives to science, as the basis of evidence, also provide weak protection against ineffective and mal- practice.*

Historically, there have been three publically and legally used criteria to determine when a treatment is effective: a) a community standard of common practice; b) a case law standard of a respectable minority; and c) a health care standard of cost effectiveness (Beutler, Clarkin, & Bongar, 2000). The first two of these socially derived and commonly accepted standards have traditionally been applied to determine when a procedure can be considered to constitute malpractice. The third is one that has long been used in health care settings to determine the clinical value and reimbursement of a treatment or treatment program.

The standard of common practice requires that one be able to justify one's practice decisions by ensuring that it is acknowledged by and is acceptable to other practitioners in one's community. Indeed, it requires that the procedure be shown to be in common or daily use. In other words, there must be evidence that the treatment is popular (Black, 1990; Klerman, 1990). This standard relies heavily on evidence that other, similarly trained and experienced individuals, practice in similar ways and with similar treatments.

The standard of a respectable minority arose from case law, specifically out of concern that the standard of common practice was insensitive to emerging but not yet popular treatments. This standard recognized that the health care fields do not always

have a consensual view of what is effective. This standard requires a demonstration that a "significant" minority of practitioners share a belief that has been articulated in legal procedures and that has defined a standard of practice (Furrow, 1980; Klerman, 1990).

Unfortunately, case law (Hood v Phillips, 1976) has defined this latter standard in a way that virtually ensures that everyone who is not protected by the common practices rule, can fall under the "respectable minority" protection. This case law defines a "Respectable Minorioty" as being as few as six individuals who share a favorable opinion of the treatment, and a single written articulation of how the treatment is done, as being adequate evidence of its value (Beutler, Clarkin, & Bongar, 2000).

The third standard was the original one used to test the efficacy of managed health care programs and in modified form, continues to be heavily weighted in contemporary discussions of a treatment's value (Aaron, 1996). It pits the number of people served by the procedure against the cost of distributing these services as a measure of effects. Thus, a "good enough" treatment is one that is delivered to many but costs nothing.

None of these three criteria for determining efficacy are based on objective outcomes. And herein lie their failing. All of them require an analysis of how the treatment is delivered, rather than one that measures how much change it has produced. There are no blinds or protections against the influence of unfounded opinion and because of this and other factors, none of the criteria provides a true protection against a treatment that is harmful but popular and cheap.

While there have been emerging improvements in both court standards and in healthcare policy deliberations, all as a function of becoming more reliant on the findings of objective and systematic scientific research (e.g., Daubert v Merrell Dow

Pharmaceuticals, 1993), the draw and attraction of these unsubstantiated criteria remain strong. Indeed, the value of scientific standards is verified by the very evidence that changes have taken place in these criteria. The increasing reliance on scientific findings have brought a concomitant increase in the stability and replicability to both legal and health care arenas.

3. *Randomized Clinical Trials (RCTs) are a viable scientific option for addressing some treatment questions.*

When the NIMH decided to support the Treatment of Depression Collaborative Program (TDCP) in the 1980s, it was an innovative and interesting idea. I recall a meeting of the Society for Psychotherapy Research (SPR) in 1986—a meeting that I organized at Wellsley College—at which the keynote speaker, Dr. Gerald Klerman, who was then head of NIMH, introduced the use of RCT designs as a viable way to study psychotherapy. He emphasized that we must come to view psychotherapy like we do aspirin. That is, each psychotherapy must have known ingredients, we must know what they are, they must be trainable and replicable across therapists, and they must be administered in a uniform and consistent way within a given study.

In the service of those objectives, RCT research methods were initiated and rose to prominence as the required methodology for determining ones status as an EST. These research methods have largely become the *gold standard* by which claims of efficacy are assessed for truth. And, from its inception in the TDCP (Elkin, 1994), RCTs have always placed a great deal of emphasis on the assurance that all therapists delivering the treatment are doing the same thing. Interestingly, some recent scholars (e.g., Addis & Cardemil, 2006), when faced with concerns that manualized training may stifle the

creativity and individuality of therapists, have argued that RCT designs are not designed to restrict the flexibility and personal creativity of the therapist (p. 151). If such an assertion were true, it is hard to imagine why such emphasis is given to training therapists to a defined level of reliability with one another—the higher the better. What is a high reliability estimate intended to do if not to ensure that all therapists are behaving in closely similar ways? If this is not true, why has such emphasis been given to identifying and retraining any outlying therapists. Particularly "creative" therapists are either trained to recapture the original reliability standard or are dropped from the analysis as being non-representative.

There was great resistance to Klerman's proposal that the research paradigm be translated into a narrow RCT model, at that 1986 SPR meeting, largely because doing so would limit psychotherapy to an assessment of what the therapist did with particular diagnostic groupings of patients. Such an approach would ignore the personal characteristics and interpersonal compatibility of the therapist and patient involved. While these early concerns have not dissipated substantially with time (e.g., Beutler, 2004; Chambless & Ollendick, 2001), we have learned some interesting things since the advent of RCTs in psychotherapy and from the empirically supported treatment (EST) movement that this methodology spawned. I am not referring here to the treatments that comprise the many lists of ESTs now available. (Beutler, Malik, et al, 2003). As I have attempted to document here, I don't believe these lists are very helpful in optimizing or even increasing the effectiveness of psychotherapy. The head-to-head comparisons of different therapies suggest that most rationally derived therapies are equivalently effective.

More interesting, I believe, is the rather paradoxical evidence that while most operationalized treatments earn equivalent results, some psychotherapies are ineffective and even harmful (Beutler, 2005; Lilienfeld, 2007; Lilienfeld, Lynn, & Lohr, 2003; Singer & Lalich, 1996). It appears to be easier to identify a bad treatment than a very good one, the latter falling prey to the Do-do bird. Many, well known treatments are ineffective or even harmful. These treatment include such treatments as Drug Abuse and Resistance Education (DARE), Recovered Memory Therapy that is often used to treat female victims of rape, Grief Counseling for Bereavement, Expressive-Experiential therapies, and the most widely used treatment for acute effects of mass trauma, Critical Incident Stress Debriefing (Lilienfield, 2007). When compared to no-treatment or placebo treatments, some accepted interventions earn ESs that are negative. That is, some treatments make people get worse. *When therapist effects are reduced to as low a level as possible,* effects often (or at least more often than desirable or expected) are negative, indicating that the average effect of the treatment is deterioriation.

If the foregoing tells us anything, it is that when some forms of psychotherapy are found to be effective, it may be *in spite of* the treatment, not because of it. The patient, the therapist, or the way they are paired, may offset the negative effects of the treatment techniques themselves, to facilitate change. That is, the beneficial effects of the therapeutic process may arise because of the resilient aspects of patients, the therapeutic qualities of people—things that cannot be randomly assigned to treatments—or from interventions that cannot be randomly trained. This realization led me to a fourth article of faith.

4. *Some research questions are not effectively addressed with RCT designs and are best answered by naturalistic and quasi-experimental studies.*

As one reads the EST literature in juxtaposition with the actual results of studies and meta-analytic comparisons of treatments, it is impossible not be impressed with the disparity between these formal results and how strongly author(s) emphasize the importance of the findings. While the results have largely failed to be terribly impressive, EST scholars continue to assure us that only random assignment studies are of sufficient scientific note as to provide believable evidence of psychotherapeutic efficacy (Chambless & Hollon, 1998). Many EST scholars seem to be oblivious to the implications of such disparities, but there are those who have recognized them as indicators that the real influences in psychotherapy include results that are associated with variables that are non-randomly distributed aspects of the therapist, the relationship, and the patient (Duncan & Miller, 2006; Castonguay & Beutler, 2006). The inability of a variable to be randomly assigned should not dictate whether studies of these variables are considered to be sufficiently scientific to warrant the attention of scientists. Indeed, there are many constructs that are central to non-psychology sciences that are not appropriately or possibly studied through random assignment. Nor should randomization be a major criterion that determines the worth or merit assigned to a variable. Rather, the nature of the variable should be looked upon as a clue that can lead us to select among the available methodological procedures, those that are appropriate and sensitive to the kinds of characteristics that are being studied. Neither the Big Bang nor the theory of trans-species evolution have been subjected to randomized controlled trials. Nor for that

matter have natural disasters, terrorist events, and star movement been excluded from scientific study because they could not be randomly assigned.

In like manner, therapist and patient personalities, interpersonal values, therapist and patient gender, social skills and attachment levels, and the like, are incapable of being randomly assigned and yet are of sufficient worth as to be given scientific consideration in assessing psychotherapeutic efficacy. The influence of some of these non-randomized variables have been subjected to meta-analyses, and the ESs can be compared to those obtained from variables that have been subject to RCT studies. For example, a meta-analysis by Beutler, Malik, and colleagues (2003) revealed that relationship factors (d = .17) and the personal and professional characteristics of therapists (d = .30) account for meaningfully more of the outcome variance than that associated with the intervention model used (mean d< .11). All of these findings suggest that many extra-intervention contributors to psychotherapy are worthy of being included within our definition of *psychotherapy*. They make stronger and equally consistent contributions to treatment outcome than the more formal aspects of the interventions themselves. Such observations underline the importance of the fifth article of faith.

5. *Changing the definitions of "psychotherapy" and of "Research-Informed Practice" (RIP)that is used in research are required to advance our understanding of their importance.*

The articles of faith as articulated in the foregoing have led my research colleagues and I to shift the research definition both of "psychotherapy" and of "Research-Informed Practice. The narrow view held by most EST research paradigms is not only unworkable when studying characteristics, qualities, and variables that are not

appropriately or even capable of being assigned to people randomly, but are inconsistent with the way that such variables are conceptualized in clinical work. Such variables as are embodied in the person of the therapist, or are captured within the patient's response dispositions, and those that index a degree of fit between the selected therapy and the patient must be considered a part of psychotherapy, itself. These variables are, or should be, central to developing effective treatments in clinical practice and should be given equal attention within the context of psychotherapy research. They deserve study as *part of* and *central* to the psychotherapeutic process, not just as interesting but incidental correlates of what is considered a psychotherapy comprised of disembodied procedures.

Accordingly, our research group has redefined psychotherapy for research purposes in order to be more consistent with the definitions operationalized within clinical practice. We have come to believe that separating the person of the therapist from the acts of psychotherapy—in the manner suggested by the medication metaphor proposed by Klerman—is unsupportable in psychotherapy research. If, as we have proposed, these aspects of character, preference, fit, and expectation, contribute more consistent and stronger predictive power in outcome assessments than the technical aspects of the interventions, then they ARE THE TREATMENT.

Specifically, we define psychotherapy, both in clinical and research applications, as: <u>The therapeutic management, control, and adaptation of patient factors, therapists' factors, relationship factors, and techniques factors that are associated with benefit and helpful change.</u>

Shifting the definitions from constructs derived from theories of psychopathology and psychotherapy, to the integration of patient, therapist, intervention, and relationship

components, has led to a marriage among treatment methods (Nathan & Gorman, 2002); participant predictors, and Empirically Supported Relationships (Norcross, 2002). Concomitantly, the change in the way that psychotherapy and research evidence is defined shifts us from relying on a narrow range of methods by which to extract "truth" and a similarly narrow range of models and patients (e.g., RCTs) to the investigation of one or more research-informed principles of effective therapy (Beutler, Clarkin, & Bongar, 2000; Castonguay & Beutler, 2006). It will also move the field to analyze the role of dimensional dynamics and interactions among therapist activities, patients and problem traits (severity, personality, etc.) and therapists, rather than maintaining the static and categorical view of the process that currently dominates the field.

To illustrate the differences that would be invoked by broadening these definitions of psychotherapy and RIP, I will briefly summarize four studies that illustrate the use of a wide array of methods and strategies to converge on reliable findings. The first study was a retrospective, quasi-RCT study that used archival data to identify and then cross-validate principles of strategic change and methods of measuring important variables (Beutler, Clarkin, & Bongar, 2000). The second study (Beutler, Moleiro, et al, 2003) was a combination of an RCT and a regression analysis which illustrated some of the strengths of these methods when used together to provide a further confirmation of the interactions among therapist, intervention, relationship, and fit of the treatment and patient. The third study (Johannsen & Beutler, in preparation) was a cross-cultural validation study of two basic principles of change, using a quasi-experimental design to inspect the fit of therapy and patient factors. And, the fourth study (Kimpara, Henderson,

& Beutler, in preparation) was a naturalistic cross-validation of two treatment principles arising from the first investigation.

Together, these studies provide a reasonable ES estimate of the gains in ESs that are associated with integrating multiple variable domains and a broadened definition of psychotherapy. These studies illustrate the application of multiple design elements to get a more comprehensive picture of optimal therapy than that which is possible using an RCT methodology, alone.

Study 1. Beutler, Clarkin, and Bongar (2000) undertook a three stage, quasi-RCT study of variables that predict and determine the effectiveness of psychotherapy. The study began with an exhaustive review of over 2000 outcome research studies in order to define and then validate the role of patient and therapist characteristics, treatment dimensions, the fit of treatment to patient, and the therapeutic relationships that are associated with outcomes. This review also provided the data to extract from extant research findings, 15 hypothetical principles that describe the relationship among these variables and outcomes.

The first phase of the study identified patient and treatment qualities that had been associated directly with outcomes as well as those that constituted well matched dyads of patient and treatment. This phase also resulted in an articulation of clinically friendly principles that predicted how outcomes would emerge as these variables interacted with one another. In the second phase of the study, instruments were developed to measure the variables that had been the bases for these strategic hypotheses. These instruments were designed to tap patient factors, qualities of the therapeutic relationship, and the dimensions that constituted a good treatment fit. Patient qualities were tapped through

independent clinical ratings; aspects of the interventions that were associated with good outcomes were assessed through rater observations; and aspects of good fit between patient and treatment were measured by combining these ratings of patient and treatment.

A third phase of the study provided a direct test of the strategic hypotheses utilizing archival data on 289 subjects. These participants represented depressive spectrum and chemical abuse disorders and were drawn from four different RCT samples along with one naturalistic sample. Seven different manualized models of psychotherapy, a manualized medication regimine, and a therapy as usual (TAU) condition represented the ways that these patients had been assigned and treated in these samples. All patients had been randomly assigned to one of a subset of the treatments, and within each data set, therapists were randomly assigned to the study interventions. Patient entry data for this study were based on the measures developed in the second phase and were applied to recordings of pre-assignment intake interviews and a variety of personality and symptom measures taken at intake. The therapy, relationship, and treatment fit variables were defined by ratings of early and late psychotherapy sessions using the therapy process measures developed in phase two. Outcomes were assessed by standard measures of psychological well being taken at pre-and post therapy. Ratings of therapy activities and the fit of the therapy to patient characteristics were all applied to patients in the 9 different treatments to ensure that all ratings were independent and uniform.

The data were analyzed by a series of structural equation models with post-hoc analyes of specific relationships. The results provided support for 13 of the 15 original hypotheses. The 13 supported hypotheses were re-framed in the form of strategic principles that could be used to inform and guide the therapist and to provide assistance

in developing a strategic plan for treatment. Five additional principles were derived from a consensual analysis of clinician ratings for dealing with dangerous patients and added to the total. Thus, 18 guiding principles were extracted from the findings, variously representing suggestions about developing a therapeutic relationship, assigning a context of therapy (treatment intensity, location, mode, and format), implementing common classes of interventions (directive and insight interventions and emotional regulation procedures), and adapting the intervention to accommodate moderating aspects of the patient's personality and problems. Some aspects of treatment were directly related to outcomes and served as direct prognostic indicators. Patient variables also served as moderators of outcome and indexed changes in therapeutic interventions. Patient factors, such as levels of functional impairment, coping styles, levels of trait-like resistance to change, and level of distress served as indicators for varying aspects of treatment (e.g., treatment intensity, insight-behavioral focus, therapist directiveness, and use of emotional confrontation).

Of greatest importance, this study exemplified the use of extant research findings to develop a model treatment that was tailored to each patient and then to cross validate the usefulness of this model. It combined RCT and naturalistic designs and in that process, gave some hints about the multiple and interactive qualities that affect outcome. By looking, at once, at therapy models, therapy procedures, patient characteristics, therapy context, and relationship factors, patterns among these variables emerged.

Study 2. Beutler, Moleiro, Malik, et al, (2003) studied 40 co-morbid depressed and chemically abusing patients using an RCT design. Patients were randomly assigned to one of three therapy models, including a prescriptive therapy (Beutler & Harwood,

2000) that was based on 10 of the 18 principles derived from Study 1. The treatments included a standardized cognitive behavioral intervention (Wright, Newman, & Liese, 1993) and a narrative intervention (Moreira, Beutler, Goncalves, 2008) in addition to the prescriptive procedure. Patients, treatments, and therapists (within treatments) were randomly assigned and matching characteristics that fit the treatment to the patient, using four dimensions were randomly determined and monitored. Analyses were undertaken in two stages: 1) analysis of treatment model and 2) an analysis of patient, therapy, relationship, and treatment fit using regression models. While the therapy models were relatively equivalent in efficacy, patient, treatment, relationship, and treatment fit variables contributed independent variance to the benefits obtained. It was the fit of the treatment to the patient that accounted for the greatest degree of long term change while treatment techniques tended to lose their effects in a relatively short period of time.

This study demonstrates the relative superiority of including patient and relationship factors within the definition of treatment. When considered only as therapist behaviors, all three therapy models produced similar effects. However, when therapy/therapist factors (d = .20), patient factors (d = .40), relationship factors (d = .40) and treatment fit factors (d = 1.40) were included within the definition of the treatments, strong effects were observed, especially for the overall compatibility between the patient qualities and the nature of the treatment. The fit of treatment accounted for the strongest effects of all variable classes by one year after treatment.

Study 3. Johannsen and Beutler (in preparation) applied a naturalistic design to a sample of 92 patients who were seen either in the United States on in Argentina. All patients were assigned to therapists and level of treatment fit using random procedures.

All outcomes were assessed using standardized pre and post measures. Patients were followed for three months or until they terminated treatment, whichever came first. Analyses of two patient-therapist matching dimensions were conducted separately. The fit between patient coping style and therapist use of symptom-focused (external coping styles) or insight (internalizing coping styles) based interventions were strongly related to outcomes in both cultural groups. The better the match, the better the outcomes (d = .61).

The strategic fit between patient level of trait like resistance and therapist use of directive (for non-resistant patients) or non-directive (for resistant patients) procedures was related to outcomes among the Argentine patients. The U.S. sample failed to show a strong effect of treatment fit on this latter dimension but this result was found to be moderated by another variable, level of severity or impairment. A relationship was only found among patients who were rated as being at least moderately distressed and impaired (d = .83).

This study confirms that patient variables serve as differential indicators for modifying aspects of psychotherapy both in Argentine and U.S. samples, though such predictors appear to be somewhat more complex among U.S. samples. Once again, if a typical RCT analysis had been undertaken of these patients, neither the effects of treatment procedures nor the moderating effects of patient factors would have been disclosed. By using a broad definition of psychotherapy, a corresponding complex analysis was suggested and more detail was revealed about the optimization of psychotherapy effects.

Study 4. A final example of varying methods and the role of broadening the definition of psychotherapy was found in a recent naturalistic investigation of a

homogeneous group of shy (avoidant and internalizing) individuals. In this study (Kimpara, Henderson, & Beutler, in preparation), a structured treatment protocol had evolved and been tailored to work with shy and avoidant individuals based on clinical experience. The treatment began with an eight week course of symptom focused, cognitive therapy and the followed by a 16-24 week course of psychodynamic psychotherapy.

Happenstantially, the structure of the therapy used in this study corresponded with two principles that had been extracted from the results of study 1. Specifically, that study suggested that an optimal treatment for internalizing/avoidant individuals would consist of a symptom oriented phase followed by an insight-oriented phase. This natural occurring experiment provided an opportunity to test these principles. Shy (internalizing) patients could be expected to benefit from treatment as a function of how closely the therapists followed the two-phase treatment.

Indeed, a multiple regression and growth curve analysis revealed that treatment benefit was related to both compliance and the dominance of patient internalization tendencies (d =.76). The use of a natural experiment to confirm previously observed findings is another condition in which multiple and flexible research methods are useful and helpful in clarifying the relationships among variables that, broadly, constitute psychotherapy.

The convergence of findings is notable among these four studies, each utilizing different research methods and designs, and all based on different samples and different levels of treatment analysis. Collectively, the results confirm the validity of many of the principles originally defined in the review of extant research (Beutler, Clarkin, & Bongar,

2000). In each case, however, the direct analysis of the therapy(ies), was strikingly uninformative without considering the role of patient, intervention, and relationship factors. Only when the mix of these variables were incorporated within the definition of "psychotherapy" were we able to see a path to optimal treatment.

Conclusions

In this paper, I have advanced the thesis that the way that psychotherapy is studied and defined in contemporary EST research is unnecessarily narrow and impedes the search for optimal clinical effects. This deficiency in contemporary research practices reflect a long standing struggle between two views of how knowledge can best and most accurately be obtained. As a result of this narrow view, scientists have tended to over rate and over sell the extent of the knowledge gained, a point is illustrated by three myths that permeate this field.

In its original form, this struggle for knowledge was one that pitted evidentiary-based knowledge against faith-based knowledge—the church against science. While this struggle continues and is expressed within factions of clinical practice, in its more recent renditions, faith based knowledge has hidden under the cloak of science. But, rather than adopting the openness to information and the search for novelty and improvement that has traditionally set science apart from faith, the new generation of top-down knowledge devotee is a worshiper of methodology. Those who worship at this alter have accepted the assumption that a single scientific method, usually the randomized clinical trial, is the only acceptable way to study the effects of psychotherapy and to render conclusions about how well it works. They make this assumption while eschewing and ignoring the

more basic principle of all science that leads one to select the method used in research in order to be responsive to the demands of the question asked and the variables studied.

By falling prey to methodolotry, research has under rated the degree of predictable and controllable change that is or can be associated with psychotherapy. It has done this by imposing a definition of psychotherapy on research that ignores and devalues the breadth of variables and factors that clinicians have implicitly acknowledged to be as important to outcomes as the procedures that are used. By ignoring the role of non- or extra-diagnostic factors, relationship factors, and how treatments might fit with the patient's experience and problems, the EST movement has ensured that its findings are weak and less than useful to clinical practice.

Drawing on a handful of studies that have focused on evaluating the level of fit between research-defined treatment and patient qualities, my students, colleagues, and I (Beutler, Moleiro, et al, 2003; Beutler, Malik, et al, 2003; Johannsen & Beutler, in preparation; Satoko, Henderson, & Beutler, in preparation), have found uniformly moderate and large effect sizes (ranging from d = .30 to .91) associated with therapy variables. Such findings confirm that research investigations of psychotherapy would be well to maintain a flexible view of the therapeutic process; one that extends beyond what the therapist does to include when and how he/she does it.

In this paper, I have identified five articles of faith that I believe support the conclusion that psychotherapy, in research as in practice, is a process that includes all variables within the network of systems that are and can be used to facilitate gains and benefits. In turn, research-informed practice, it is argued, must be more than RCT-informed practice and must draw findings from all relevant, reliable, and systematic

scientific methods into our understanding of treatment effects. Through examples of research that integrates multiple research and statistical methods into psychotherapy studies, I have attempted to lay the groundwork for a discourse among well-meaning scientists about the nature of knowledge. It is my hope that this discourse will expand and extend the role of science in clinical practice in ways that will advance both our knowledge and psychotherapeutic practice.

References

Aaron, H. (1996). End of an era. *Brookings Review, 14*(1), 35-37.

Addis, M. E., & Cardemil, E. V. (2007a). Psychotherapy manuals can improve outcomes. In J. C. Norcross, L. E. Beutler, & R. F. Levant (Eds), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions* (pp. 131-140), Washington, DC: American Psychological Association.

Addis, M. E., & Cardemil, E. V. (2007b). Dialogue: Convergence and contention. In J. C. Norcross, L. E. Beutler, & R. F. Levant (Eds), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions* (pp. 149-151), Washington, DC: American Psychological Association.

Berman, J. S., Miller, C., & Masserman, P. J. (1985). Cognitive therapy versus systematic desensitization: Is one treatment superior? *Psychological Bulletin, 97,* 451-461.

Beutler, L. E. (2000). David and Goliath: When psychotherapy research meets health care delivery systems. *American Psychologist, 55,* 997-1007.

Beutler, L. E. (2004). The Empirically-Validated treatments movement: A scientist-practitioner's perspective. *Clinical Psychology: Science and Practice, 11,* 225-229.

Beutler, L. E., Clarkin, J. F., & Bongar, B., (2000). *Guidelines for the systematic treatment of the depressed patient.* New York: Oxford University Press.

Beutler, L. E., & Harwood, T. M. (2000). *Prescriptive Therapy: A Practical Guide to Systematic treatment selection.* New York: Oxford University Press.

Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., &

Wong, E. (2003). Therapist variables. In M. J. Lambert (Ed), *Handbook of psychotherapy and behavior change* (5th ed.)(pp. 227-306). New York: John Wiley and Sons.

Beutler, L. E., Moleiro, C., Malik, M., Harwood, T.M., Romanelli, R., Gallagher-Thompson, D., & Thompson, L. (2003). A comparison of the Dodo, EST, and ATI indicators Among Co-Morbid Stimulant Dependent, Depressed Patients. *Clinical Psychology & Psychotherapy, 10,* 69-85.

Beutler, L. E., Williams, R. E., Wakefield, P. J., & Entwistle, S. R. (1995). Bridging scientist and practitioner perspectives in clinical psychology. *American Psychologist, 50,* 984-994.

Black, H. C. (1990). *Black's law dictionary (6th ed.).* St. Paul, MN: West.

Castonguay, L. G., & Beutler, L. E. (Eds)(2006). *Principles of Therapeutic Change that work: Integrating relationship, treatment, client, and therapist factors*. New York: Oxford University Press.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66,* 7-18.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52,* 685-716.

Cummings, N. A., & O'Donohue, W. T. (2008). *Eleven Blunders that Cripple Psychotherapy in America.* New York: Routledge Press.

Daubert v Merrell Dow pharmaceuticals (1993)(92-102), 509, U. S. 579.

Duncan, B. L., & Miller, S. D. (2000). *The heroic client.* San Francisco: Jossey-Bass.

Duncan, B. L., & Miller, S. D. (2006). Treatment manuals do not improve outcomes. In J. C. Norcross, L. E. Beutler, & R. F. Levant (Eds), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions* (pp. 140-149), Washington, DC: American Psychological Association.

Elkin, I. (1994). The NIMH treatment of depression collaborative research program: Where we began and where we are. In A. E. Bergin and S. L. Garfield (Eds.) *Handbook of Psychotherapy and Behavior Change* (4th ed., pp. 114-139). New York: John Wiley and Sons.

Emmelkamp, P. M., Bouman, T. K., & Blaauw, E. (1994). Individualized versus standardized therapy: A comparative evaluation with obsessive-compulsive patients. *Clinical Psychology and Psychotherapy, 1*, 95-100.

Furrow, B. R. (1980). *Malpractice in psychotherapy.* Lexington, MA: Lexington Books.

Grissom, R. J. (1996). The magical number .7 ±.2: Meta-meta-analysis of the probability of superior outcome in comparisons involving therapy, placebo, and control. *Journal of Consulting and Clinical Psychology, 64,*973-982.

Hood v. Phillips, 537 s.w.2d 291 (Tex. Civ. App. 1976).

Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. Journal of Counseling Psychology, 38, 139-149.

Johannsen, B. E. & Beutler, L. E. (in preparation). *Matching therapy procedures to patient characteristics among U.S. and Argentine samples.* An unpublished study, Pacific Graduate School of Psychology.

Kimpara, S., Henderson, L., & Beutler, L. E. (in preparation). *Shyness and systematic*

treatment selection. An unpublished study, Pacific Graduate School of Psychology.

Klerman, G. L. (1986). *Keynote address*. Delivered to the annual meeting of the Society for Psychotherapy Research. Wellesley, MS.

Klerman, G. L. (1990). The psychiatric patient's right to effective treatment: Implications of *Osheroff v. Chestnut Lodge. American Journal of Psychiatry, 147 (4),* 409-418.

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2,* 53-70.

Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (2003). *Science and Pseudoscience in Clinical Psychology.* New York: Guilford Press.

Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies. *Archives of General Psychiatry, 32,* 995-1008.

Martin, D. J., Garske,, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. Journal of Consulting and Clinical Psychology, 68, 438-450.

Mercer, J., Sarner, L., & rosa, L. (2003). *Attachment therapy on trial.* Westport, CT: Praeger Press.

Moreira, P., Beutler, L. E., & Goncalves, O. F. (2003). Narrative change in psychotherapy: Differences between good and bad outcome cases in cognitive, narrative, and prescriptive therapies. *Journal of Clinical Psychology, 64,* 1181-1194.

Nathan, P. E., & Gorman, J. M. (Eds.). (2002). *A guide to treatments that work* (2nd ed.*).*

New York: Oxford University Press.

Norcross, J.C. (Ed) (2002). *Psychotherapy Relationships that Work: Therapist Contributions and Responsiveness to Patient Needs.* New York: Oxford University Press.

Norcross, J. C., & Lambert, M. J. (2006). The therapy relationship. In J. C. Norcross, L. E. Beutler, & R. F. Levant (Eds), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions* (pp. 208-218), Washington, DC: American Psychological Association.

Ollendick, T. H., & King, N. J. (2007). Empirically supported treatments typically produce outcomes superior to non-empirically supported treatments. In J. C. Norcross, L. E. Beutler, & R. F. Levant (Eds), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions* (pp. 308-317), Washington, DC: American Psychological Association.

Public Broadcasting System (1995, Sept.). Frontline. New York: Public Broadcasting Co.

Schulte, D., Kunzel, R., Pepping, G., & Schulte-Bahrenbert, T. (1992). Tailor-made versus standardized therapy of phobic patients. *Advanced Behavior Research and Therapy, 14,*67-92.

Shapiro, D. A. & Shapiro, D. (1982). Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin, 92,* 581-604.

Smith, M. L., Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32,* 752-76 0.

Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy.* Baltimore: Johns Hopkins University Press.

Stevens, S. E., Hynan, M. t., & Allen, M. (2000). A meta-analysis of common factor and specific treatment effects across oucome domains of the phase model of psychotherapy. *Clinical Psychology: Science and Practice, 7,* 273-290.

Thaler, M. & Lalich, J. (1996). *Crazy Therapies.* San Francisco, CA: Jossey-Bass.

Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings.* Hillsdale, NJ: Erlbaum.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, Aall must have prizes@. Psychological Bulletin, 122, 203-215.

Westin, D., Novotny, C. M., & thompson-Brenner, H. (2004). Ther empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin, 130,* 631-663.

Wright, F. D., Beck, A. T., Newman, C. F., & Liese, B. S. (1993). Cognitive Therapy of substance abuse: Theoretical rationale. In L. S. Onken, J. D. blaine, and J. J. Boren (Eds.) *Behavioral treatments for drug abuse and dependence.* (NIDA Monograph No. 137). Washington, D.C.: National Institute of Drug Abuse.