

## Identifying Empirically Supported Treatments What If We Didn't?

**Larry E. Beutler**

Counseling/Clinical/School Psychology Program University of California, Santa Barbara

### ABSTRACT

The conclusion of the Division 12 Task Force's report on empirically supported treatments raises 3 questions: (a) Is it desirable for the profession to specify what treatments are effective? (b) Do the criteria, either selected by the Task Force or modified by others, represent a reasonable way of identifying effective treatments? (c) Would different and less controversial conclusions have been reached if the criteria used were broadened to include naturalistic and quasi-experimental studies? It is concluded that the Task Force's selection of criteria, particularly as modified by [D. L. Chambless and S. D. Hollon \(1998\)](#), was a reasonable response to these pressures. Findings from studies using less stringent and controlled research designs suggest that the proposals may have resulted in less palatable conclusions than those offered in its original report.

---

Correspondence may be addressed to Larry E. Beutler, Counseling/Clinical/School Psychology Program, Department of Education, University of California, Santa Barbara, California, 93106.

Received: November 7, 1996

Revised: March 14, 1997

Accepted: April 8, 1997

---

The report of the Division 12 Task Force on Promotion and Dissemination of Psychological Procedures (Task Force; 1995) set a precedent. At a time when concerns with being politically correct have made it unconventional to fly in the face of popular opinion, the report espoused a clearly unpopular viewpoint among the populace of psychotherapists. It identified a small collection of treatments as *efficacious*, and by their omission, it implied that the others had ostensibly failed the test of clinical efficacy. The Task Force report was a response to intense social, economic, and political forces that demanded to know which of the 400-plus psychotherapies were effective. The Task Force responded in what seemed to be a fairly simple and straight-forward way—they identified a set of criteria and set about determining which of the many psychological procedures practiced were in compliance with these criteria. They made no claims that these were the only effective treatments, that these treatments should take precedence over

others, or that the list was finite and unchanging. Indeed, the Task Force mission mandated an annual review and update of the conclusions.

Those criteria by which efficacy was judged were adapted from those established and used for many years by the Federal Drug Administration (FDA), the committee represented major figures in the field, the conclusions were reached by discussion and review of countless research studies, and compliance with those criteria by the interventions that were included in either the first or the second ( [Chambless et al., 1996](#) ) annual reports were carefully documented. However, publication was met with strenuous, sometimes even angry, opposition.

The many voices raised in critique of the Task Force report, as well as the widespread resentment and outrage that it stimulated, focused on a number of assumptions and implied assumptions that were thought to underwrite the report ( [Drozd & Goldfried, 1996](#) ; [Fay & Lazarus, 1993](#) ; [Garfield, 1996](#) ; [Silverman, 1996](#) ; E. W. L. [Smith, 1995](#) ). Collectively, the major criticisms can be summarized as follows: (a) An unrepresentative and small number of studies constituted the basis of the review; (b) the restrictive criteria misrepresented the broad range of research findings; (c) overreliance on studies that used manualized therapies and random assignment led to inaccurate conclusions about the nature of psychotherapy and its effects; (d) the recommendations were likely to be misused both by managed health care (MHC) and training institutions to limit practice, training, or reimbursement to therapies favored by academicians; and (e) any conclusions beyond the general one that generic psychotherapy is effective, are premature.

The strength and severity of these criticisms give cause to reconsider the implicit and explicit assumptions that constituted the understructure of the Task Force's deliberations. Accordingly, the criteria were clarified and extended in key ways for the first annual Task Force update ( [Chambless et al., 1996](#) ). [Chambless and Hollon \(1998\)](#) propose the use of an even more refined and thoughtful set of criteria. This article references these various criteria to address three basic questions: (a) Is (was) it necessary or even important for the profession to specify what treatments are effective? (b) If so, do those criteria either originally selected by the Task Force or modified by Chambless and Hollon represent a reasonable way of responding to this need? (c) Would the selection of even more liberal criteria result in different conclusions about the nature and effectiveness of treatments? In this presentation, I focus largely on the original Task Force criteria, noting when and how the implications of conclusions may be tempered or changed with the revised and liberalized criteria suggested by [Chambless and Hollon \(1998\)](#) .

## **Should the Profession Identify Scientifically Supportable Treatments?**

Among the criticisms of the Task Force report, a major controversy has arisen between those who argue that any listing of effective treatments is premature (e.g., [Drozd & Goldfried, 1996](#) ; [Fay & Lazarus, 1993](#) ; [Garfield, 1996](#) ) and those who assert that the creation of such a list is timely, but that the criterion of inclusion should be expanded to be more clinically relevant ( [Beutler, Kim, Davison, Karno, & Fisher, 1996](#) ; [Silverman, 1996](#) ). A resolution of this conflict may be facilitated by inspecting the circumstances

that initially gave rise to the formation of the Division 12 Task Force. Placing the Task Force in historical perspective will help to determine if, indeed, a significant need existed or continues to exist for such a document. Resolving this issue may help allay the concerns of those who might see the report as a malicious and premeditated effort to disenfranchise the practitioners of certain methods.

In 1992, when the Task Force was conceived, the United States was in the throes of a debate about implementing a national health care policy. Psychotherapy was in the very real danger of not being covered in any of the proposed plans. The first of three stages of the MHC system ( [Aaron, 1996](#) ) was coming to an end. This stage had infused MHC since the early 1980s and emphasized cost containment through restriction of patient access and the establishment of closed provider panels. Hence, the Task Force was conceived in an environment in which treatment effectiveness was assessed by a formula that included immediate cost and general access, but that paid little or no attention to whether the treatment actually produced some clinical benefit. In the managed care systems of the early 1990s, psychotherapy was considered to be a homogeneous treatment and psychotherapists were considered to be interchangeable. Thus, doctoral practitioners of psychotherapy were being systematically replaced either by those who could manage psychopharmacological treatment or by those whose level of training justified cheaper service.

Both legislation and case law reinforced the nonempirical criteria of effectiveness adopted by MHC organizations, but unlike these latter groups, the law forced a distinction among the psychotherapies. Recognizing that psychotherapy practitioners do not hold consensual opinions about what treatments work, beginning in the late 1970s, the courts adopted two standards for determining whether a treatment was effective for purposes of addressing allegations of malpractice or when adjudicating claims against MHC procedures. Both of these standards were based on a review of the theoretical model that the therapist used. These standards of assessing effectiveness were (a) the principle of *the community standard* and (b) the doctrine of *respectable minority*. The first of these principles reduced the task of selecting effective treatments to matters of popularity and common use. Treatments that are frequently practiced within a given community are held, by this principle, to be valid and true, regardless of their clinical effectiveness or their potential for harm. Reliance on this principle obviates the need for scientific research at all, subserving it to the whims of popular appeal to professionals within a given region.

Yet, the court also attempted to address the fact that profound splits in the mental health field frequently prevented the emergence of a clear community standard. They developed the doctrine of respectable minority to address the fact that none of the professionally initiated efforts to establish diagnostic and treatment standards have enjoyed widespread acceptance within the mental health professions. This doctrine assumes the necessary task of defining effective treatments when the professions fail to develop such guidelines.

This respectable minority doctrine holds that where there are disputes based on theoretical differences and methods of practice, the clinician is entitled to be judged according to the school he or she professes to follow. This school must be one with definite principles, and it must be the line of thought of a respectable minority of the profession. Case law established the rule that as few as six individuals constitute a respectable minority, however. Of further note, medical case law established that such a

minority consensus is a valid test of a treatment's effectiveness, even when research clearly revealed that the practices are either of no benefit or actually dangerous to patients' health, safety, and well-being (see [Beutler, Bongar, & Shurkin, in press](#) ).

The practice of psychotherapy was faced with the possibility that the courts and legislative bodies would define, nonempirically, what types of psychotherapy could and could not be practiced and reimbursed in a system of managed competition. In response, Division 12 elected to convene a group of scientist—practitioners with the commission that they use research data to craft answers to the questions of what treatments are of value. Although the press from national health care has abated since the inception of the Task Force, there continue to be persuasive sociopolitical forces that auger for a response from the profession. And, again, the most likely and obvious price of nonresponse is to turn the decision of what treatments should and should not be covered in a managed competition environment, over to the law and other non-psychotherapist bodies.

Two contemporary examples illustrate the continuing need for the profession to take a stand in identifying effective treatments: (a) MHC programs are beginning to shift their perspective from a focus on cost to a focus that would require a treatment to prove itself empirically, as one of the criteria for reimbursement ( [Aaron, 1996](#) ; [Sperry, Brill, Howard, & Grissom, 1996](#) ); <sup>1</sup>and (b) professional organizations in medicine and primary care have already grasped the initiative from psychotherapy specialty groups in defining treatment guidelines. These medical groups have, understandably, given priority to medical treatments over psychotherapeutic ones as front-line interventions. They have done so even though devaluing psychotherapy is not consistent with the relative empirical value of psychological and psychopharmacological approaches to treating either anxiety or depression (e.g., [Barlow, 1994](#) ). Scientific comparisons of psychosocial and psychopharmacological treatments have been favorable to psychotherapy, particularly when long-term costs and benefits are considered (e.g., [Evans et al., 1992](#) ; [Hollon et al., 1992](#) ; [Steinbrueck, Maxwell, & Howard, 1983](#) ).

In sum, even a modest understanding of the current sociopolitical environment provides ample reason to conclude that identifying those treatments that are effective was and continues to be a viable response to critical social developments in MHC. Although some may still doubt whether any such list can be valid, it would be difficult to conclude that attempting to construct a list of effective treatments was not a logical response to both the emerging social issues of the time and to the continuing health care environment. The selection of an empirical standard by which to address questions of treatment benefit, moreover, both fits the pragmatic zeitgeist of our time and has traditionally been favorable to psychotherapy. The alternative may well have been to rely on implicit or tacit knowledge as managed by legislators, the judiciary, and health care administrators. That this may be quite unacceptable to psychotherapy practitioners and scientists alike goes without saying. So why is a scientific standard of such concern to psychotherapists themselves? It is to this question that I now turn.

## **Are Contemporary Criteria Reasonable and Fair?**

A substantial minority of the criticisms leveled against the Division 12 Task Force have argued that

although some effort to identify beneficial treatments, on the part of those who best understand psychotherapy, was justified and necessary, those criteria adopted were too narrow (e.g., [Beutler et al., 1996](#); [Silverman, 1996](#)). The point of these arguments emphasizes concerns with generalization, balance, and representativeness in the conclusions reached by the Task Force, because of the reliance on randomized clinical trials (RCT) research methodologies. Three specific concerns have been expressed: (a) RCT designs place undue weight on questions of efficacy rather than effectiveness (e.g., [Howard et al., 1996](#); [Seligman, 1995](#)); (b) the selection of uncomplicated diagnostic groups in these research designs fail to represent the needs and responses of the usual, complex, comorbid, patient seen in practice ([Silverman, 1996](#)); and (c) the manualized therapies used in such designs do not adequately or validly represent the psychotherapy that is conventionally practiced (E. W. L. [Smith, 1995](#)).

A frequent theme in these criticisms is that the set of criteria should include quasi-experimental studies rather than those that only use RCT designs or a critical mass of controlled,  $N = 1$  designs. These studies would replace the use of manualized treatments with monitored but naturalistic treatments; and uncomplicated diagnostic groups would be replaced by both more representative, comorbid patient groups and monitoring of nondiagnostic but treatment-relevant patient variables. [Chambless and Hollon \(1998\)](#) propose what amounts to a compromise by liberalizing the criteria of efficacious treatments both to reduce the number of  $N = 1$  replications needed and to reduce the standard applied to group designs from an independently replicated demonstration of specificity (superiority of treatment to a control group that used nonspecific interventions only) to one that required only a demonstration of efficacy (superior to no treatment or equal to an established treatment). They continue to reject the use of quasi-experimental and naturalistic (i.e., effectiveness or clinical utility) studies, however. In the midst of the many and often vehement criticisms sounded against the valuation given to RCT group designs by both of these proposals, few of the subsequent critiques specify what limits should be imposed on the additional criteria to ensure accuracy. It has also become clear that expanding the criteria would not necessarily mean losing the ability to study either select patient samples or purified forms of treatment ([Howard et al., 1996](#); [Seligman, 1995](#)).

Although I am certainly no devotee of randomized clinical trials, manualization, those criteria originally defined by the Task Force, or even of the assumption that there are diagnosis-specific treatments ([Beutler, 1996](#); [Beutler & Davison, 1995](#); [Beutler et al., 1996](#)), I have become convinced that many of the criticisms are overdetermined and nonobjective. The exaggerated reactions tend to cloud, rather than clarify, relevant issues, and leave the impression that the concerns are self-serving and arise from the fear that the ineffectiveness of one's favorite viewpoint will be found out.

[Silverman \(1996\)](#), for example, correctly observes that a number of the critiques that have been raised against the Task Force report are lacking in substance. He then presents arguments of his own, at least two of which are of equally weak substance, and both of which distract the focus of argument away from the critical issue of those criteria of inclusion selected. For example, he claims that the Task Force relied on nominations of favored therapies advanced by committee members. This claim misrepresents both the way nominations were solicited and the complex process that followed the nomination of a therapeutic approach ([Chambless, 1996](#); Chambless, et al., 1996). Specifically, this claim ignores two facts of Task

Force functioning: (a) The process of nominating treatments included but was not restricted to Task Force members—nominations were accepted from any interested party or group who believed that a particular intervention met the requisite criteria; and (b) nominating a therapy was only the first step in a time-intensive and extensive process of becoming listed as probably efficacious or well-established in efficacy. The Task Force process was clearly different from the more conventional procedure of selecting a panel of experts from among a special interest organization and then developing guidelines by a combination of panel and membership vote, such as was used in the construction of the "Primary Care Guidelines for the Treatment of Depression" developed by the Agency for Health Care Policy and Research (AHCPR; see [Barlow, 1994](#)).

In a related critique, [Silverman \(1996\)](#) also expressed chagrin over the fact that the report contained only 30 references. This critique reflects Silverman's failure to understand the difference between a Task Force report and a scientific paper. His pronouncement of the report as "poor scholarship" (p. 213) ignores the fact that such a report is a summary of recommendations based on the body of literature. It is not a review article and, thus, does not cite all of the evidence considered in reaching a conclusion. It addresses only enough literature to satisfy the criteria that support inclusion of a given treatment. In this case, the Task Force report dealt specifically with the question of what treatments met minimal criteria of benefit. A count of citations is an inaccurate measure of the literature reviewed en route to the conclusions. Responding to the mission assigned, the Task Force cited only enough examples from the literature searched to ensure that the listed treatments were in compliance with criteria. The categorical criterion of inclusion was of two independent, RCT studies or a series of  $N = 1$  studies, in the absence of conflicting data. Uncited literature was reviewed until the Task Force members could agree that a treatment had met the set of criteria. Two studies and meta-analytic reviews that supported (not determined) this determination were then cited in the report. Further citations would have been superfluous to the mission of the Task Force and would only have been necessary if this report had really been intended to be a scholarly review of literature.

As a reluctant, new (since 1995), initially skeptical, continuing disagreeable, and now cautious member of the Task Force, for example, I have reviewed dozens of studies for the 1996 and 1997 annual committee reports. These studies have focused on nominated treatments. Most of the time, these reviews resulted in a decision to delay advancing the approach to the report, pending additional research being available. In these instances, the work done will not be recognized or mentioned in the report, nor should it. On those rare occasions on which I recommended that a treatment be listed in the report, and on those even rarer occasions when my recommendations have been accepted by the other Task Force members, only the two or three studies that support this decision are included in the published report. Unacknowledged are the dozens of other studies reviewed but found wanting.

When one sees beyond such misguided criticisms, a telling critique remains around the question of whether the mission of the Task Force would have been better served by selecting a criterion that was more reflective of the question of treatment effectiveness than of treatment efficacy, a distinction made popular by [Seligman \(1995\)](#). From those who have offered this substantive critique (e.g., [Beutler et al., 1996](#); [Drozd & Goldfried, 1996](#); [Fay & Lazarus, 1993](#); [Silverman, 1996](#)), there have emerged two rather different suggestions for changing the criteria used: (a) Include quasi-experimental investigations of

treatment effectiveness, and (b) incorporate nonempirical evidence as one criterion.

The second of these alternatives is particularly interesting. Some (e.g., [Fay & Lazarus, 1993](#) ; [Silverman, 1996](#) ) have suggested that empirical guidelines derived from group research are insufficient to capture the individual needs of patients for tailored treatments. Implied by the examples offered, but largely left unstated, is the suggestion either that nonempirical standards be used to compensate for these deficiencies in empirical research or that efforts to identify effective treatments be abandoned altogether. The adoption of a standard that includes criteria that are lacking evidence of reliability and validity, in the equation that determines what treatments will be identified as scientifically supported, seems dubious at best and dangerous at worst.

In contemporary mental health practice, there are many examples of unfortunate occurrences arising from adopting a standard that can move between implicit and explicit (scientific) knowledge at the will of those who interpret the data. When clinical impressions and popular beliefs are at variance with empirical findings, the empirical data are almost surely to be ignored even when they are demonstrably more accurate. Clinicians' faith in their own perceptions and opinions far surpass their accuracy ( [Houts & Graham, 1986](#) ), and they are frequently persuaded by illusory relationships between observations, even when these relationships have no foundation in fact ( [Chapman & Chapman, 1969](#) ).

The results of relying on clinical impressions without benefit of scientific test are often disastrous. Witness the emerging stories of family disruption and disintegration that have arisen when the common clinical assumptions—that there are symptom-specific signs that indicate early abuse has occurred and that uncovering or recovering the memories of these experiences is desirable—are accorded a status in excess of that warranted by empirical literature on the topic.

The critical question in broadening the criteria in this way is how much weight to accord these nonempirical beliefs when they are at odds with scientific research. The danger is that such unstandardized observations would be given greater weight than standardized ones and that this could undermine the scientific credibility of the field. Witness, for example, one recent result of relying on procedures that have long been advocated by many psychotherapists who work to uncover repressed memories using clinical impressions as criteria of effects. Several months ago, the Public Broadcasting System presented an expo&sacute;e of the mythical relationship between multiple personality disorder and repressed memories of satanic abuse and indoctrination that was advanced by several psychologists and psychiatrists at Rush-Presbyterian Medical School (see Beutler et al., in press).

The relative credibility accorded implicit knowledge and belief over explicit and empirical knowledge places the problem of determining truth with those who advance the most persuasive arguments, who have the most charismatic personalities, or who retain the most social power. By the exercise of these forces, King Henry VIII overruled the church on matters of God's acceptance of divorce, and in turn, the church overruled thinkers of the likes of Galileo and Martin Luther on matters of both science and theology. A more recent example of the power of nonempirical forces over the voice of science is to be found in the evolution, acceptance, and reification of the *Diagnostic and Statistical Manual of Mental*

*Disorders*, 4th ed. ( [American Psychiatric Association, 1994](#) ). In the starkly political process that resulted in that document, symptoms were clustered and labeled through the power of persuasion and vote, in frank contradiction with scientific evidence indicating that neither a categorical system nor the particular symptom clusters identified represented the nature of human problems (see Beutler et al., in press; [Follette & Houts, 1996](#) ).

The proposal (e.g., [Beutler et al., 1996](#) ; [Drozd & Goldfried, 1996](#) ; [Silverman, 1996](#) ) that the mission of the field would be better served by broadening the set of criteria used to define empirical support to include quasi-experimental and naturalistic research is more defensible, logically, than the proposal to include clinical impressions and beliefs. In brief, these latter proposals are based on the following arguments: (a) Patient diagnosis is a questionable differential treatment predictor and its importance is overemphasized in the Task Force report; (b) the manualized therapies used in RCT are not representative of the treatment strategies, integration of theories, and breadth of procedures that are most frequently used in practice; (c) the patient samples used in the methodologies favored by the Task Force are not representative of those patients who most frequently seek and receive mental health services; and (d) the therapists represented in these studies are not representative of those in professional practice.

Notably, [Chambless and Hollon \(1998\)](#) acceded to the suggestion (e.g., [Beutler et al., 1996](#) ) that criteria be broadened to include nondiagnostic patient dimensions that may interact with attributes of therapy. They were less responsive to the suggestions that the criteria include well-controlled effectiveness studies as well as efficacy studies ( [Seligman, 1995](#) ; [Silverman, 1996](#) ) and that naturalistic designs of therapist effects be considered as offering evidence of effectiveness ( [Howard et al., 1996](#) ; [Silverman, 1996](#) ). Acknowledging that some of these directions may be appropriate, however, the Task Force ( [Chambless et al., 1996](#) ) is still holding in reserve the possibility that dimensional aspects of treatment may eventually supplement the criteria.

The Task Force report also stimulated a concern with building increased flexibility into the treatment manuals used in research. Inspecting the variations that are observed within therapies and therapists has been used increasingly in the development of flexible treatment guidelines to direct the use of treatment manuals ( [Beutler, 1996](#) ; [Mahrer, 1995](#) ; [Norcross, 1995](#) ). The question remains, however, as to whether broadening the criteria in these ways will substantially change the results presented by the original Task Force.

## **Would the Selection of Broad Criteria Result in Different Conclusions?**

In hallway discussions, I have frequently heard it said that the Division 12 Task Force was designed to disenfranchise those who adopted an orientation to psychotherapy that is not behavioral or cognitive—behavioral. If the adoption of broadened criteria would result in conclusions that were more favorable to the approaches that are valued by practitioners, some credence may be lent to these claims. On the other hand, if adopting broadened and clinically representative criteria does not substantially alter



the conclusions, or even makes them less palatable to nonacademics, cause for concern with the constitution of the Task Force would, logically, be attenuated. It is noteworthy that, as the Task Force criteria have become more liberal along the lines suggested by [Chambless and Hollon \(1998\)](#), a wider array of treatments are being included in the efficacious and probably efficacious lists. However, it is uncertain if this is a result of the altered criteria or of the accumulating work of reviewing research in the area over time.

In an effort to address this question, [Beutler et al. \(1996\)](#) reviewed summaries of research results and meta-analyses that incorporated RCT and quasi-experimental studies to determine the degree to which conclusions based on these latter reviews reached different conclusions than those presented by the Task Force. [Beutler et al. \(1996\)](#) reached the following conclusion: "It is doubtful that changing the criteria of evidence would make a substantial change in the results" (p. 201). They also considered several specific conclusions. The ones that are relevant to the questions addressed here are paraphrased below compared with the Task Force criteria:

1. Conclusions from broadened criteria would provide stronger evidence of the nonspecific effectiveness of psychotherapy, especially long-term treatment, but would offer fewer conclusions about the effectiveness of specific approaches (e.g., [Howard et al., 1996](#); [Lambert, Shapiro, & Bergin, 1986](#); [M. L. Smith, Glass, & Miller, 1980](#)).
2. Cognitive therapy would still predominate as among the strongest treatments, but therapist factors would be given relatively more weight than they are in the Task Force report when RCT methods are the dominant design ([Beutler, Machado, & Neufeldt, 1994](#); [Lambert et al., 1986](#)).
3. Whereas advanced training has clearly been found to be necessary to develop proficiency in the use of manualized treatments ([Jacobson, 1995](#); also, see the comment by [Chambless & Hollon, 1998](#)), the question of whether expert and advanced training is needed to practice psychotherapy would go unresolved with the inclusion of quasi-experimental and naturalistic studies. Although there are some notable suggestions that training and experience are helpful for long-term and difficult patients, for most patients, the results would probably favor lesser trained individuals and may even lend themselves to the conclusion that intensive supervision, as done in clinical practice, has failed to demonstrate its effectiveness as a training method ([Beutler, 1995](#); [Holloway & Neufeldt, 1995](#); [Stein & Lambert, 1995](#)).
4. Psychoanalytic therapies would continue to make a poor showing if quasi-experimental designs were included among the criteria. It may even be concluded that sometimes those who get this treatment become systematically worse off than when they started (e.g., [Svartberg & Stiles, 1991](#)).

As suggested by these conclusions, it may be that broadening the criteria would result in conclusions that are even less favorable to those methods that are valued by clinical practitioners than the conclusions reached in the two Task Force reports, to date. The modified conclusions might provide support to MHC agencies to continue considering all psychotherapy and psychotherapists as equivalent and

interchangeable. Although the question would remain unresolved, little support would be generated for continuing to train psychotherapists at the doctoral level. The modified conclusions would also support arguments against the support of psychodynamic psychotherapies.

On a positive note, the modified conclusions might encourage the support of longer term therapy than is currently in vogue in MHC (e.g., [Howard, Kopta, Krause, & Orlinsky, 1986](#) ; [Seligman, 1995](#) ). One must ask, however, if this single advantage justifies liberalizing the criteria.

## Conclusions

I have argued for three conclusions. The first conclusion is that the political environment in which we now live demands that the profession take the initiative to specify the nature and limits of effective treatment. The penalty for failing to initiate such actions may well be the demise of specialized and advanced training in psychotherapy.

The second conclusion advanced here is that those criteria selected by the Task Force, and even subsequently by [Chambless and Hollon \(1998\)](#) , are reasonable but probably not optimal ways of responding to this mandate. The criteria had precedence in FDA considerations of drug treatments and represented an advance over the politicized methods of selecting a consensus panel and subjecting such guidelines to popular vote and persuasive appeal.

Finally, the arguments presented here warrant the conclusion that the use of a broader set of criteria than that either used by the Division 12 Task Force or suggested by [Chambless and Hollon \(1998\)](#) would likely result in a number of less palatable conclusions. Whereas the modified conclusions may be more supportive of long-term treatment than the old ones, they would also be at least as lacking in support of psychoanalytic therapy and of doctoral training and experience as the extant ones. The relative emphasis of cognitive therapies over psychodynamic ones, which has been of concern to many clinicians, would not be corrected by expanding the criteria set.

The [Task Force \(1995\)](#) took a bold initiative. In a day of being politically correct, it proposed the unheard-of conclusion that some specific psychotherapeutic approaches are effective. The Task Force conclusions were bolder because they were unlikely to be ones that would have been accepted if their adoption had depended on membership vote. Electing not to rely on a consensus panel or to subject the findings to a vote of the membership departed from the procedure most often used to develop diagnoses and treatment guidelines, as well as from the time-honored process by which major proactive societies had established the nature of truth from long before the advent of Martin Luther.

The strategy undertaken by Division 12 was to select and publish a finite and incomplete list of those treatments, and the patient-based conditions for which they were demonstrably beneficial, that met a predetermined standard of effectiveness. This tactic sought to identify the state of knowledge within the limits imposed by the criteria selected and to serve as a guide to training and treatment selection. However, as many critiques have argued, identifying anything less than a completely exhaustive and

comprehensive list of the treatments that are practiced, runs the risk of serving as an invitation to third-party payers to restrict covered services to those treatments that are listed (e.g., [Silverman, 1996](#)). In a scenario that has been passed around in the hallways of influence, externally imposed restrictions arising from these recommendations would ensure that no new knowledge or treatments would ever evolve to see the light of day. This is a serious concern, although the Task Force ([Chambless et al., 1996](#)) and its advocates ([Chambless & Hollon, 1998](#)) have gone to great lengths to provide disclaimers. Although serious, the problem is not the Task Force report but the health care industry. Therefore, there may be no alternative that is more acceptable than some listing of empirically supported treatments.

As it stands, health care agencies are restricting service requests for far more capricious reasons without the benefit of empirical research at all. Would it really be preferable that the decisions be made in the absence of any semblance of science? I think not.

There are some things that might be accomplished to dilute this latter effect without compromising the rigor of criteria suggested by [Chambless and Hollon \(1998\)](#). One of the care-limiting strategies used in MHC agencies rests on the assumption that pharmacotherapy is cheaper and at least as effective as psychotherapy for the treatment of most conditions. As observed, this assumption is not consistent with current research ([Barlow, 1994](#); [Steinbrueck et al., 1983](#)). The extant Task Force criteria is silent on this point and could be broadened to consider research that compares psychotherapy to pharmacotherapy or combinations of psychotherapy and pharmacotherapy, to some great advantage.

In the final analysis, the question of whether the societal demands and pressures present at the time, or even now, justify a listing of empirically supported treatments, may depend on the probable consequences of taking the various alternative courses of action. One alternative to the tactic used by the Task Force would be to construct a list that included all of the treatments that are used among the community of practitioners. Selecting this option would overcome the exclusiveness of the Task Force strategy because all psychotherapists could be safe in the assurance that their pet methods were recommended. However, this approach would mean abandoning the scientific basis of evidence and, on its face, would have ensured the loss of whatever credibility that psychotherapy and psychotherapy research may have accrued among the general public. Moreover, this tactic may have opened the door to litigation if any of the treatments were recommended, despite evidence that they were sometimes dangerous. Finally, this strategy may have subjected the Task Force to intensive lobbying efforts that may have required adding a host of unacceptable treatments without benefit of scrutiny.

A variant of this approach would be to list all treatments for which there is not persuasive evidence of harmfulness. This task would be extremely complex and difficult, probably surpassing that of assigning effectiveness. There is little research that addresses the issue of harm, and the nature of harm is even more difficult to define than the nature of benefit. It is not certain that liability would be significantly attenuated by such a tactic.

The other alternative would be to resist the pressures to construct such a list altogether and to adopt the view advocated by many (e.g., [Drozd & Goldfried, 1996](#); [Garfield, 1996](#)) that such a listing is premature.

This option would eliminate the danger of litigation and would avoid the intraprofessional conflict that has followed the Task Force report. Recommending a delay would reinforce the assumption, already made by many health care and political bodies, that effectiveness can be equated with the cost and popularity of the treatment.

This stance may also have resulted in a substantial loss of credibility among policymakers and health care colleagues, alike. With now well over 1,000 well-conceived studies available ( [Bergin & Garfield, 1994](#) ; [Orlinsky, Grawe, & Parks, 1994](#) ), knowledgeable voices that are reluctant to make judgments may well convey the impression that practitioners know of hidden dangers that are not visible to those outside the field. Certainly, it is reasonable for a politician or health care manager to ask what is to be concluded from the 1,000-plus studies that have been conducted, if not to say what treatments are effective. One may weakly reply that these studies demonstrate that psychotherapy is effective, that kind therapists are better than unkind ones, and that relationships are important ( [Lambert, 1991](#) ). To suggest that these are the only conclusions warranted after an accumulated expenditure of the millions of tax dollars committed to research over the past 40 years, is decidedly uninteresting. Moreover, it understates the facts.

Reluctance of identified experts to draw more exact conclusions than the foregoing from the extant body of research, while continuing to promote and charge for the practice of specific forms of psychotherapy, may have struck some observers as a peculiar contradiction. Delaying in the face of pressing social needs would likely convey the impression that professional psychotherapists, themselves, lack confidence either in those who carry out contemporary research or in those who practice psychotherapy as a profession. Requiring more evidence, when so much research is already available, may well lead to the withdrawal of funding both for the practice and research of psychotherapy.

If, as one might conclude from the failure to take a stand, all of the 400-plus varieties of psychotherapy are equally effective for all people and all problems, then of what value are our training programs? Under these circumstances, it is understandable that the public would choose to withdraw support from the education and training of practitioners and to turn the work of being psychotherapists over to those who are the cheapest to train and who cost the least to hire.

Personally, I am much more optimistic about what research shows. I believe that there is ample evidence to suggest that there are limiting and facilitating conditions that stimulate the benefit of most psychotherapies. Taking the initiative to identify treatments that have and have not been supported by empirical research, as the Division 12 Task Force has done, may well force those who favor and practice psychotherapy to give up favored opinions, practices, and theories. But, it may also open the door to those practitioners who are willing to accept the responsibility of bridging science and practice, a credo to which those in the field have long paid lip service. As long as such task forces and the reports they produce can be seen as dynamic rather than static processes, the nature of the criteria by which treatments will be judged will continue to improve and expand, and research will continue to offer the hope of facilitating and refining the practice of psychotherapy. The payoff for initially living with processes and recommendations that are considered to be less than optimal may well be increased credibility and expanded opportunities to practice and refine psychotherapy.

## References

- Aaron, H. (1996). End of an era.( *The Brookings Review*, 14(1), 35—37.)
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.).(Washington DC: Author)
- Barlow, D. H. (1994). Psychological interventions in the era of managed competition.( *Clinical Psychology: Science and Practice*, 1, 109—122.)
- Bergin, A. E. (Ed.) & Garfield, S. L. (Ed.) (1994). *Handbook of psychotherapy and behavior change* (4th ed.).(New York: Wiley)
- Beutler, L. E. (1995). The germ theory myth and the myth of outcome homogeneity.( *Psychotherapy*, 32, 489—494.)
- Beutler, L. E. (1996, August). *Manualizing flexibility: The training of eclectic therapists*. (Paper presented at the 104th Annual Convention of the American Psychological Association, Toronto, Ontario, Canada)
- Beutler, L. E., Bongar, B. & Shurkin, J. C. (in press). *Am I crazy or is it my shrink?* (New York: Oxford University Press)
- Beutler, L. E. & Davison, E. H. (1995). What standards should we use?(In S. C. Hayes, V. M. Follette, R. M. Dawes, & K. E. Grady (Eds.), *Scientific standards in psychological practice: Issues and recommendations* (pp. 11—24). Reno, NV: Context Press.)
- Beutler, L. E., Kim, E. J., Davison, E., Karno, M. & Fisher, D. (1996). Research contributions to improving managed health care outcomes.( *Psychotherapy*, 33, 197—206.)
- Beutler, L. E., Machado, P. P. P. & Neufeldt, S. (1994). Therapist variables.(In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 259—269). New York: Wiley.)
- Chambless, D. L. (1996). In defense of dissemination of empirically supported psychological interventions.( *Clinical Psychology: Science and Practice*, 3, 230—235.)
- Chambless, D. L. & Hollon, S. D. (1998). Defining empirically supported therapies.( *Journal of Consulting and Clinical Psychology*, 66, 7—18.)
- Chambless, D. L., Sanderson, W. C., Shoham, V., Johnson, S. B., Pope, K. S., Crits-Christoph, P., Baker, M., Johnson, B., Woody, S. R., Sue, S., Beutler, L. E., Williams, D. A. & McCurry, S. (1996). An update on empirically validated therapies.( *The Clinical Psychologist*, 49, (2), 5—14.)
- Chapman, L. J. & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs.( *Journal of Abnormal Psychology*, 74, 193—204.)
- Drozd, J. F. & Goldfried, M. R. (1996). A critical evaluation of the state-of-the-art in psychotherapy outcome research.( *Psychotherapy*, 33, 171—180.)
- Evans, M. D., Hollon, S. D., DeRubeis, R. J., Piasecki, J. M., Grove, W. M., Garvey, M. J. & Tuason, V. B. (1992). Differential relapse following cognitive therapy and pharmacotherapy for depression.( *Archives of General Psychiatry*, 49, 802—808.)
- Fay, A. & Lazarus, A. A. (1993). On necessity and sufficiency in psychotherapy.( *Psychotherapy in Private Practice*, 12, 33—39.)
- Follette, W. C. & Houts, A. C. (1996). Models of scientific progress and the role of theory in taxonomy development: A case study of the DSM. ( *Journal of Consulting and Clinical Psychology*, 64, 1120—1132.)

- Garfield, S. L. (1996). Some problems associated with "validated" forms of psychotherapy.( *Clinical Psychology: Science and Practice*, 3, 218—229.)
- Hollon, S. D., DeRubeis, R. J., Evans, M. D., Wiemer, M. J., Garvey, M. J., Grove, W. M. & Tuason, V. B. (1992). Cognitive therapy and pharmacotherapy for depression: Singly and in combination.( *Archives of General Psychiatry*, 49, 774—781.)
- Holloway, E. L. & Neufeldt, S. A. (1995). Supervision: Its contributions to treatment efficacy.( *Journal of Consulting and Clinical Psychology*, 63, 207—213.)
- Houts, A. C. & Graham, K. (1986). Can religion make you crazy?(Impact of client and therapist religious values on clinical judgments. *Journal of Consulting and Clinical Psychology*, 54, 267—271.)
- 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., Moras, K., Brill, P. L., Martinovich, Z. & Lutz, W. (1996). The evaluation of psychotherapy: Efficacy, effectiveness, patient progress.( *American Psychologist*, 10, 1059—1064.)
- Jacobson, N. S. (1995, March). The overselling of therapy.( *Networker*, pp. 41—47.)
- Lambert, M. J. (1991). An introduction to psychotherapy research.(In L. E. Beutler & M. Crago (Eds.), *Psychotherapy research: An international review of programatic studies* (pp. 1—11). Washington, DC: American Psychological Association.)
- Lambert, M. J., Shapiro, D. A. & Bergin, A. E. (1986). The effectiveness of psychotherapy.(In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change*. (3rd ed., pp. 157—211). New York: Wiley.)
- Mahrer, A. R. (1995). An introduction to some disposable myths, how to detect them, and a short list.( *Psychotherapy*, 32, 484—488.)
- Norcross, J. C. (1995). Dispelling the Dodo Bird verdict and the exclusivity myth in psychotherapy.( *Psychotherapy*, 32, 500—504.)
- Orlinsky, D. E., Grawe, K. & Parks, B. K. (1994). Process and outcome in psychotherapy—Noch Einmal.(In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 270—376). New York: Wiley.)
- Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The *Consumer Reports* study.( *American Psychologist*, 50, 965—974.)
- Silverman, W. H. (1996). Cookbooks, manuals, and paint-by-numbers psychotherapy in the 90's.( *Psychotherapy*, 33, 207—215.)
- Smith, E. W. L. (1995). A passionate, rational response to the "manualization" of psychotherapy.( *Psychotherapy Bulletin*, 30(2), 36—40.)
- Smith, M. L., Glass, G. V. & Miller, T. I. (1980). *The benefits of psychotherapy*. (Baltimore: Johns Hopkins University Press)
- Sperry, L., Brill, P. L., Howard, K. I. & Grissom, G. R. (1996). *Treatment outcomes in psychotherapy and psychiatric interventions*. (New York: Brunner/Mazel)
- Stein, D. M. & Lambert, M. J. (1995). Graduate training in psychotherapy: Are therapy outcomes enhanced?( *Journal of Consulting and Clinical Psychology*, 63, 182—196.)
- Steinbrueck, S. M., Maxwell, S. E. & Howard, G. S. (1983). A meta-analysis of psychotherapy and drug therapy in the treatment of unipolar depression with adults.( *Journal of Consulting and Clinical Psychology*, 51, 856—863.)
- Svartberg, M. & Stiles, T. C. (1991). Comparative effects of short-term psychodynamic psychotherapy: A

meta-analysis.( *Journal of Consulting and Clinical Psychology*, 59, 704—714.)

Task Force on Promotion and Dissemination of Psychological Procedures (1995). Training in and dissemination of empirically-validated treatments: Report and recommendations.( *The Clinical Psychologist*, 48 (1), 3—23.)

## 1

This movement toward an empirical standard clearly preceded the release of the first Division 12 Task Force report, although it is likely that some will and do blame this development on that latter report.