How Would We Know If Psychotherapy Were Harmful?

Sona Dimidjian *University of Colorado*
Steven D. Hollon *Vanderbilt University*

Patients can be harmed by treatment or by the decisions that are made about those treatments. Although dramatic examples of harmful effects of psychotherapy have been reported, the full scope of the problem remains unclear. The field currently lacks consensus about how to detect harm and what to do about it when it occurs. In this article, we define the ways in which treatment (or the inferences about treatment) can do harm and discuss factors that complicate efforts to detect harm. We also recommend methods to detect and understand harm when it occurs, drawing from and modifying many of the same strategies that are used to detect benefit. Specifically, we highlight the value of establishing independent systems for monitoring untoward events in clinical practice, reporting descriptive case studies and qualitative research, and making use of information from randomized clinical trials, including examining potential active ingredients, mechanisms, moderators, and a broad range of outcomes measured over time. We also highlight the value of promoting discussion in the field about standards for defining and identifying harm.

Keywords: harm, psychotherapy, methods, adverse effects, patient safety

If psychotherapy is powerful enough to do good, it may be powerful enough to do harm. Although it is now well established that most psychotherapies provide benefit, recent discussions have raised concerns about examples of psychotherapies that may do harm (Lilienfeld, 2007). It is still unclear whether such harm is more widespread than is currently recognized and how psychologists would know if that were the case. With few exceptions, psychologists have been primarily concerned with determining whether psychotherapy has benefit. Such an emphasis is necessary and important; however, failure to detect harm can have serious consequences.

This article represents an attempt to address this imbalance by defining harm and identifying ways in which harm can be detected and understood. Attempts to classify specific treatments as harmful have generated considerable controversy, and we encourage such debate in the field. Progress in identifying harmful treatments likely will evolve from a combination of rigorous empirical research and lively discussion in the field about optimal conventions and standards. We seek to contribute to this endeavor by outlining methods for research that can address the question of harm. We also identify questions and processes that may advance consensus building in the field. Thus, our goal here is not to identify specific treatments as harmful or helpful but to lay out a method for detecting such effects. We consider psychotherapy to be broadly defined as any psychosocial intervention intended to aid a client with mental health or life problems. We first define the ways that treatments can cause harm directly and the ways in which decisions made about treatments can cause harm indirectly. We then examine the conceptual complexities involved in the detection of such harm. We next recommend methods for detecting harm, drawing on what has been learned from the science and practice of both psychotherapy and medicine. Finally, we close with a discussion of some strategies that can be used to understand harm when it occurs.

What Do We Mean by Harm?

Patients can be harmed by treatments themselves or by the decisions that are made about those treatments. We first discuss ways in which treatments directly cause harm and then turn our attention to the indirect harm that may stem from inferences made about treatments. Harmful treatments have a causal effect, producing outcomes that are worse than they would have been in the absence of treatment. We underscore that harmful effects are more than simply unhelpful. As defined by *Webster's Third New International Dictionary* (1993), *unhelpful* is "offering no assistance" (p. 1034), whereas *harmful* is "damaging, troublesome, injurious" (p. 2498).

There are many examples of treatments that fail to make things better and of the failure to provide treatments that might have made things better. It is widely accepted that poor detection or undertreatment of depression is a widespread problem. However, failing to provide interventions that may provide relief is not the same as providing interventions that cause harm. The latter is our focus. Related to this are many examples of treatments that are unhelpful (e.g., a depressed suicidal patient who commits

Sona Dimidjian, Department of Psychology and Neuroscience, University of Colorado; Steven D. Hollon, Department of Psychology, Vanderbilt University.

We are grateful to Scott Lilienfeld for helpful comments on a draft of this article and to Marsha Linehan for many valuable conversations about psychcotherapy practice and research.

Correspondence concerning this article should be addressed to Sona Dimidjian, Department of Psychology and Neuroscience, University of Colorado, 345 Muenzinger, Office 313B, Boulder, CO 80309-0345. E-mail: sona.dimidjian@colorado.edu

 0003-066X/10/$12.00
Vol. 65, No. 1, 21–33 DOI: 10.1037/a0017299

Sona Dimidjian

suicide despite being in treatment). In this case, treatment clearly was not helpful; however, such a clearly negative outcome does not by definition imply that treatment was harmful. Finally, we recognize a distinction between harm that can be caused by a disorder and harm that can be caused by the application of a treatment. Many disorders exist that clearly are associated with harmful consequences. People with antisocial personality disorder or psychotic disorders may commit violent acts, and depressed people may commit suicide; however, our focus is harm that results from treatment, not harm that is an inherent feature of the disorder itself.

Thus, what is at question here is whether a treatment makes things worse. There are at least two ways in which a treatment can be harmful. A treatment can make the target problem worse. A treatment also can make worse other domains that were not the initial target of intervention. The matter is further complicated in five ways.

First, the same treatment can have both beneficial and harmful effects. Use of some selective serotonin reuptake inhibitors as antidepressant medications among adolescents has been associated with both improvement in depression and possible increased risk for suicide (Whittington et al., 2004). Similarly, perinatal women and their treatment providers often are presented with complicated cost–benefit decisions regarding treatment options during pregnancy and lactation. As Wisner et al. (2000) explained,

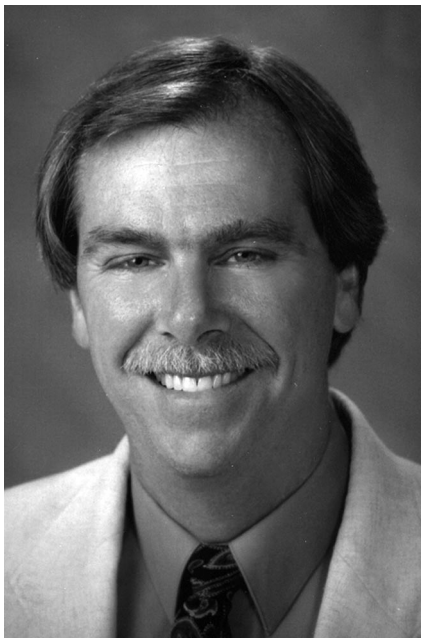
> There are unique aspects to this decision-making process that involve potential consequences for two individuals. The mother has to consider the risk information not only for herself, but also for her fetus. The outcomes can be positive for both, negative for one but positive for the other, or negative for both. (p. 1936)

Thus, considering the maternal/fetal system, the same intervention may have effects that are both positive (e.g., resolution of maternal symptoms) and harmful (e.g., fetal toxicity). Although these examples are taken from pharmacological interventions, the same logic applies to psychotherapy.

Second, different parties can view the same outcome in different ways (Strupp & Hadley, 1977). For example, ending a troubled marriage following marital therapy may be viewed as beneficial by one spouse and as harmful by the other. Disagreement about the outcome of clinical intervention also may occur in professional and political domains. The use of conversion therapy for gay and lesbian clients provides a controversial example. Large professional organizations and individual experts suggest that such practices are harmful not only to the individual client, but also to society as a whole, because they reinforce stigma and prejudice (Davidson, 2001; Haldeman, 1994). The American Psychiatric Association, for example, strongly supports this position, stating in a position paper on psychiatric treatment and sexual orientation that "the potential risks of reparative therapy are great, including depression, anxiety, and self-destructive behavior, since therapist alignment with societal prejudices against homosexuality may reinforce self-hatred already experienced by the patient" (American Psychiatric Association, 2000b). In contrast, some claim that the use of conversion therapy is a legitimate, potentially beneficial treatment option for some gay and lesbian clients (Throckmorton, 2002; Yarhouse & Burkett, 2002). Such controversial opinions underscore the reality that different parties may view the same outcome in radically different ways.

Third, the nature of the outcomes can change over time. An intervention that seems harmful initially may be beneficial over time, and vice versa. For example, exposure therapy is often thought to lead to increased levels of distress during its implementation, an issue that has been raised as a cause for concern regarding the use of exposure for posttraumatic stress disorder. Empirical data suggest that exposure is associated with symptom exacerbation for some patients. A small minority of patients evidenced reliable increases in general anxiety in one trial (Foa, Zoellner, Feeny, Hembree, & Alvarez-Conrad, 2002) and exacerbation of posttraumatic stress disorder symptoms in another (Nishith, Resick, & Griffin, 2002). Such deterioration, however, appears to be transient and is not associated with higher rates of dropout or nonresponse (Foa et al., 2002). Thus, although some patients may experience worsening at the outset of exposure treatment, such patients appear to improve over time and experience long-term resolution of the target problem.

Conversely, the current controversy regarding grief therapy represents an example in which questions are raised about later harm produced by an intervention that seemed initially beneficial. Concerns about grief therapy were raised in a review by Neimeyer (2000), in which he suggested that grief therapy may interfere with the normal process of recovering from loss over time. Recent reports have questioned the methods used in Neimeyer's review (Larson & Hoyt, 2007), and active debate continues regarding the status of grief therapy (Bonanno & Lilienfeld, 2008;



Steven D. Hollon

Hoyt & Larson, 2008). Perhaps a less controversial illustration is the inducement of tardive dyskinesia as a consequence of treating patients with schizophrenia with antipsychotic medication. This well-known example illustrates a harmful effect of a beneficial treatment that emerges only over time.

Fourth, a treatment can be universally harmful in its effects, or such effects can be moderated by patient characteristics. For example, there are indications that relaxation training can induce panic attacks among a minority of patients (Adler, Craske, & Barlow, 1987). Thus, whereas relaxation has been generally beneficial for most patients, it may be harmful for a few.

Fifth, ambiguity regarding what is meant by treatment complicates efforts to define harmful effects. The same psychotherapy can be applied in a range of different ways, depending, for instance, on the content of the intervention and the skill level of the practitioner. One manner of delivery of a treatment may be helpful, whereas another may be harmful. Given the multifaceted nature of most psychological treatments and the possible variability in delivery, it is important to specify at a more nuanced level the specific ingredients of the approach. Such considerations also highlight the potential importance of clinician skill. An unskilled surgeon may kill a patient, whereas a skilled surgeon may save someone's life. Whether this holds true with respect to psychotherapy is controversial (Jacobson & Christensen, 1996), but there are indications that expertise may influence outcome, at least in cognitive therapy for depression (Coffman, Martell, Dimidjian, Gallop, & Hollon, 2007; DeRubeis et al., 2005; Jacobson & Hollon, 1996). Although such variability would not necessarily meet standards for harm, the sensitivity of outcome to clinician expertise suggests the possibility of harm resulting from the incompetent application of a treatment.

Finally, although it clearly is possible for patients to be harmed directly by psychotherapy, patients also can be harmed by the decisions that are made about psychotherapy. The field has paid inadequate attention to direct harmful effects (Lilienfeld, 2007), and it has also ignored the indirect harm that can result from inaccurate inferences drawn about treatment efficacy. An inert treatment that is falsely assumed to be beneficial can be costly in terms of time, expense, and other resources. A beneficial treatment that is falsely assumed to be inert or worse can result in opportunities lost. These instances do not represent causal effects of the treatment but are consequences of erroneous inference regarding treatment efficacy and merit consideration in a discussion of harm.

Efforts to develop consensus in the field regarding how to define harm undoubtedly will require consideration of the following factors: iatrogenic effects on the target problem or other domains, the simultaneous presence of beneficial and harmful effects, variability in the perspectives of different parties, the dimension of time, moderating patient characteristics, and a detailed understanding of the intervention. Moreover, it is necessary to consider carefully the basis on which inferences about benefit and harm are drawn.

What Complicates the Detection of Harm?

Detecting whether treatments cause harm is a complicated endeavor. It is easy to overlook the harmful effects of some treatment interventions, to attribute falsely harmful or beneficial effects to others, and to neglect the beneficial effects of still others. In this section, we attempt to clarify the conceptual complexities associated with detecting treatment effects that have beset the field to date.

The detection of whether a treatment can cause harm may be complicated by the natural course of the untreated disorder (*X*). Table 1 depicts the relations between three kinds of course for a disorder (constant, deteriorating, and improving) and three levels of treatment effects (harmful, beneficial, and absent). We also have added a table note to depict the possible effects of treatment on outcomes other than the target problem (*Y*).

Harmful treatment effects are most easily detected in the context of courses of illness that are constant (No. 1 in Table 1) or deteriorating (No. 2). Disorders with a constant course do not change over time and provide a backdrop against which changes induced by treatment should be relatively easy to detect; harmful treatments leave the patients worse than when they started. For disorders that follow a deteriorating course, it is not sufficient to observe that patients are worse off after treatment than when they started, because change for the worse could have been a consequence of natural processes that would have unfolded independent of treatment. For such a course, it must be observed that patients are worse off than they would have been in the absence of treatment.

Harmful effects are particularly difficult to detect when the disorder improves over time (spontaneous remis-

Table 1
Harmful Effects and Harmful Inferential Errors: Treatment Effect by Natural Course of Target Problem

Natural course	Harmful treatment effect	No treatment effect	Beneficial treatment effect
Constant course of target problem	1. Worse than started	4. Same as started	7. Better than started
Deteriorating course of target problem	2. Worse than started	5. Worse than started (might falsely assume iatrogenic effect)	8. Same, better, or worse than started (might miss the beneficial effect of treatment)
Spontaneous remission of target problem	3. Same, better, or worse than started (might miss the harmful effect of treatment)	6. Better than started (might falsely assume beneficial effect of treatment)	9. Better than started

Note. In some cases, a new problem is created and patients are worse off than when they started treatment.

sion; No. 3). In such a situation, not only is the treatment actively harmful, but worse still, the fact that patients improve as a matter of course can serve to mask the harmful effects. Thus, the inferential error compounds the problem by allowing a harmful treatment to escape detection. In such cases, the intervention itself may interfere with normal developmental processes that can facilitate recovery from the target problem among some individuals. For example, it has been suggested that critical incident stress debriefing, a brief group intervention performed within a day or two of a traumatic incident, in which participants are encouraged to process their negative emotions and are educated about the nature of the symptoms that might develop, can actually impede the process of spontaneous remission that occurs for many trauma survivors (McNally, Bryant, & Ehlers, 2003). Additionally, Dishion, McCord, and Poulin (1999) have argued that the group treatment format provides a context for delinquency training that perpetuates problematic behaviors in adolescents who might otherwise outgrow such behaviors; however, subsequent reports have not provided support for harmful effects of such group interventions (Weiss et al., 2005).

Inferential errors can be harmful in and of themselves, even if the treatments themselves are not. This can be true in the case of inert treatments that are seen as being beneficial (No. 6) and beneficial treatments that are seen as inert or harmful (No. 8). Instituting treatment at the height of a disorder with an improving baseline can easily lead to the potentially unwarranted inference that subsequent improvement was a consequence of that intervention. If an inert treatment is inaccurately assumed to be beneficial, it still may be costly in terms of time, expense, and other resources. Although such harm is not a causal effect of the treatment, it is a consequence of making an inaccurate inference about treatment efficacy. That many patients survived and thrived in the past after being treated with bleeding undoubtedly led to an unjustified faith in its effects. Similarly, recent studies with sham procedures as controls have shown that the purported benefits of orthopedic surgery have been greatly overrated (Moseley et al., 2002).

Conversely, beneficial effects can be hard to detect in the context of a deteriorating course, because patients may be worse off than when they started, and yet they may still have benefitted from treatment (cell 8). A classic example can be found from the field of early childhood education. The Head Start program was nearly abandoned in its early years because children in the program continued to fall behind their peers. It was not until it was determined that they lost less ground than they would have lost if they had not received the program that its future was secured (Oden, Schweinhart, & Weikart, 2000). Strictly speaking, this is not an iatrogenic effect of treatment, but it may be a harmful consequence of drawing inaccurate inferences about the efficacy of treatments. Similarly, an inert treatment may be seen as being harmful in the context of a deteriorating course (No. 5), but this inferential error has little cost. Finally, treatments that have beneficial effects on a constant (No. 7) or improving course (No. 9) or inert treatments that have no effects on constant course (No. 4) are not likely to be subject to inferential error.

As indicated in the note to Table 1, problems that were not the initial target can develop or worsen as a consequence of treatment. Although the effect of the treatment on the target problem may range from harmful to beneficial, nontarget problems all share the feature that harm is caused by the creation of new problems that would not otherwise have occurred or the exacerbation of problems that already exist. In some instances, such harmful effects are clear cut. The suffocation of children during rebirthing therapy is a clear and obvious example of harm caused that was not an initial target of intervention (Mercer, Sarner, & Rosa, 2003). Not everyone would agree that rebirthing therapy is a form of psychotherapy, but all would agree that the death of a child is a form of harm. In other instances, however, such harm can be difficult to detect, particularly if attention is focused exclusively on the target problem or

if the effects of treatment on the target problem are beneficial. Electroconvulsive therapy is a powerful treatment for severe depression (Sackeim et al., 2000), but some patients complain of persistent memory deficits as a consequence (Hollon, Thase, & Markowitz, 2002). The general consensus among the medical community is that these concerns are a natural consequence of being depressed and not an iatrogenic effect of treatment (American Psychiatric Association, 2000a); however, concerns about persistent memory deficits have not been adequately addressed by the data. There may be a risk on the part of the medical community to overlook or minimize potentially harmful effects of interventions that are known to be beneficial, particularly for difficult-to-treat target problems. This example highlights the controversies that can arise regarding the detection of harm alleged to be associated with beneficial treatments.

As can be seen, the ease with which harmful treatment effects can be detected and the kinds of inferential errors most likely to be made depend, in part, on the natural course of the disorder. Harm is easiest to detect when disorders show a constant course because any change can be attributed to treatment. Harm is more likely to be missed when the natural course of the disorder changes over time. This is especially problematic for disorders that tend to improve over time, because the propensity for spontaneous remission can obscure the negative effects of treatment. Moreover, although psychologists have some knowledge of the natural course of many disorders, within any disorder there is considerable variability in natural course across given individuals.

Patients also can be harmed as a result of inferential errors, even when treatments themselves are not harmful. Inert treatments can be misperceived as beneficial in the context of spontaneously remitting problems, and beneficial treatments can be missed in the context of a deteriorating course. Finally, it may be all too easy to miss treatments that do harm by creating new problems or exacerbating problems that already exist, if vigilance is not exercised. We turn now to a discussion of methods that can be used to minimize these risks.

What Methods Aid the Detection of Harm?

The same methods that are used to identify when a treatment is beneficial can be used to determine when it is harmful. Many of these methods are designed to offset the complications previously discussed. Just as empirical inquiry may proceed from anecdotal reports of favorable outcomes to descriptive case studies to controlled clinical trials to determine benefit, many of the same strategies can be used to detect possible harm. In this section, we recommend methods that can be used to aid the detection and explanation of harm: (a) develop systems for monitoring untoward events, (b) report descriptive case studies and qualitative research, (c) make use of randomized controlled trials, (d) examine a broad range of outcomes over time, (e) examine the active ingredients of treatments that cause harm, (f) examine the mechanisms by which harm is produced, (g) examine whether harm is universal or moderated, and (h) promote discussion about standards and replication.

Develop Systems for Monitoring Untoward Events

Anecdotal evidence represents the first line of defense but, by its very nature, tends to be unsystematic. Reports of death due to suffocation during rebirthing therapy were sufficient to raise concerns in the minds of many. Suffocation is not a common outcome of psychotherapy, and even rare instances that can plausibly be attributed to the treatment are a cause for concern. The problem with anecdotal information is that it needs to be noticed and reported. Unusual or unexpected outcomes may not be attributed to an intervention (even if they should). Practitioners of an approach may have a vested interest in not entertaining the possibility of a connection or even reporting an event. At the same time, individual outcomes are causally ambiguous. Simply because an untoward outcome occurs does not mean that the treatment (or the clinician) was the cause. Systematically monitoring such events can help to identify whether standard interventions in the field evidence harm, which might otherwise be attributed erroneously to lack of skill on the part of the individual clinician.

The Food and Drug Administration monitors unexpected outcomes by means of a postmarketing surveillance system in which doctors who prescribe a medication are encouraged to report unexpected negative outcomes. It was this reporting system that first brought to light concerns about the effects of serotonergic medications on increasing rates of violence and suicide (Breggin, 2004). Although these concerns were largely ignored by the medical community for over a decade (although not by the legal community), subsequent placebo-controlled trials have indicated that selective serotonin reuptake inhibitors might indeed increase risk for suicidal behaviors in adolescents and young adults (Whittington et al., 2004). This led the Food and Drug Administration to add a black box warning to the packaging of selective serotonin reuptake inhibitors when they are prescribed (Emslie et al., 1997; Gibbons et al., 2007; Richmond & Rosen, 2005). The wisdom of this move remains controversial, because the early indications are that suicide rates among adolescents may be going up in concert with the subsequent reduction in medication prescriptions written for this age group (Gibbons, Hur, Bhaumik, & Mann, 2006); nonetheless, it appears that postmarketing surveillance has had an effect on practice. There is no comparable mechanism for monitoring whether psychotherapy has unintended negative effects, but the detection of harmful effects (if they exist) would be facilitated by such a system.

Ideally, a monitoring system would gather information about a full range of untoward outcomes, independent of their presumed relationship to the intervention. Minimally, systematic monitoring of the target problem is essential to determine if the problem is worsening over time. The use of systematic monitoring of client outcomes has

been recommended widely to counter problems with cognitive biases, heuristics, and memory in clinical decision making (e.g., Kazdin, 2008), and such concerns are as important for considerations of harm as they are for benefit. A wide range of brief reliable and valid self-report measures for common mental health problems are available (e.g., Antony, Orsillo, & Roemer, 2007; Nezu, Ronan, Meadows, & McClure, 2000). Asking clients to complete a structured self-report measure for the target problem at the outset of each session offers an accessible, practical option for many clinicians. The use of such measures in an ongoing fashion helps to monitor whether the client's target problem has worsened. Moreover, studies suggest that monitoring status in such ways is itself associated with improvements in clinical outcomes (Lambert, Harvey, & Poland, 2007).

Clinicians also might use standard Food and Drug Administration definitions as a starting point for discussion about the range of outcomes that should be monitored; these definitions outline both types of events as well as requirements for the promptness with which they should be reported. Serious adverse events typically include death, life-threatening events, hospitalization, persistent or significant disability, and birth defects and are subject to expedited reporting. For psychotherapy, events including suicide, homicide, arrests, hospitalizations, accidents, and serious medical illnesses might well be included. Such reporting is important for all psychotherapies and is perhaps even more important for those that are novel. For example, in a study using mindfulness meditation with pregnant women at high risk of depression, which is being conducted by one of the authors (Sona Dimidjian), events such as miscarriage are tracked and reported, even though it is unlikely that such outcomes are attributable to the intervention.

Such a system could be implemented at the state or national level, and it could be implemented within existing care systems, such as clinical trials networks or large managed care organizations. Such systems already have the benefit of an infrastructure that collects patient information on a systematic basis. Whenever centralized records are kept, there is a greater chance for untoward effects to be detected. Whether those negative effects are, in fact, a consequence of the treatment provided would remain to be determined, but such a practice might at least alert the field that a problem might exist.

Finally, it is important to have independent impartial observers monitor patient safety. The oversight of individuals who have no investment in the treatment is a key element of effective monitoring systems. The treating clinician or the research investigator may have a tendency to minimize or overlook negative outcomes (Lambert et al., 2007). Even when that is not the case, responsible clinicians and investigators will welcome the notion of independent oversight for the protection of their patients.

The use of data safety and monitoring boards (DSMBs), now required by the National Institutes of Health for clinical trials, provides an informative model from which psychotherapy clinicians and researchers can draw. DSMBs are composed of individuals who have knowledge of the patient population and no vested interest in the outcome of the study. They are charged with overseeing patient safety and the responsible conduct of the study, and they have the power to stop trials if harm is evident. For example, a DSMB stopped the major study of hormone replacement therapy (widely used in the treatment of menopausal symptoms) when interim findings suggested that the intervention increased risk for cancer and heart disease (Writing Group for the Women's Health Initiative Investigators, 2002). Currently, there exists minimal direct monitoring of psychotherapy practitioners. Psychotherapists rarely are directly observed in their clinical practices, and the onus of responsibility for reporting negative events falls on the public through such routes as filing complaints with licensing boards. Knowledge of the potential harm associated with psychotherapy would be facilitated by requiring clinicians and researchers to report negative events to independent monitoring boards. Whether the outcomes reported reflected harm caused by treatment is a more complicated matter to determine, but a monitoring system would facilitate considerations of harm.

Report Descriptive Case Studies and Qualitative Research

Case studies can be an extremely valuable source of information that builds on the simple reporting of untoward events. A rich clinical narrative provides a context for exploring a possible relationship between untoward events and the interventions used. Such uncontrolled accounts cannot tell whether the untoward event was a consequence of the treatment, particularly given the variability in the natural course of disorders across different individuals. However, uncontrolled case studies help identify possible hypotheses for more systematic study and may guide the adaptation of interventions to minimize future harm. There are publication outlets for such case studies, and although few in number, there are examples in the literature (Linehan & Dexter-Mazza, 2008; Marsh & Hunsley, 1993; Strupp, 1980). Instead of burying mistakes, reporting descriptive case studies allows psychologists to learn from them.

Similarly, qualitative research also can highlight possible indications of harm and directions for future research. For instance, concerns regarding harmful effects of conversion therapy for gay and lesbian clients were highlighted in a qualitative study of experiences reported by 202 individuals who had received sexual conversion therapy interventions (Shidlo & Schroeder, 2002). Experiences of self-defined harm were identified across multiple domains by the majority of participants. Although such methods cannot determine whether harm was caused by such interventions, they can identify clearly areas of concern to be investigated in subsequent studies.

Make Use of Randomized Controlled Trials (RCTs)

RCTs provide the best means for drawing causal inferences about treatment effects. They are not necessarily the best

means to alert the field to possible negative effects (that would fall to the less systematic strategies just described) or to explain such effects (that would fall to the strategies described next), but they are the most certain way to determine whether a negative effect can be attributed to treatment, particularly given the potential variance in the natural course of the disorder previously described.

If researchers knew for sure that the course of a disorder was constant, as discussed previously, then an RCT would not be needed; it would be sufficient to observe whether change occurred over the course of treatment. The determination of whether that treatment was harmful would simply be a consequence of whether things changed for the worse. Unfortunately, psychologists rarely know the course of a disorder for a given individual and therefore must rely on RCTs or their logical equivalents. RCTs provide an opportunity to determine whether negative outcomes are more common in the presence of treatment than in the absence of treatment. Because participants are randomly assigned to the treatment conditions, causal inferences can be drawn from differences in outcome, protecting against potentially harmful inferential errors.

Seizure therapies were first developed as a treatment for schizophrenia (on the basis of the mistaken notion that epilepsy protected persons at risk from becoming psychotic), and some maintained the belief that electroconvulsive therapy should help reverse the course of the disorder as late as the 1950s. A medical researcher named Peter King reasoned that if a brief course of electroconvulsive therapy was not sufficient to cure schizophrenia, then perhaps increasing treatment frequency was required. He developed an approach called regressive electroshock therapy (REST) that induced seizures in schizophrenic patients several times a day over the course of several weeks. An RCT was conducted in which inpatients with schizophrenia were randomly assigned to have either antipsychotic medications or REST. In findings published in the *American Journal of Psychiatry*, not only was REST inferior to antipsychotic medication, it also was inferior to standard care (King, 1958). Moreover, two patients assigned to REST died during the course of treatment. This example highlights the value of RCTs in detecting harmful effects. Although REST might have been abandoned eventually on the basis of accumulating case reports of patient deaths, the fact that the RCT established its deleterious effect in comparison to standard care no doubt hastened the demise of this approach.

In addition to aiding the detection of harmful effects, RCTs also are valuable in protecting against the potential harm that can result from inferential errors regarding treatment effects. In the early 1950s, Hans Eysenck shook the professional community by purporting to show that rates of response to psychotherapy were no greater than rates of change brought about by spontaneous remission in people who did not seek therapy at all (Eysenck, 1952). The article was something of a polemic; Eysenck was an early advocate of more behavioral approaches (which he considered to be more scientific), and his critique was directed largely at the traditional dynamic approaches in vogue at the time. Nonetheless, it set off a firestorm of counterclaims purporting to demonstrate the value of psychotherapy (as it was then construed), and it served as the impetus for a considerable number of RCTs intended to test the issue in a more scientific fashion. The vast majority of these trials demonstrated that psychotherapy was more efficacious than its absence (Luborsky, Singer, & Luborsky, 1975; M. L. Smith & Glass, 1977). Although some have argued against the use of RCTs, claiming that they represent the medicalization of the psychotherapy field (Goldfried & Wolfe, 1998), the humanistic psychotherapist Carl Rogers, in fact, used such trials to establish the efficacy of psychotherapy several years ahead of the first published medication trials among psychiatric populations (Rogers & Dymond, 1954).

To make use of an RCT for identifying potential harmful effects, it is necessary as a first step to ensure that randomization did not fail and that comparable kinds of patients were distributed equally across the conditions. Before examining outcomes, one must first check the adequacy of randomization by testing for comparability between conditions on relevant patient characteristics. In subsequent analyses, it is useful to try to control for observed differences that predict outcome. At that point, one may examine whether differences exist between the conditions; such differences may be manifest in differences between the means or the variances. Mean differences as a function of treatment indicate constant harm across the full range of patients.

However, in some instances, negative effects may not be readily detected by comparison of group means. When effects are moderated, tests for interactions are recommended but are notoriously underpowered if the interactions are ordinal (B. Smith & Sechrest, 1991). In such cases, it can be helpful to plot individual cases as a function of treatment condition. Visual inspection of such plots can facilitate the detection of outliers who may indeed have been harmed by treatment (the fourth quartile of patients in the treatment group). If there are indications of patients who were harmed, the next step is to identify their characteristics and determine whether comparable proportions of those patients were represented in the control or comparison conditions and, if so, whether they were free from harm. If that is the case, then it may be plausible to attribute the negative outcomes to the treatment. We encourage investigators to examine their data in these ways, starting at the level of the individual patient. We further concur with recommendations of Lilienfeld (2007) that investigators report the full range of scores on dependent variables distributed by quartile or the number needed to harm for dichotomous outcomes.

Examine a Broad Range of Outcomes Over Time

It also is important to be open to a wide range of possible negative outcomes both in clinical practice and in RCTs. As the note for Table 1 indicates, problems that were not the initial target but that develop or worsen as a conse-

quence of treatment are important to assess directly. Systematic measurement of a broad range of outcomes helps to identify such potential harmful effects. Harmful effects may be missed if measurement strategies are limited to detecting change in the target problem exclusively.

Unanticipated harmful effects may be experienced by the client or by others in the client's life. For example, McLennan and Offord (2002) have raised concerns that efforts to prevent postpartum depression may lead to unintended adverse outcomes, such as increased guilt or anxiety among at-risk mothers. Although such claims are purely speculative, McLennan and Offord cited examples of studies targeting the prevention of other disorders in which adverse effects were reported (e.g., increased anxiety among parents of children with developmental disabilities who received a screening and referral intervention; Cadman et al., 1987).

Examination of a broad range of potential effects on the client's interpersonal network also may be important. For example, outcomes of assertiveness training may include not only desired increases in assertiveness for the client, but also potentially undesired effects on others in the patient's life. As discussed previously, pharmacological treatment of antenatal depression may improve a mother's depressive symptoms, while simultaneously exposing the fetus to possible adverse effects (Wisner et al., 2000).

Measurement strategies for high probability adverse outcomes might be planned a priori; however, it also is important to have a system in place that can detect negative outcomes that are not anticipated. It is impossible to anticipate all the things that could go wrong. Monitoring and reporting all serious adverse events without regard to whether they are clearly linked to the treatment is one way to ensure that possible harmful effects are not overlooked. This is a standard part of monitoring in medication trials. It is easier to decide after the fact whether a pattern of differences exists as a function of treatment than to anticipate in advance the full range of harmful effects a treatment might have.

Attrition can sometimes mask harmful effects. It is important to do as much as possible to learn why patients drop out of treatment and specifically to assess whether patients stop treatment because it was having a deleterious effect. Moreover, in the context of an RCT, the basic principle of "once randomized, always analyzed" means that all patients who enroll in a trial are included in outcome analyses even if they drop out of treatment. This principle encourages investigators to collect as much information as possible about the course of patients who drop out of treatment, thereby maximizing the likelihood of identifying harm that is masked by attrition. Although such naturalistic observations are clearly uncontrolled, they may provide useful information that highlights possible problems with treatment. Finally, follow-up periods are similarly necessary to assess whether harmful effects emerge over time. As discussed earlier, some of the treatments that have generated controversy regarding potential harm are ones in which the onset of the adverse effects was delayed (e.g., grief therapy, critical incident stress debriefing).

Examine the Active Ingredients of Treatments That Cause Harm

Most psychotherapies are complex, multifaceted packages. They typically contain strategies that are specific to the approach (e.g., cognitive restructuring) and that are common to many approaches (e.g., empathy). To say that such a package is harmful may not provide sufficient information about what specifically caused harm. In fact, Bootzin and Bailey (2005) have encouraged attention to the ways in which specific and common components of an intervention may interact to cause harm. Descriptive and experimental methods are important in providing more informative detail about what may have led to harm.

Descriptively, it is important for investigators to report measurement of treatment fidelity and quality in RCTs. Variability in treatment implementation can occur across studies or within a study at the level of therapists or patients. Delivery of the intervention in ways that are not faithful to the basic approach or that diverge in important ways may be a potential cause of harm. Reliance on treatment manuals that characterize the nature of the intervention can be valuable. Although the use of treatment manuals has generated controversy in the field of psychotherapy research, we underscore that such manuals are not required to be strict session-by-session guides for the clinician (protocol driven manuals; e.g., Barlow & Craske, 2006). In contrast, treatment manuals may be principle driven, providing an explanation of broad principles or phases that guide the application of the treatment (e.g., Beck, Rush, Shaw, & Emery, 1979; Kabat-Zinn, 1990; Linehan, 1993). In addition, it is critical to conduct and report careful ongoing checks on the ways in which such manuals are implemented to make the best possible use of RCTs in identifying harm.

The need for checks on fidelity and quality is underscored by data suggesting that the same psychological intervention (delivered slightly differently) can produce highly variable results. An interesting example comes from the recent development of social–psychological interventions to mitigate the effects of negative stereotypes. Self-affirmation interventions have been proposed as an intervention for countering the negative performance effects associated with stereotype threat (the stress arising from the perception that one could be judged poorly on the basis of group membership; Steele, 1988). In a double blind study with seventh-grade students, Cohen, Garcia, Apfel, and Master (2006; Cohen, Garcia, Purdie-Vaugns, Apfel, & Brzustoski, 2009) found that African American students randomly assigned to complete a series of structured 15-minute identity affirmation writing assignments—in which they wrote about an important personal value, such as relationships with friends or religion—evidenced significantly higher academic performance over the course of middle school compared to control students.

Self-affirmation interventions, however, are not universally beneficial in counteracting negative effects of prejudice. In fact, some appear to backfire. In some cases, the affirmation intervention may be effective only when it is

unrelated to the domain of threat (e.g., affirming not academics but relationships when in school). For instance, there is robust evidence that self-affirmation interventions mitigate defensive resistance to threatening information, such as threatening health risk information (Harris & Napper, 2005; Sherman, Nelson, & Steele, 2000), making people more open to changing destructive personal behavior or to adopting health-promoting behavior. However, such positive effects are nullified or even reversed when people affirm in the same domain as the one in which they are being threatened (e.g., affirming the personal importance of health conscientiousness before reading a health message that suggests one is engaging in risky health behavior; see Blanton, Cooper, Skurnick, & Aronson, 1997; Jacks & O'Brien, 2004; Sherman & Cohen, 2006). Likewise, certain types of affirmations, such as affirming one's morality or objectivity, can backfire in the interpersonal domain by giving people a sense of impunity and thus licensing them to stereotype and denigrate others (see Sherman & Cohen, 2006, for a review).

The field undoubtedly would demand to know if a chemical in the composition of a pharmaceutical drug had been modified or replaced in a test of its efficacy. It appears that the outcomes of even seemingly simple interventions, such as brief self-affirmations, can vary depending on the domain that individuals are asked to affirm. Complex psychotherapies clearly can be applied in a range of ways. The use of treatment manuals and ongoing tests of fidelity in the conduct of an RCT help to address the wide possible variability of treatment implementation. Clinicians in routine practice also can use such manuals and measures to guide and assess in an ongoing manner the nature of their treatment implementation.

Investigators can address such questions more rigorously with the use of experimental methods. Efforts to examine the active ingredients of benefit have turned increasingly to the use of component analysis and additive designs. For example, in a well-known component analysis study of cognitive therapy, Jacobson et al. (1996) dismantled cognitive therapy into its components: behavioral activation, cognitive restructuring focused on automatic thoughts, and cognitive restructuring focused on core beliefs. They then tested the efficacy of these components in an effort to identify the necessary and sufficient conditions for benefit. Such methods also can be used for identifying active ingredients of harm.

Examine the Mechanisms by Which Harm Is Produced

RCTs allow psychologists to draw causal inferences about the relationship between an intervention and harmful outcomes. RCTs are an essential tool in the effort to detect harmful effects; however, detecting an effect is not synonymous with explaining an effect. Such explanations require an understanding of the mechanisms by which a treatment produces a harmful outcome. Specifically, identifying the mechanism refers to the variable that is causally responsible for the harm produced by the treatment. Identification of the mechanisms by which harm is produced clearly can help to make interventions safer. As Kazdin (2008) recently observed, "The study of mechanisms of change has received the least attention even though understanding mechanisms may well be the best long-term investment for improving clinical practice and patient care" (p. 151). This is equally true in the case of harm. Identifying mechanisms can improve clinical practice and care by specifying precisely what to avoid. In addition, attention to mechanisms may help to protect against the harm that may result from inaccurate inferences. For example, in their criticism of the research on the harmful effects of group interventions for antisocial youths, Weiss et al. (2005) suggested that the processes that occur during group treatment fail to provide a compelling rationale for the purported harmful effects.

Identifying mechanisms of harm requires experimental manipulation of the variable in question, and such studies are critically important. Typically, however, such studies are preceded by investigations of potential mediators of change (i.e., variables that account for some of the effects of the intervention but that have not been experimentally manipulated). RCTs provide ideal settings in which to examine potential mediators of change, and such efforts may help to identify candidate mechanisms for further study. Multiple recommendations for investigating mediators of change have been offered (Kraemer, Wilson, Fairburn, & Agras, 2002; Murphy, Cooper, Hollon, & Fairburn, 2009); such methods should occupy a central focus of efforts to detect harm.

Examine Whether Harm Is Universal or Moderated

It is possible that a standard treatment causes harm for all patients; however, if not everyone evidences harm, then it may be important to look further. As discussed previously, it is possible that some patients may have received a nonstandard delivery of the treatment that caused harm. It also is possible that some patients are particularly susceptible to being harmed by the standard treatment. We operate from the basic assumption that treatments fail patients, not that patients fail treatments (see also Linehan, 1993). If harm is less than universal, then it is possible that the particular treatment is contraindicated for some patients. For such patients, treatment was harmful, and it is important to determine their characteristics so that others with such characteristics can be protected from future exposure.

A moderator is a variable that influences the direction or magnitude of the relationship between treatment and outcome. A number of studies have suggested important moderators of treatment across patient characteristics and treatment interventions. For instance, in an analysis of data across multiple trials comparing antidepressants and placebo, pretreatment depressive severity moderated the effect of antidepressant use (Fournier et al., 2009). Antidepressants demonstrated only small effects among patients with less severe forms of major depressive disorder (those with Hamilton Rating Scale for Depression scores less than or equal to 19; Hamilton, 1960); however, large effects were evident among patients with more severe forms of major

depressive disorder (Hamilton Rating Scale for Depression scores greater than or equal to 20).

The best way to test hypotheses about patient characteristics that moderate outcome is to conduct an RCT that is adequately powered to test the interaction of the patient characteristic and treatment. The key point is that patients with the same characteristics of interest are randomly assigned to different interventions. This can be as readily accomplished by randomization (stratifying on the relevant characteristics) as by matching strategies. The goal at this point is not to assign patients to the intervention that is expected to work best for them, but rather to be sure that some patients are assigned to the intervention that is expected to work best for them and others are assigned to the intervention that is expected to work less well. Designs that hold treatment constant and allow individual differences to vary can tell researchers which patients to select to make treatments look good, whereas designs that hold individual differences constant and allow treatments to vary (as experimental manipulations) can tell researchers which treatment is best for a given individual given his or her individual difference characteristics.

Unfortunately, studies to detect the interaction of treatment and moderator typically require large samples, and many trials are underpowered to detect interactions that are ordinal (B. Smith & Sechrest, 1991). Moreover, given that what psychologists know about possible harmful effects is vastly outstripped by what they do not know, as they build the knowledge base in this area, the goal will likely be to detect effects that have not been anticipated.

Controversial methods may have a role in such research at this point in the development of the knowledge base. For instance, the use of subgroup analyses to examine variables that may be possible moderators has been a source of particular controversy. Many have argued that it is not permissible to test for subgroup differences in the absence of significant interactions between patient and treatment (Freemantle, 2001). The purpose of this convention is to guard against data dredging and subgroup findings that are not likely to replicate. The utility of this convention must be balanced against the risk of overlooking potential differences in response across different patients, particularly when those differences relate to possible harmful effects. For example, in a recent cogent discussion of moderation and mediation, Kazdin (2007) highlighted the role of subgroup analyses, explaining,

> Rather than looking for main effects of an intervention and a uniform mechanism of change, we may need to identify and characterize subgroups, very much in the way that genetic researchers often profit from looking at special groups and individual outliers. (p. 22)

It also is possible that hypothesis-generating studies may be more important at this stage of the research in order to identify possible variables to test in future designs. Exploratory strategies must proceed on a post hoc basis. Within this context, however, numerous writers have urged researchers to ground their exploration in sound theory. Kazdin (2007), for example, underscored the importance of theory in identifying potential moderators, explaining

> In much of treatment research and moderator research in clinical psychology more generally, moderators of convenience are used, such as information routinely obtained and global indices (e.g., socioeconomic class, ethnicity, comorbidity). There is little sound theory behind the research or predictions that derive from proposing precisely what facets of the moderator might be important in explaining the relation. . . . This is fine as a start, but much of the research never gets past the 'start'. (p. 14)

Theory can guide the field in identifying variables that may moderate harmful effects of treatments. Identification of moderators may be critical in resolving debate over, say, harmful effects of grief therapy. In a recent discussion of grief therapy, Bonanno and Lilienfeld (2008) argued that reference to the normative developmental course of grief may help to inform an understanding of for whom grief interventions may be efficacious. They emphasized that most people are resilient in the face of loss, whereas a small minority experience complicated grief reactions. Despite significant debate, there appears to be emerging consensus that it is important to examine factors that may influence the course of treatment response (e.g., time since loss, self-referred vs. recruitment initiation of treatment; Bonanno & Lilienfeld, 2008; Hoyt & Larson, 2008). The use of both hypothesis-testing and hypothesis-generating methods and reliance on sound theory to identify potential moderators represents an important part of the process of understanding the full effects of psychotherapy and protecting patients from unanticipated harm.

Promote Discussion About Standards and Replication

It is incumbent on the field to begin systematic discussions about potentially harmful effects of psychotherapy. It is critical that psychologists discuss complicated questions, such as whether there may be times when it is desirable to adopt less rigorous standards for raising concerns about possible harm than are required to provide proof of therapeutic benefit. There is consensus in the field that a claim of treatment efficacy requires a treatment to exceed a control treatment beyond a probability of .05 or to demonstrate equivalency with an established treatment; such claims also must be replicated by at least one other independent group to be considered empirically supported (Chambless & Hollon, 1998). Additional attention has been focused on determining whether treatment effects are clinically significant. To define clinical significance, researchers have focused on the considerations of measurement error and normative standards (Jacobson & Truax, 1991; Kendall, 1999). Some investigators have argued that methods of clinical significance should be similarly applied to an analysis of potential harm to ensure that purported evidence of harm is not simply measurement error (Devilly & Foa, 2001). Such methods may have much to offer; however, researchers have not yet developed consensus regarding use of these methods in evaluating clinical benefit, much less regarding their application to questions of harm.

Moreover, it is not clear whether the same standards should be applied to considerations of harm. If death is a legitimate risk of rebirthing therapy, how many children would have to die to meet standards that parallel those proposed for efficacy, and would the field insist on an independent replication before this intervention was determined to be harmful? In addition, that which would be considered a harmful effect of intervention for one disorder might be considered a commonplace occurrence for another. For example, DSMBs for studies on depression that we are currently leading closely monitor and evaluate any suicide attempt made by participants. In contrast, suicide attempts are common among people with borderline personality disorder and, in fact, often are considered a primary outcome variable of research (Linehan et al., 2006). One of the authors (Steven D. Hollon) serves on a DSMB for a study on the treatment of suicidal behavior in which suicide attempts do not rise to the level of an adverse event that must be reported in an expedited manner to the institutional review board.

Finally, decisions about the potential harmful effects of a treatment must always be weighed in the context of the potential benefit and harm of alternative approaches, including failing to treat the target problem. For example, although there has been considerable debate in the literature and the popular press about the potential adverse effects of fetal exposure to pharmacological treatment for maternal depression during pregnancy, such discussions often ignore the adverse effects of fetal exposure to untreated maternal depression during pregnancy. Recent reports have called for greater attention to weighing the potential harm and benefit from treatment and lack of treatment (Henry, Beach, Stowe, & Newport, 2004).

These are matters of judgment and value about which there is not yet full consensus. We encourage full and frank discussion about standards that ought to be applied and the need for and risks of replication. Such discussions would benefit from the involvement of multiple parties, including practitioners, researchers, patients, ethicists, public policy experts, attorneys, and statisticians. The inclusion of multiple stakeholders should help to protect individual clinicians from false accusations of harm while simultaneously protecting patients from treatments that cause harm. Organizations such as the American Psychological Association or the National Institutes of Health could take the lead in promoting discussions of these important matters.

Conclusion

Treatments cause harm when they make people worse or prevent them from getting better. Incorrect inferences regarding treatment efficacy also can cause harm. Efforts to detect harm are complicated by the fact that different target problems follow different natural courses. Moreover, treatments can affect other outcomes that were not the initial target of intervention; such outcomes are easily overlooked. Many of the same strategies that are used to detect benefit can be used to detect harm. However, these strategies must sometimes be modified to reflect the fact that harm can be rare and unanticipated. We offered recommendations for how to increase the likelihood that harmful effects will be detected if present. In particular, we recommend establishing independent systems for monitoring untoward events in clinical practice, reporting descriptive case studies and qualitative research, and making full use of randomized clinical trials, including examining potential active ingredients, mechanisms, moderators, and a broad range of outcomes measured over time. We also highlighted the value of promoting full and frank discussion in the field about standards for defining and identifying harm. These recommendations were offered on the principle that treatments fail patients and not the other way around. It is incumbent on the field to address more directly the potential for harm, as well as benefit, that may result from psychotherapy.

REFERENCES

Adler, C. M., Craske, M. G., & Barlow, D. H. (1987). Relaxation-induced panic (RIP): When resting isn't peaceful. *Integrative Psychiatry, 5,* 94–100.

American Psychiatric Association. (2000a). Practice guideline for the treatment of patients with major depressive disorder (revision). *American Journal of Psychiatry, 157,* 1–45.

American Psychiatric Association. (2000b). *Therapies focused on attempts to change sexual orientation (reparative or conversion therapies): Position statement.* Retrieved from http://www.psych.org/Departments/EDU/Library/APAOfficialDocumentsandRelated/PositionStatements/200001.aspx

Antony, M. M., Orsillo, S. M., & Roemer, L. (2007). *Practitioner's guide to empirically based measures of anxiety*. New York: Springer.

Barlow, D. H., & Craske, M. G. (2006). *Mastery of your anxiety and panic: Client workbook* (4th ed.). New York: Oxford University Press.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression.* New York: Guilford.

Blanton, H., Copper, J., Skurnick, I., & Aronson, J. (1997). When bad things happen to good feedback: Exacerbating the need for self-justification through self-affirmation. *Personality and Social Psychology Bulletin, 23,* 684.

Bonanno, G. A., & Lilienfeld. (2008). Let's be realistic: When grief counseling is effective and when it's not. *Professional Psychology: Research and Practice, 39,* 377–380.

Bootzin, R. R., & Bailey, E. T. (2005). Understanding placebo, nocebo, and iatrogenic treatment effects. *Journal of Clinical Psychology, 61,* 871–880.

Breggin, P. R. (2004). Suicidality, violence and mania caused by selective serotonin reuptake inhibitors (SSRIs): A review and analysis. *International Journal of Risk and Safety in Medicine, 16,* 31–49.

Cadman, D., Chambers, L., Walter, S., Ferguson, R., Johnson, N., & McNamee, J. (1987). Evaluation of public health preschool child developmental screening: The process and outcomes of a community program. *American Journal of Public Health, 77,* 45–51.

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

Coffman, S. J., Martell, C. R., Dimidjian, S., Gallop, R., & Hollon, S. D. (2007). Extreme nonresponse in cognitive therapy: Can behavioral activation succeed where cognitive therapy fails? *Journal of Consulting and Clinical Psychology, 75,* 531–541.

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006, September 1). Reducing the racial achievement gap: A social–psychological intervention. *Science, 313,* 1307–1310.

Cohen, G. L., Garcia, J., Purdie-Vaugns, V., Apfel, N., & Brzustoski, P. (2009, April 17). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science, 324,* 400–403.

Davidson, G. C. (2001). Conceptual and ethical issues in therapy for the psychological problems of gay men, lesbians, and bisexuals. *Journal of Clinical Psychology, 57,* 695–704.

DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young,

P. R., Salomon, R. M., et al. (2005). Cognitive therapy vs. medications in the treatment of moderate to severe depression. *Archives of General Psychiatry, 62,* 409–416.
Devilly, G. J., & Foa, E. B. (2001). The investigation of exposure and cognitive therapy: Comment on Tarrier et al. (1999). *Journal of Consulting and Clinical Psychology, 69,* 114–116.
Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm. Peer groups and problem behavior. *American Psychologist, 54,* 755–764.
Emslie, G. J., Rush, A. J., Weinberg, W. A., Kowatch, R. A., Hughes, C. W., Carmody, T., et al. (1997). A double-blind, randomized, placebo-controlled trial of fluoxetine in children and adolescents with depression. *Archives of General Psychiatry, 54,* 1031–1037.
Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology, 16,* 319–324.
Foa, E. B., Zoellner, L. A., Feeny, N. C., Hembree, E. A., & Alvarez-Conrad, J. (2002). Does imaginal exposure exacerbate PTSD symptoms? *Journal of Consulting and Clinical Psychology, 70,* 1022–1028.
Fournier, J. C., DeRubeis, R. J., Hollon, S. D., Dimidjian, S., Amsterdam, J. D., Shelton, R. C. & Fawcett, J. (2009). *Do antidepressants have a "true" drug effect across the range of depression severity?* Manuscript in preparation.
Freemantle, N. (2001). Interpreting the results of secondary end points and subgroup analyses in clinical trials: Should we lock the crazy aunt in the attic? *British Medical Journal, 322,* 989–991.
Gibbons, R. D., Brown, C. H., Hur, K., Marcus, S. M., Bhaumik, D. K., Erkens, J. A., et al. (2007). Early evidence on the effects of regulators' suicidality warnings on SSRI prescriptions and suicide in children and adolescents. *American Journal of Psychiatry, 164,* 1356–1363.
Gibbons, R. D., Hur, K., Bhaumik, D. K., & Mann, J. J. (2006). The relationship between antidepressant prescription rates and rate of early adolescent suicide. *American Journal of Psychiatry, 163,* 1898–1904.
Goldfried, M. R., & Wolfe, B. E. (1998). Toward a more clinically valid approach to therapy research. *Journal of Consulting and Clinical Psychology, 66,* 143–150.
Haldeman, D. (1994). The practice and ethics of sexual orientation conversion therapy. *Journal of Consulting and Clinical Psychology, 62,* 221–227.
Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry, 23,* 56–61.
Harris, P. R., & Napper, L. (2005). Self-affirmation and the biased processing of threatening health-risk information. *Personality and Social Psychology Bulletin, 31,* 1250–1263.
Henry, A. L., Beach, A. J., Stowe, Z. N., & Newport, D. J. (2004). The fetus and maternal depression: Implications for antenatal treatment guidelines. *Clinical Obstetrics and Gynecology, 47,* 535–546.
Hollon, S. D., Thase, M. E., & Markowitz, J. C. (2002). Treatment and prevention of depression. *Psychological Science in the Public Interest, 3,* 39–77.
Hoyt, W. T., & Larson, D. G. (2008). A realistic approach to drawing conclusions from the scientific literature: Response to Bonanno and Lilienfeld. *Professional Psychology: Research and Practice, 39,* 378–379.
Jacks, J. Z., & O'Brien, M. E. (2004). Decreasing resistance by affirming the self. In E. S. Knowles & J. A. Linn (Eds.), *Resistance and persuasion* (pp. 235–257). Mahwah, NJ: Erlbaum.
Jacobson, N. S., & Christensen, A. (1996). Studying the effectiveness of psychotherapy: How well can clinical trials do the job? *American Psychologist, 51,* 1031–1039.
Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., et al. (1996). A component analysis of cognitive–behavioral treatment for depression. *Journal of Consulting and Clinical Psychology, 64,* 295–304.
Jacobson, N. S., & Hollon, S. D. (1996). Cognitive behavior therapy versus pharmacotherapy: Now that the jury's returned its verdict, its time to present the rest of the evidence. *Journal of Consulting and Clinical Psychology, 64,* 74–80.
Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12–19.
Kabat-Zinn, J. (1990). *Full catastrophe living: Using the wisdom of your mind to face stress, pain and illness.* New York: Dell.
Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology, 3,* 1–27.
Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63,* 146–159.
Kendall, P. C. (1999). Clinical significance. *Journal of Consulting and Clinical Psychology, 67,* 283–285.
King, P. D. (1958). Regressive EST, chlorpromazine, and group therapy in treatment of hospitalized chronic schizophrenics. *American Journal of Psychiatry, 4,* 354–357.
Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, 59,* 877–883.
Lambert, R. A., Harvey, I., & Poland, F. (2007). A pragmatic, unblinded randomised controlled trial comparing an occupational therapy-led lifestyle approach and routine GP care for panic disorder treatment in primary care. *Journal of Affective Disorder, 99,* 63–71.
Larson, D. G., & Hoyt, W. T. (2007). What has become of grief counseling? An evaluation of the empirical foundations of the new pessimism. *Professional Psychology: Research and Practice, 38,* 347–355.
Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2,* 53–70.
Linehan, M. M. (1993). *Cognitive behavioral treatment of borderline personality disorder.* New York: Guilford Press.
Linehan, M. M., Comtois, K. A., Murray, A. M., Brown, M. Z., Gallop, R. J., Heard, H. L., et al. (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs. therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry, 63,* 757–766.
Linehan, M. M., & Dexter-Mazza, E. T. (2008). Dialectical behavior therapy for borderline personality disorder. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders: A step-by-step treatment manual* (4th ed., pp. 365–420). New York: Guilford Press.
Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies. Is it true that "everyone has one and all must have prizes"? *Archives of General Psychiatry, 32,* 995–1008.
Marsh, E. J., & Hunsley, J. (1993). Assessment considerations in the identification of failing psychotherapy: Bring the negatives out of the darkroom. *Psychological Assessment, 5,* 292–302.
McLennan, J. D., & Offord, D. R. (2002). Should postpartum depression be targeted to improve child mental health? *Journal of the American Academy of Child and Adolescent Psychiatry, 41,* 28–35.
McNally, R. J., Bryant, R. A., & Ehlers, A. (2003). Does early psychological intervention promote recovery from posttraumatic stress? *Psychological Science in the Public Interest, 4,* 45–79.
Mercer, J., Sarner, L., & Rosa, L. (2003). *Attachment therapy on trial.* Westport, CT: Praeger.
Moseley, J. B., O'Malley, K., Petersen, N. J., Menke, T. J., Brody, B. A., Kuykendall, D. H., et al. (2002). A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *New England Journal of Medicine, 347,* 81–88.
Murphy, R., Cooper, Z., Hollon, S. D., & Fairburn, C. G. (2009). How do psychological treatments work? Investigating mediators of change. *Behaviour Research and Therapy, 47,* 1–5.
Neimeyer, R. A. (2000). Searching for the meaning of meaning: Grief therapy and the process of reconstruction. *Death Studies, 24,* 541–558.
Nezu, A. M., Ronan, G. F., Meadows, E. A., & McClure, K. S. (2000). *Practitioner's guide to empirically based measures of depression.* New York: Kluwer Academic/Plenum.
Nishith, P., Resick, P. A., & Griffin, M. G. (2002). Pattern of change in prolonged exposure and cognitive-processing therapy for female rape victims with posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 70,* 880–886.
Oden, S., Schweinhart, T. F., & Weikart, D. (2000). *Into adulthood: A study of the effects of Head Start.* Ypsilanti, MI: High Scope Press.
Richmond, T. K., & Rosen, D. S. (2005). The treatment of adolescent depression in the era of the black box warning. *Current Opinion in Pediatrics, 17,* 466–472.
Rogers, C. R., & Dymond, R. F. (1954). *Psychotherapy and personality change.* Chicago: University of Chicago Press.
Sackeim, H. A., Prudic, J., Devanand, D. P., Nobler, M. S., Lisanby, S. H.,

Peyser, S., et al. (2000). A prospective, randomized, double-blind comparison of bilateral and right unilateral electroconvulsive therapy at different stimulus intensities. *Archives of General Psychiatry, 57,* 425–434.

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 38, pp. 183–242). San Diego, CA: Academic Press.

Sherman, D. K., Nelson, L. D., & Steele, C. M. (2000). Do messages about health risks threaten the self? Increasing the acceptance of threatening health messages via self-affirmation. *Personality and Social Psychology Bulletin, 26,* 1046–1058.

Shidlo, A., & Schroeder, M. (2002). Changing sexual orientation: A consumers' report. *Professional Psychology: Research and Practice, 33,* 249–259.

Smith, B., & Sechrest, L. (1991). Treatment of Aptitude × Treatment interactions. *Journal of Consulting and Clinical Psychology, 59,* 233–244.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32,* 752–760.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261–302). New York: Academic Press.

Strupp, H. (1980). Success and failure in time-limited psychotherapy: A systematic comparison of two cases: Comparison 1. *Archives of General Psychiatry, 37,* 595–603.

Strupp, H., & Hadley, S. W. (1977). A tripartite model of mental health and therapeutic outcomes. With special reference to negative effects in psychotherapy. *American Psychologist, 32,* 187–196.

Throckmorton, W. (2002). Empirical and clinical findings concerning the change process for ex-gays. *Professional Psychology: Research and Practice, 3,* 242–248.

Webster's third new international dictionary. (1993). Springfield, MA: Merriam-Webster.

Weiss, B., Caron, A., Ball, S., Tapp, J., Johnson, M., & Weisz, J. (2005). Iatrogenic effects of group treatment for antisocial youth. *Journal of Consulting and Clinical Psychology, 73,* 1036–1044.

Whittington, C. J., Kendall, T., Fonagy, P., Cottrell, D., Cotgrove, A., & Boddington, E. (2004). Selective serotonin reuptake inhibitors in childhood depression: Systematic review of published versus unpublished data. *Lancet, 363,* 1341–1345.

Wisner, K. L., Zarin, D. A., Holmboe, E. S., Appelbaum, P. S., Gelenberg, A. J., Leonard, H. L., & Frank, E. (2000). Risk–benefit decision making for treatment of depression during pregnancy. *American Journal of Psychiatry, 157,* 1933–1940.

Writing Group for the Women's Health Initiative Investigators. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the Women's Health Initiative randomized controlled trial. *Journal of the American Medical Association, 288,* 321–333.

Yarhouse, M. A., & Burkett, L. A. (2002). An inclusive response to LGB and conservative religious persons: The case of same-sex attraction and behavior. *Professional Psychology: Research and Practice, 33,* 235–241.