

What Does It Mean for an Intervention to “Work”? Making Sense of Conflicting Treatment Outcomes for Youth Facing Emotional Problems

Jacob Z. Hess & Jeffrey R. Lacasse

As public and professional attention to outcome and evaluation research grows, focus commonly remains centered on the question, “Does this treatment work or not?” Consequently, much less emphasis is given to what exactly it means for a treatment to be effective. This article examines 5 issues relevant to whether an intervention is deemed successful or not: (a) Sponsorship: Who generates the empirical evidence? (b) Sensitivity: How deep does the evidence gathering go? (c) Scope: Over what time period does the evidence span? (d) Source: According to which data is a determination reached? (e) Fairness: How seriously are negative cases examined? After exploring each issue, larger implications for youth treatment are discussed. Ultimately, we propose that current trends warrant more thoughtful deliberation and education among citizens and practitioners about treatment outcomes generally.

IMPLICATIONS FOR PRACTICE

- Families exploring treatment for their child or adolescent deserve to know about conflicting professional evaluations associated with a particular treatment.
- Comprehensive education to the broader public, to professional helpers, and to individual patients is crucial to facilitating a more open and healthy collective deliberation about youth treatment.

In January 2002, the No Child Left Behind Act was signed into law by President George W. Bush, establishing a series of new evaluation guidelines across U.S. public schools. A centerpiece of this initiative was new standardized exams that would purportedly determine whether a class or school was effective (or not). As policy implications unfolded, however, some in the educational community began to raise questions about what it actually *meant* for a child or school to pass or fail the tests: Were positive scores necessarily indicative of a successful youth outcome, and negative scores the reverse? How well did this particular way of measuring kids and classrooms legitimately capture the most important aspects of teaching effectiveness (Chrismer, Hodge, & Saintil, 2006)?

Questions about the precise meaning of effectiveness and success are not unique to the field of education. As the push for empirical measurement of outcomes has spread across disciplines, similar questions have been raised across a variety of interventions and treatments—from deliberations about how to determine effective counseling, psychotherapy, and residential treatment, to contested outcomes associated with medications, supplements, and other health interventions (e.g., Adams, Matto, & LeCroy, 2009; Gambrell, 1999, 2010). Although the professional and public benefits of treatment evaluation studies are obvious, at closer view, complexities and nuances regarding the precise nature of successful outcomes become apparent.

For instance, across fields, deliberations of how to evaluate and track outcomes typically revolve around a few questions taken for granted as straightforward and crucial, such as: “Is this particular intervention effective?” or “Does this treatment work?” Other, more specific

questions include, “Where does this treatment work—and for what populations, in what settings?” Although these queries are valuable, they each assume discrete, dichotomous options: The treatment either “works” or it does not (within a certain context, for a particular patient, for a specific disorder, etc.). While technical expertise required to answer such questions is commonly acknowledged by researchers, overall, the format of such a study is understood to be a fairly uncontroversial and clear-cut process.

As many research efforts continue examining whether or not a treatment or intervention works, a related, additional question has received much less attention—namely, what exactly does it *mean* for an intervention to work? For a particular setting, condition, or population, how is one to know whether a behavioral or medical treatment has been truly successful or effective? By what standards or criteria of outcome evidence for a given treatment can this kind of a conclusion be legitimately reached? Romyn et al. (2003) commented that while there is general scholarly agreement that “practice should be based on the best available evidence, there is a lack of agreement [in the research literature as to] a) what the term *evidence* means...b) the ends for which evidence is to be sought and...c) the means by which it is to be acquired” (p. 184, emphasis in original).

This article explores several nuances associated with the meaning of outcomes in social and medical services for serious emotional problems. Though we believe these issues to be relevant across disciplines and age groups, we largely bound our exploration to behavioral and psychiatric treatments for youth with severe emotional problems. Although subtle, the following five questions are proposed as significant enough to potentially shape research findings in this area in tangible ways: (a) Sponsorship: Who generates the empirical evidence? (b) Sensitivity: How deep does the evidence gathering go? (c) Scope: Over what time period does the evidence span? (d) Source: According to which data is a determination reached? (e) Fairness: How seriously are negative cases examined? For each question, diverging stances will be illustrated from outcome studies of various treatments and interventions for serious emotional problems. Although drawing intervention examples from diverse fields, we also highlight our own research specialties of psychiatric treatment and

other interventions into depression and attention-deficit/hyperactivity disorder (ADHD), both of which are particularly relevant as the need for effective youth treatment continues to expand at an alarming rate.

Intervention Issues

Sponsorship: Who Generates the Empirical Evidence?

The first issue concerns the funding source of research. Although the importance of this factor seems obvious, the ramifications of industry sponsorship do not always receive the full scholarly attention they deserve. Recent studies have been clarifying the picture in this regard. In a systematic review of biomedical research, for instance, Lexchin, Bero, Djulbegovic, and Clark (2003) found that industry-funded studies were more likely to reach outcome conclusions favoring the sponsor's product, when compared with research independent of corporate interests, $OR = 4.05$, 95% CI [2.98, 5.51]. In the wake of other similar findings (e.g., Sismondo, 2008), articles in major medical journals have issued warnings regarding the prevalence of industry-funded research efforts and the "uneasy alliance" they entail with the helping professions (Bodenheimer, 2000; Campbell et al., 2007).

It would, of course, be misguided to dismiss research entirely because it was sponsored by those marketing a product. When scientific standards are met, industry-sponsored research may provide valid and legitimate results. However, it is in relation to some of these fundamental scientific standards that concerns still exist. In reviews of industry-funded research into psychotropic medications, for instance, problems range from inadequate blinding and/or use of inactive placebo (Antonuccio, Burns, & Danton, 2002), to the selective reporting of endpoints (Jureidini, McHenry, & Mansfield, 2008), selective publication (Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008), and "ghost authorship" of peer-reviewed publications by companies and their subcontractors (Lacasse & Leo, 2010; Sismondo, 2009). In spite of such concerns, most data for industry-funded psychiatric research are not readily available for external reanalysis. Similar questions could also be asked of outcome results generated by other for-profit organizations such as therapeutic schools and residential treatment centers.

Recent years have seen modest progress in attempting to build a legitimate firewall between industry and academia. Following a number of proposed recommendations (Antonuccio, Danton, & McClanahan, 2003; Pachter, Fox, Zimbaro, & Antonuccio, 2007), many journals now require authors to register clinical trials prior to publication, disclose any potential conflicts of interests, and delineate each author's contribution to the manuscript. With such measures still in their infancy, it appears premature to conclude whether authentic reform is being implemented on the issue. Reflecting perhaps both growing accountability and continuing surreptitious influence, recent scandals have surfaced regarding some of the most prestigious and influential research psychiatrists in the United States receiving unreported payments from the pharmaceutical industry (Harris & Carey, 2008; Tanne, 2008).

Sensitivity: How Deep Does the Evidence Gathering Go?

As a second issue relevant to the meaning and significance of treatment outcomes, we now turn to an issue rarely examined in the evaluation literature: study intensity or depth. More often than not, discussions of general research quality center on a study's level of

structure and control, with carefully monitored experiments such as randomized controlled trials (RCTs) often viewed as the "gold standard." Standardized rating scales, structured surveys, control groups, and statistical measures can be very useful and valuable in demonstrating quantitatively that some kind of change is occurring. Even so, in the wake of such a study, the precise meaning of these bounded, numerical changes (whether positive or negative) is often challenging to interpret.

To illustrate, outcome studies of psychotherapy and residential treatment generally use standardized measures or rating scales to track client progress over time. In some cases, these measures remain the only measure of discharge or postdischarge success. Though providing a helpful snapshot of behavior at a given moment, the bounded focus of these measures on individual behavior change contains some inherent limitations. For instance, it is not always clear what these numerical behavioral ratings say about cognitive or affective shifts happening during treatment, such as changes in motivation, attitude, or interpretation. Larger environmental changes happening (or not) in the family setting are also typically not addressed. These kinds of limitations, in a recent review of outcome studies on child and family services, led Berry et al. (2006) to propose that researchers attempt to track long-term outcomes on a deeper or more sensitive level.

The lead author has been conducting simultaneous quantitative and qualitative outcome analyses for various intervention programs for abused children. When juxtaposed, results from 3,000 behavior-oriented ratings and 175 interviews both complement and challenge each other in striking ways. While statistical results suggest what and how much youth are changing, we found interview data confirming details of how and why the changes are occurring. And while short-term outcome findings document immediate program effects, retrospective interviews conducted 1–7 years post-treatment have delineated distinct trajectory patterns of change over time and allowed examination of interrelated factors that appear to play a role in differentiating enduring outcomes (Hess, Bjorklund, Preece, & Draper, under review).

As reflected here, perhaps the simplest way to increase the depth, intensity, and sensitivity of research is to complement statistical inquiries with rigorous qualitative investigations of these same intervention endpoints. Molloy, Woodfield, and Bacon (2002) proposed that a qualitative approach can add a unique perspective regarding "people's decision-making, experiences and behavior grounded in the experiences and world view of those likely to be affected [by a treatment]" (p. 1). In spite of its potential, in the past qualitative research has been stereotyped by some as a subjective inquiry that involves mere anecdotes. This notion was reinforced by a historic bias that numbers represented the exclusive "language of science" (Dantzig, 1930). Over the last 20 years, however, many in psychology, sociology, nursing, and social work have embraced qualitative research as a valuable complement to conventional quantitative approaches (Drisko, 2008; Gilgun, 1994; Padgett, 2008). This growing contingent recognizes that the systematic and rigorous use of methods such as interviewing and participant observation can yield rich and compelling findings that complement other scientific methods (Denzin & Lincoln, 2005). The National Registry of Evidence-Based Programs and Practices (NREPP; 2009) recently acknowledged that "there is a wide spectrum of possible definitions of 'evidence'" and that in addition to "controlled clinical trials...other methods of establishing evidence are considered valid as well."

The implications of research depth or sensitivity can be substantial. In the context of psychiatric medication outcome research, Jacobs (1999) pointed out a “disparity which exists between side effects established in randomized, placebo-controlled clinical trials” versus a “much broader range and severity of adverse drug reaction reports which emanate from non-RCT formats” (p. 322). In their discussion of psychosocial side effects associated with psychiatric treatments, Moses and Kirk (2005) noted a similar research disparity:

We know comparatively little about the overall experience of being treated with psychotropic drugs. In addition to symptomatic physical changes produced by medications, what is the subjective experience...of taking psychotropic drugs? Are there direct or indirect social and psychological effects on esteem, identity, and other views of the self? (pp. 387, 392–393)

Jacobs went on to point out that controlled clinical trials are designed for the primary purpose of documenting immediate effects for a specified psychiatric condition and suggested that in these studies, generally, “Much less thought and effort is directed at...establishing the full range of psychological alterations [associated with the medication]” (p. 322). Instead, he noted that most of the attention in these studies is directed at either “somatic distress or what could be called the lowest level of drug-induced psychological disturbances (restlessness, agitation, nervousness).” Consequently, it often appears that “drug effects in the realm of psychosocial functioning are ruled out in clinical trials...by virtue of non-investigation” (Jacobs, p. 312).

As reflected in the preceding examples, depending on the sensitivity of a study, a very different picture of outcomes and effectiveness can emerge. In light of these consequences, more serious attention to qualitative studies as a complement to ongoing statistical reviews seems both helpful and necessary. In the context of psychiatric treatment, for instance, there have been increasing numbers of in-depth examinations into the experience of taking medication (Floersch et al., 2009; Karp, 2006; Knudsen, Hansen, Traulsen, & Eskildsen, 2002; Pound et al., 2005; Stoppard & McMullen, 2003; Verbeek-Heida & Mathot, 2006; Venarde, 1999), with one team of researchers calling for qualitative studies to be employed as a formal part of outcome research in psychiatry generally (Crawford, Weaver, Rutter, Sensky, & Tyrer, 2002). Similar progress may offer additional insight across other treatment modalities as well.

Scope: Over What Time Period Does the Evidence Span?

As a third issue, we turn to the temporal scope of a study. When an intervention is presented in marketing or news reports as “effective,” it is likely that most lay people assume the intervention has shown some kind of evidence documenting effectiveness over a meaningful period of time. While an increasing number of longer-term studies are appearing across treatments, the bulk of outcome studies, unfortunately, especially in relation to youth and adolescents, continues to reflect primarily shorter-term results. For instance, although a handful of residential treatment studies provide evidence of some enduring effects for youth (Behrens & Satterfield, 2007), critiques have been raised that most such studies “fail to measure outcome after discharge” (McKay, 2007, p. 74). And in the context of outpatient residential care, Knorth, Harder, Zandberg, & Kendrick (2008) noted that “it is remarkable that there are so few reviews and meta-analyses of outcomes of residential child and youth care services,” before emphasizing, in particular, “there is very little evidence on long-term outcomes” (p. 123).

Similar concerns have been raised in pharmacological studies as well, with new medications routinely approved on the basis of two RCTs, lasting 8 to 12 weeks (see Bridge et al., 2007). After noting a 7.6-week average trial length for one of the latest ADHD drugs, Strattera® (atomoxetine HCl), Cohen, Hughes, and Jacobs (2009) remarked: “One must note that although Strattera was approved, and is marketed and promoted, to treat a ‘chronic’ condition, the only studies reviewed by the FDA were short-term studies” (p. 323; see also Simon et al., 2002). After acknowledging that more children are being treated at even younger ages, developmental neuroscientists Andersen and Navalta (2004) noted similarly that “surprisingly, despite the obvious need for such information, the long-term effects of therapeutic drug exposure on an immature brain have not been adequately assessed at either the clinical or preclinical stage” (p. 424).

In 2005, in pursuit of a higher standard of evidence for drugs used over the long term, the U.S. Food and Drug Administration (FDA) implemented a new requirement that drug companies submit longer-term efficacy data as part of the drug approval process. Within a few months, however, the Psychopharmacologic Drugs Advisory Committee voted 12–0 for a resolution instructing the agency to reverse its decision after 10 pharmaceutical companies submitted testimony that this action would “slow drug development and timely approval of new medications for the treatment of mental illness” (Rosack, 2005, p. 1).

Decisions such as this have kept the number of longer-term studies for psychiatric drugs limited. In the absence of more extensive research, it is subsequently the short-term studies that continue to be promoted among the general public as primary evidence of effectiveness. Cohen et al. (2009) went so far as to suggest that conventional medication studies of “initial, ‘selective’ action” can potentially distract and draw attention away from the more extensive picture of complex, rippling drug effects, including “subsequent cascades of transient and long-lasting neurochemical changes involving other neurotransmitters” (p. 318).

In the absence of more serious and systematic longer-term research, what can legitimately be said about an intervention’s true effects? In the context of residential treatment programs for adolescents, researchers Henggeler and Schoenwald (1994) suggest that these centers “may affect behavior change in a controlled environment, but are not likely to maintain their effectiveness when the individual reenters his/her unchanged family, peer, and neighborhood environment” (cited in Cervenka, Dembo, & Brown, 1996, p. 207). Concerns about temporary surface effects have also been raised across other interventions as well, including certain kinds of psychotherapy, juvenile programs, and community-based educational interventions (Henggeler et al., 1999; Lynam et al., 1999). With psychiatric treatment, Jacobs (1999) observed among his own research participants that “it took time for patients to realize” that psychological side effects associated with a medication had become problematic—specifically, more time than was allowed in “the six week time frame of RCTs conducted for FDA approval” (p. 330). In light of this kind of limitation, Cohen et al. (2009) pointed out an alarming “inability of conventional clinical trials to provide a true picture of a drug’s...[full] effects” (p. 318).

Fortunately, as with previous issues, there appears to be some progress in this area as well. Although the practical meaning of “long-term” research naturally varies across settings, there is a growing realization that for a youth treatment or intervention program to be called “effective,” it should be able to demonstrate those effects in more than the immediate beginning or ending of a formal intervention. This en-

tails finding ways to document over at least a 6 to 12 month period to what degree changes have endured—that is, “was there a change... that we can plausibly attribute to what we set in motion?” (McCardle & Chhabra, 2004).

Although longitudinal research continues to be labeled as an “emergent method” in a recent handbook of social science methodology (Ruspini, 2008, p. 1), long-term studies have substantial histories in many fields and growing relevance in others. Long-term studies have been a central part of the proposed agenda for foster care research since the 1990s (Goerge, Wulczyn, & Fanshel, 1994), for instance, and have an especially impressive record in family-based interventions generally (Cairns, 2004; McDonald, Allen, Westerfelt, & Piliavin, 1996). Long-term studies are also appearing in the area of proctor care for preschool-aged children (Whitmore, Ford, & Sack, 2003) and occurring more often for in-home intensive interventions (Butler, 2006; Lewis, 2004). There are also increasing numbers of longer-term studies of medication effects, in some cases conducted independently of industry funding.

Before continuing, it is worth asking more directly why we do not see more examples of these studies, if they offer such rich and telling information. Perhaps it is this very depth and power of the information that provides a partial answer to the question. The potential for (even brutal) accuracy of long-term research was illustrated well in the 10-year follow-up of the youth-focused Drug Abuse Resistance Education (DARE) program, where an otherwise popular and lauded initiative was shown to have dubious long-term effects (Lynam et al., 1999). More recently, a 5-year follow-up on a “continuum of care” innovation similarly showed no difference in children’s improvement across demonstration and comparison sites (Bickman, Lambert, Andrade, & Penaloza, 2000).

Of course, study results such as these can potentially prompt significant revamping and improvement to programs. They can be a powerful way to adjust and upgrade interventions to better address the needs of individuals and families. On the other hand, disappointing results can also clearly impact profits if a proprietary intervention disappoints over the long term. Just as surface-level findings were suggested earlier as potentially distracting from more in-depth evidence, it appears that short-term research may likewise draw attention away from longer-term views. In the context of psychotropic treatment, for instance, Whitaker (2010) argued that in the modern psychiatric treatment paradigm, dependence on short-term trials as evidence may have inflated the perceived efficacy of medications and obfuscated longer-term realities, since studies of the same drugs over time often show untreated groups faring better. The importance of a longer-term view is thus providing families with a clear understanding of what can be expected from an intervention over time.

Source: According to Which Data Is a Determination Reached?

Whose voice and which data are considered valid evidence? Overlapping with issues reviewed so far is another distinction deserving attention. In the context of depression treatment, sociologist David Karp (1997) suggested that “the essential problem with nearly all studies of depression is that we hear the voices of a battalion of mental health experts...and never the voices of depressed people themselves” (p. 11). In his review of *Journal of Affective Disorders*, he noted that in 12 volumes of this journal, he could not find one word spoken by a person who lives with depression. He concluded that “research about a feeling disorder that does not get at people’s feelings seems, to put it kindly, incomplete” (p. 12).

As with issues discussed earlier, depending on whose views and judgments are held as credible and legitimate, researchers may naturally arrive at very different findings. This is, once again, illustrated pointedly in psychiatric studies over recent decades. In 1997, Emslie and colleagues conducted research on the efficacy of fluoxetine (Prozac®) in children. It was heralded as a “landmark study,” with the results presented in the media as the first evidence that “Prozac works for children.” This study corresponded with a sharp increase in antidepressant prescriptions for children across the nation, with rates nearly doubling (49%) over the next 5 years (Delate, Gelenberg, Simmons, & Motheral, 2004). Several years later, mounting reports by parents to the FDA on adverse effects for youth on Prozac, including atypical child suicides, led to a review of the original study. Although five different measures were originally used to compare medication and placebo (three parent/child self-report scales and two clinician ratings), it was discovered that Emslie et al.’s original conclusions were drawn almost exclusively from clinician ratings. Since patterns of clinician ratings differed considerably from what the parents and children themselves reported, a separate analysis giving equal weight to all scales reached very different conclusions about the safety and efficacy of antidepressant use in children (Safer, 2006).

A similar dynamic was observed in Study 329, an industry-funded trial of paroxetine (Paxil®) for youth diagnosed with major depression, where the medication was first reported to be “generally well tolerated and effective for major depression in adolescents” (Keller et al., 2001, p. 762). Throughout the last decade, Study 329 was cited as evidence that paroxetine was effective in children (Jureidini & Jureidini, 2008). Later reanalysis prompted by litigation, however, revealed that outcomes had been redefined and selectively reported, and that contrary to the published conclusions, the study data were “negative on all protocol-specified outcomes and positive for harm” (Jureidini et al., 2008, p. 73). It was further discovered that the article had been written by a drug company employee, before being submitted for publication by an academic psychiatrist (“ghostwriting”). In this case, then, it was the corporate voice that emerged dominant at the expense of what the actual patient data were suggesting.

Laying aside the influence of sponsorship, we see here illustrations of research conclusions emerging from a selective emphasis on certain data (and voices) as more legitimate than others. In the wake of such patterns, formal attention to the worth of all voices—including those of parents and youth themselves—becomes especially crucial. In contrast to viewing researchers and academicians as the exclusive experts on phenomena, for instance, fields such as community psychology have come to emphasize that citizens who have lived through the trauma of depression, schizophrenia, and so forth, possess an “expertise from experience” that offers valuable insight beyond that which can be obtained from abstract study or research. And by its emphasis on personal language and narratives, the move toward qualitative research, reviewed earlier, contributes to the highlighting of layperson experiences. Among other things, by ensuring that the voices of patients, consumers, and clients are heard and valued more consistently as a part of outcome and evaluation research, overall conclusions may arguably become more accurate and reliable (Crawford, et al., 2002; Lietz, 2009).

Fairness: How Seriously Are Negative Cases Examined?

The fifth and final issue discussed here, fairness, is clear in its importance but often difficult to detect. Given the significant investment in research and development involved in creating a new treatment, it is

natural for individuals and organizations to assume that their own interventions are positive and helpful. While this belief is understandable, if it prevents an organization or individual from seeing or acknowledging counterevidence to their intervention, then problems may arise.

Under the Freedom of Information Act, Cohen et al. (2009) obtained all 11 studies originally submitted to the FDA for approval of the ADHD drug Strattera®. In dissecting the details of these studies, these researchers found that the average number of measures for positive drug efficacy exceeded measures of possible adverse effects by a significant margin: While there were between three and seven measures of positive short-term drug effects across studies, most studies had few, if any, rigorous measures of negative drug effects:

In nine published reports, measurement of adverse effects was elicited only via “self-report” (one study), “spontaneous reports from parent or child” (one study), “unsolicited adverse event reports” (two studies), and “open-ended questions” (four studies). One study did not report on how measurements of adverse effects were collected. (Cohen et al., 2009, p. 325)

In light of such evidence, Cohen et al. decried a troubling pattern of “selective presentations and publication of adverse effect data” (p. 316), wherein “published results...tend to distort or conceal negative findings and emphasize positive findings” (p. 320). Based on this review, they went on to share an overall impression from these studies that “ascertaining harm from treatment takes a distinctly subordinate position to the goal of establishing superiority of a tested drug to placebo” (p. 324). This same research team compiled evidence suggesting similar problems across many other psychiatric drug trials as well (Hughes & Cohen, 2010).

True fairness, of course, would require that researchers actively seek out negative cases and evidence. It is well acknowledged in methodological discussions of both qualitative and quantitative research that exploring a phenomenon in a reasonably complete way requires seeking contradictory evidence (Drisko, 1997; Frankel, 1999). This reflects one of the fundamental criteria of good scientific research: *testability*, or the basic openness to being shown otherwise (Flew, 1995).

Conclusion

After reviewing over 4,000 “scientifically rigorous studies of family-based services in children’s health and mental health” since 1980, Hoagwood (2005) noted, “the range of outcomes that are typically assessed in clinical treatment studies is *too* narrow to afford an adequate view of the impact of family-based interventions. A broader view of outcomes is needed” (p. 708, emphasis in original).

It is toward this broader view of outcomes that we have proposed more carefully examining results from youth intervention studies generally, calling for specific attention to the neglected question, “What does it mean for an intervention to work?” To facilitate a more thoughtful collective deliberation regarding this question, several key issues have been reviewed: “Who is sponsoring the study and how are associated researchers handling any potential conflicts of interest?” “At what depth and duration is the purview of the study?” “To what degree are nonprofessional, client voices also privileged in the study?” and, “How are researchers ensuring that both negative and positive effects are being equally examined?”

In the absence of attention to these kinds of questions, we submit that researchers may, even unaware, set up studies that are predisposed toward positive results—for example, short-term investigations using surface-level measures that minimize both participant voices and negative outcomes. Ensuing results may then be presented to potential clients and the broader public as evidence that X therapy or Y medication works. As similar dynamics are replicated across many studies, claims of consensus may then eventually appear, with general statements on the safety and efficacy of a particular treatment appearing in evidence-based treatment guidelines and academic textbooks, and even forming the basis for institutional standards of care (see Healy, 2006; Turner et al., 2008; Whitaker, 2010).

Of course, if rigorous research evidence ultimately shows that a given intervention is truly effective, then such information should be broadly disseminated. For a treatment that has demonstrated safety and efficacy, distressed individuals and families deserve to know. On the other hand, if an intervention is deemed effective according to the dubious and superficial methodological standards reviewed here, serious questions should remain for anyone involved. To the degree that this is happening, we conclude with three simple recommendations arising from our investigation: (a) more comprehensive, in-depth research; (b) more comprehensive public and professional education; and (c) more comprehensive informed consent.

More Comprehensive, In-Depth Research

First of all, implicit in this review is a call for more long-term, in-depth, and balanced studies. In current deliberations of outcome research across professions, much emphasis has been given to better specifying the conditions under which treatments are effective (e.g., Hinshaw, 2007; Kazdin & Nock, 2003). According to this argument, if researchers are more skillful in delineating in what population, for what conditions, and within what setting a treatment is being evaluated, results of an outcome study will be generally clear and straightforward.

As illustrated earlier, in the absence of attention to the issues we raise, this kind of an effort to better specify outcome terms may not be enough, on its own, to produce clear results. Two research teams studying the same outcome question, for instance, could carefully target the same population under the same conditions and insist on the same settings and instruments. Yet depending on how issues of sponsorship, depth, scope, and balance are handled, study conclusions could still vary dramatically.

In light of this, more long-term, in-depth, and balanced studies might provide for clients and the general public a more accurate and honest assessment of what effects can legitimately be expected from a particular treatment. We are currently developing a research quality scale to allow more systematic, empirical examination of the range of variables specified in this article. More than simply longer-term research, we highlight a need for more thoughtful attention to the voices and experiences of those facing the problems themselves. One basic way this might be accomplished is through the use of more mixed-method research designs (Creswell, 2003; Padgett, 2008).

All methods have limitations, and the qualitative and longitudinal approaches reviewed here are no exception. The scientific method itself acknowledges such limitations and the fallibility of any one study’s conclusions. It is for this reason that these methodological suggestions are proposed as complementary to ongoing conventional statistical studies. There is a safety in methodological pluralism comparable to the checks and balances in U.S. governmental systems. Within the

open dialogue between results from multiple research approaches a collective consensus can emerge that is much stronger than a meta-analysis of 30 studies employing the same methodology (see Slife & Gantt, 1999). In the absence of such openness to other approaches, however, conclusions from malformed studies may continue to gain more traction than they deserve. In the end, of course, those who lose out by such practices are those who are most vulnerable: youth facing emotional challenges themselves.

More Comprehensive Public and Professional Education

In addition to better research, we thus propose a second way to promote more thoughtful public and professional deliberations about treatment. Like much of the treatment research itself, there is a growing awareness that education about treatments has been funded, to varying degrees, by corporations that manufacture and sell those treatments. Following a U.S. Senate probe, for instance, it was recently discovered that one of the premier mental health educational and advocacy organizations, the National Alliance on Mental Illness (NAMI), has received more than 75% of its budget from pharmaceutical companies (Harris, 2009). Like industry-linked research, education associated with industry does not necessarily mean it is invalid. In light of evidence of distortion in industry marketing (Lacasse, 2005; Leo & Lacasse, 2008), however, it is difficult to believe that a similar influence does not occur when this same industry funds advocacy and educational efforts.

In light of this trend, a second logical initiative would be advancing industry-independent education. If the general public is to make treatment decisions that are fully informed of the entire scope of costs and benefits, they deserve a picture independent of those standing to profit from those treatments. As an alternative to classes from industry-linked groups, some scholars and practitioners are beginning to create more independent resources (e.g., classes, books) directed at a public audience (e.g., Gordon, 2008; Williams, Teasdale, Zindel, & Kabat-Zinn, 2007).

Since many professional educational opportunities are also linked to industry, the same need for industry-independent education exists for doctors and therapists as well. Florida International University received funding in 2007 from the General Consumer and Prescriber Grant Program, administered by the Special Committee of State Attorneys, to develop a broad, research-based curriculum on psychotropic medications for nonmedical helping professionals. In order to reduce financial conflicts of interest, all investigators and consultants on the project agreed to forego any level of industry funding. The stated goals of the CriticalThinkRx project, headed by David Cohen of Florida International University, are to “sharpen critical skills of mental health and child welfare professionals assessing and practicing with children and adolescents who may be medicated with psychiatric drugs” and “close gaps between research and practice to maximize opportunities to help clients and avoid harm” (CriticalThinkRx website: <http://www.criticalthinkrx.org>). Compared to other resources available, the quality and fairness of the resulting educational program are striking.

More Comprehensive Informed Consent

Based on the controversies detailed earlier, our final recommendation revolves around treatment itself. After reviewing some of these research limitations cited in this article with a medical doctor colleague recently, we were asked, “Well, what do you propose should happen?” One of us answered, “Let people know about the contro-

versies.” His response surprised us: “Oh, no! You don’t want to do that. Families and individuals facing emotional problems shouldn’t be asked to hear all this.” He went on to explain that it was doctors and researchers who were best trusted to be able to handle and navigate the full scope of findings, with such complexities better left to their deliberation.

With all due respect to our doctor friend, we strongly suggest that withholding relevant information about treatment is both unwise and unethical. In our view, it is the families exploring treatment for their child or adolescent—perhaps more than anyone else—that deserve to know about any controversies and constraints associated with a particular treatment. In contrast to an individual or family who is led to believe that science has incontrovertibly proven the worth of a particular intervention, those made aware of real complexities can, we would argue, be significantly helped to think through their own treatment decisions. While clear research results would be ideal, in the absence of such clarity families at least deserve to be made aware of the full complexity of the issues at hand. In medicine, these tenets are congruent with both a model of shared decision making and an emphasis on evidence-based practice (Edwards & Elwyn, 2009).

In addition to the enhanced public educational opportunities already described, one way to convey more comprehensive information about medical and therapeutic interventions is through a bolstered informed consent process. A more comprehensive informed consent process could be a significant advance in assisting families in these deliberations. As in other realms, the absence of a thorough and complete disclosure of known risks and benefits naturally constitutes a barrier to authentic choice (Gambrell, 2001). Of course, the ability of professionals to create such informed consent documents relies on the fairness and thoroughness of their own education, a variable that, by some measures, cannot be taken for granted (e.g., Lacasse & Gomory, 2003).

In summary, alongside more thoughtful and comprehensive research, an insistence on more comprehensive education to the broader public, to professional helpers, and to individual patients is crucial to facilitating a more open and healthy collective deliberation about youth treatment. More than simply improving overall discussion about intervention outcomes, it is the impact upon real-life outcomes for children—their depth, their scope, and their power—that would reflect the full potential of such a shift.

References

- Adams, K. B., Matto, H. C., & LeCroy, C. W. (2009). Limitations of evidence-based practice for social work education: Unpacking the complexity. *Journal of Social Work Education, 45*(2), 1–22.
- Andersen, S. L., & Navalta, C. P. (2004). Altering the course of neurodevelopment: A framework for understanding the enduring effects of psychotropic drugs. *International Journal of Developmental Neuroscience, 22*, 423–440.
- Antonuccio, D. O., Burns, D. D., & Danton, W. G. (2002). Antidepressants: A triumph of marketing over science? *Prevention & Treatment, 5*(1), article 25.
- Antonuccio, D. O., Danton, W. O., & McClanahan, T. M. (2003). Psychology in the prescription era: Building a firewall between marketing and science. *American Psychologist, 58*, 1028–1043.
- Behrens, E. N., & Satterfield, K. M. (2007). Longitudinal family and academic outcomes in residential programs: How students function in two important areas of their lives. *Journal of Therapeutic Schools & Programs, 2*(1), 95–105.
- Berry, M., Brandon, M., Chaskin, R., Fernandez, E., Grietens, H., Lightburn, A., . . . Zeira, A. (2006). Identifying sensitive outcomes of interventions in community-based centres. *International Journal of Child & Family Welfare, 9*, 1–10.
- Bickman, L., Lambert, E. W., Andrade, A. R., & Penalzoza, R. (2000). The Fort Bragg continuum of care for children and adolescents: Mental health outcomes over 5 years. *Journal of Consulting and Clinical Psychology, 68*(4), 710–716.

- Bodenheimer, T. (2000). Uneasy alliance—Clinical investigators and the pharmaceutical industry. *New England Journal of Medicine*, 342(20), 1539–1544.
- Bridge, J. A., Iyengar, S., Salary, C. B., Barbe, R. P., Birmaher, B., Pincus, H. A.,... Brent, D. A. (2007). Clinical response and risk for reported suicidal ideation and suicide attempts in pediatric antidepressant treatment: A meta-analysis of randomized controlled trials. *JAMA*, 297(15), 1683–1696.
- Butler, S. M. (2006, June 30). *Implementing and evaluating multi-systemic therapy with families of persistent young offenders in the UK: Initial reflections and observations from a longitudinal trial*. Paper presented at Jack Tizard Memorial Lecture and Day Conference on Evidence-Based Psychotherapies in Child and Adolescent Mental Health Practice, Royal College of Surgeons, London, UK.
- Cairns, B. (2004). *Fostering attachments: Long-term outcomes in family group care*. London, UK: BAAF.
- Campbell, E. G., Weissman, J. S., Ehringhaus, S., Rao, S. R., Moy, B., Feibelmann, S., & Goold, S. D. (2007). Institutional academic-industry relationships. *JAMA*, 298, 1779–1786.
- Cervenka, K. A., Dembo, R., & Brown, C. H. (1996). A family empowerment intervention for families of juvenile offenders. *Aggression and Violent Behavior*, 1(3), 205–216.
- Chrismser, S. S., Hodge, S. T., & Saintil, D. (Eds.). (2006). Assessing “No Child Left Behind”: Perspectives and prescriptions [Special issue]. *Harvard Educational Review*, 76, Winter.
- Cohen, D., Hughes, S., & Jacobs, D. J. (2009). The deficiencies of drug treatment research: The case of Strattera. In S. Timimi & J. Leo (Eds.), *Rethinking ADHD: From brain to culture* (pp. 313–333). New York and London, UK: Palgrave Macmillan.
- Crawford, M. J., Weaver, T., Rutter, D., Sensky, T., & Tyrer, P. (2002). Evaluating new treatments in psychiatry: The potential value of combining qualitative and quantitative research methods. *International Review of Psychiatry*, 14(1), 6–11.
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches*. New York: SAGE.
- Dantzig, T. (1930). *Number: The language of science*. New York: Macmillan.
- Delate, T., Gelenberg, A. J., Simmons, V. A., & Motheral, B. (2004). Trends in the use of antidepressants in a national sample of commercially insured pediatric patients, 1998 to 2002. *Journal of Psychiatric Services*, 55, 387–391.
- Denzin, N. K., & Lincoln, Y. (2005). *The SAGE handbook of qualitative research* (3rd ed.). London, UK: SAGE.
- Drisko, J. (1997). Strengthening qualitative studies and reports: Standards to promote academic integrity. *Journal of Social Work Education*, 33(1), 185–197.
- Drisko, J. (2008). How is qualitative research taught at the master's level? *Journal of Social Work Education*, 44(1), 85–101.
- Edwards, A., & Elwyn, G. (2009). *Shared decision-making in health care: Achieving evidence-based patient choice* (2nd ed.). New York: Oxford University Press.
- Emslie, G. J., Rush, A. J., Weinberg, W. A., Kowatch, R. A., Hughes, C. W., Carmody, T., & Rintelmann, J. (1997). A double-blind, randomized, placebo-controlled trial of fluoxetine in children and adolescents with depression. *Archives of General Psychiatry*, 54(11), 1031–1037.
- Flew, A. (1995). *Thinking about social thinking*. Amherst, NY: Prometheus.
- Floersch, J., Townsend, L., Longhofer, J., Munson, M., Winbush, V., Kranke, D.,... Findling, R. (2009). Adolescent experience of psychotropic treatment. *Transcultural Psychiatry*, 46(1), 157–179.
- Frankel, R. (1999). Standards of qualitative research. In B. Crabtree & W. Miller (Eds.), *Doing qualitative research* (2nd ed.; pp. 333–346). Los Angeles, CA: SAGE.
- Gambrill, E. (1999). Evidence-based practice: An alternative to authority-based practice. *Families in Society: The Journal of Contemporary Human Services*, 80(4), 341–350.
- Gambrill, E. (2001). Social work: An authority-based profession. *Research on Social Work Practice*, 11(2), 166–175.
- Gambrill, E. (2010). Evidence-informed practice: Antidote to propaganda in the helping professions? *Research on Social Work Practice*, 20, 302–320.
- Gilgun, J. (1994). A case for case studies in social work research. *Social Work*, 39, 371–380.
- Goerge, R., Wulczyn, F., & Fanshel, D. (1994). A foster care research agenda for the '90s. *Child Welfare*, 73(5), 525–549.
- Gordon, J. S. (2008). *Unstuck: Your guide to the seven-stage journey out of depression*. New York: Penguin Press.
- Harris, G. (2009, October 22). Drug makers are advocacy group's biggest donors. *The New York Times*, A23.
- Harris, G., & Carey, B. (2008, June 8). Researchers fail to reveal full drug pay. *The New York Times*, A1.
- Healy, D. (2006). Manufacturing consensus. *Culture, medicine, and psychiatry*, 30(2), 135–156.
- Henggeler, S. W., Rowland, M. D., Randall, J., Ward, D. M., Pickrel, S. G., Cunningham, P. B.,... Santos, A. B. (1999). Home-based multisystemic therapy as an alternative to the hospitalization of youths in psychiatric crisis: Clinical outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*, 38, 1331–1339.
- Henggeler, S. W., & Schoenwald, K. (1994). Boot camps for juvenile offenders: Just say no. *Journal of Child and Family Studies*, 3, 243–248.
- Hess, J. Z., Bjorklund, E., Preece, N., & Draper, S. (under review). Poison apples, big bad wolves and other ‘happy ending’ spoilers: Challenges to enduring change following youth residential treatment.
- Journal of Therapeutic Schools and Programs. Hinshaw, S. (2007). Moderators and mediators of treatment outcome for youth with ADHD: Understanding for whom and how interventions work. *Journal of Pediatric Psychology*, 32(6), 664–675.
- Hoagwood, K. E. (2005). Family-based services in children's mental health: A research review and synthesis. *Journal of Child Psychology and Psychiatry*, 46(7), 690–713.
- Hughes, S., & Cohen, D. (2009). A systematic review of long-term studies of drug treated and non-drug treated depression. *Journal of Affective Disorders*, 118(1), 9–18.
- Hughes, S., & Cohen, D. (2010). Understanding the assessment of psychotropic drug harms in clinical trials to improve social workers' role in medication monitoring. *Social Work*, 55(2), 105–111.
- Jacobs, D. H. (1999). A close and critical examination of how psychopharmacotherapy research is conducted. *Journal of Mind and Behavior*, 20(3), 311–349.
- Jureidini, J. N., & Jureidini, E. S. (2008, February). 226 papers/letters that cite Keller et al.'s report of GlaxoSmithKline's Study 329 of antidepressants for adolescents. *Healthy Skepticism International News*. Retrieved June 1, 2010, from http://www.healthyskepticism.org/global/news/extra/keller_citations_update
- Jureidini, J. N., McHenry, L. B., & Mansfield, P. R. (2008). Clinical trial and drug promotion: Selective reporting of Study 329. *International Journal of Risk and Safety in Medicine*, 20, 73–81.
- Karp, D. A. (1997). *Speaking of sadness: Depression, disconnection, and the meanings of illness*. London: Oxford University.
- Karp, D. A. (2006). *Is it me or my meds? Living with antidepressants*. Cambridge, MA: Harvard University Press.
- Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry*, 44, 1116–1129.
- Keller, M. B., Ryan, N. D., Strober, R. G., Klein, S. P., Kutcher, B., Birmaher, B.,... McCafferty, J. P. (2001). Efficacy of paroxetine in the treatment of adolescent major depression: A randomized, controlled trial. *Journal of the American Academy of Adolescent and Child Psychiatry*, 40, 762–772.
- Knorth, E. J., Harder, A. T., Zandberg, T., & Kendrick, A. J. (2008). Under one roof: A review and selective meta-analysis on the outcomes of residential child and youth care. *Children and Youth Services Review*, 30(2), 123–140.
- Knudsen, P., Hansen, E. H., Traulsen, J. M., & Eskildsen, K. (2002). Changes in self-concept while using SSRI antidepressants. *Qualitative Health Research*, 12(7), 932–944.
- Lacasse, J. R. (2005). Consumer advertising of psychiatric medications biases the public against non-pharmacological treatment. *Ethical Human Psychology and Psychiatry*, 7(3), 175–179.
- Lacasse, J. R., & Gomory, T. (2003). Is graduate social work education promoting a critical approach to mental health practice? *Journal of Social Work Education*, 39(3), 383–408.
- Lacasse, J. R., & Leo, J. (2010). Ghostwriting at elite academic medical centers in the United States. *PLoS Medicine*, 7(2), e1000230. doi: 10.1371/journal.pmed.1000230
- Leo, J., & Lacasse, J. R. (2008). The media and the chemical imbalance theory of depression. *Society*, 45, 35–45.
- Lewis, R. E. (2004). The effectiveness of Families First services: An experimental study. *Children and Youth Services Review*, 27, 499–509.
- Lexchin, J., Bero, L. A., Djulbegovic, B., & Clark, O. (2003). Pharmaceutical industry sponsorship and research outcome and quality: A systematic review. *BMJ*, 326, 1167–1176.
- Lietz, C. A. (2009). Examining families' perceptions of intensive in-home services: A mixed method study. *Children and Youth Services Review*, 31(12), 1337–1345.
- Lynam, D. R., Milicha, R., Zimmerman, R., Novak, S. P., Logan, T. K., Martin, C.,... Clayton, R. (1999). Project DARE: No effects at 10-year follow-up. *Journal of Consulting and Clinical Psychology*, 67(4), 590–593.
- McCardle, P., & Chhabra, V. (2004). The voice of evidence in reading research. In B. K. Keogh (Ed.), *The importance of longitudinal research for early intervention practices* (pp. 81–102). Baltimore, MD: Paul H. Brookes.

- McDonald, T. P., Allen, R. I., Westerfelt, A., & Piliavin, I. (1996). *Assessing the long-term effects of foster care: A research synthesis*. Washington, DC: CWLA Press.
- McKay, J. (2007). So you want to run an outcome study? The challenges to measuring adolescent residential treatment outcomes. *Journal of Therapeutic Schools & Programs*, 2(1), 63–80.
- Molloy, D., Woodfield, K., & Bacon, J. (2002). *Longitudinal qualitative research approaches in evaluation studies* (Working Paper No. 7). Department for Work and Pensions, London, UK. Retrieved October 10, 2010, from <http://research.dwp.gov.uk/asd/asd5/WP7.pdf>
- Moses, T., & Kirk, S. A. (2005). Psychosocial side effects of drug treatment of youth. In S. A. Kirk (Ed.), *Mental disorders in the social environment: Critical perspectives. Foundations of social work knowledge* (pp. 385–407). New York: Columbia University Press.
- National Registry of Evidence-Based Programs and Practices. (2009). *What is evidence-based?* U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration (SAMHSA). Retrieved May 5, 2010, from <http://www.nrepp.samhsa.gov/about-evidence.asp>
- Pachter, W. S., Fox, R. E., Zimbardo, P., & Antonuccio, D. O. (2007). Corporate funding and conflicts of interest: A primer for psychologists. *American Psychologist*, 62, 1005–1015.
- Padgett, D. (2008). *Qualitative methods in social work research*. Los Angeles, CA: SAGE.
- Pound, P., Britten, N., Morgan, M., Yardley, L., Pope, C., Daker-White, G., & Campbell, R. (2005). Resisting medicines: A synthesis of qualitative studies of medicine taking. *Social Science & Medicine*, 61, 133–155.
- Romyn, D. M., Allen, M. N., Boschma, G., Duncan, S. M., Edgecombe, N., Jensen, L. A., . . . Warnock, F. (2003). The notion of evidence in evidence-based practice by the nursing philosophy working group. *Journal of Professional Nursing*, 19(4), 184–188.
- Rosack, J. (2005, November 18). FDA panel rejects new rule for psychiatric drug testing. *Government New Psychiatric News, American Psychiatric Association*, 40(22), 1. Retrieved June 23, 2009, from <http://pn.psychiatryonline.org/content/40/22/1.1.full>
- Ruspini, E. (2008). Longitudinal research: An emergent method in the social sciences. In S. N. Hesse-Biber & P. Leavy (Eds.), *Handbook of emergent methods* (pp. 437–460). New York, NY: Guilford Press.
- Safer, D. J. (2006). Should selective serotonin reuptake inhibitors be prescribed for children with major depressive and anxiety disorders? *Pediatrics*, 118(3), 1248–1251.
- Simon, N. M., Safren, S. A., Otto, M. W., Sharma, S. G., Lanka, G. D., & Pollack, M. H. (2002). Longitudinal outcome with pharmacotherapy in a naturalistic study of panic disorder. *Journal of Affective Disorders*, 69(1/3), 201–208.
- Sismondo, S. (2008). How pharmaceutical industry funding affects trial outcomes: Causal structures and responses. *Social Science and Medicine*, 66(9), 1909–1914
- Sismondo, S. (2009). Ghosts in the machine: Publication planning in the medical sciences. *Social Studies of Science*, 39(2), 171–198.
- Slife, B. D., & Gantt, E. (1999). Methodological pluralism: A framework for psychotherapy research. *Journal of Clinical Psychology*, 55(12), 1–13.
- Stoppard, J. M., & McMullen, L. M. (Eds.). (2003). *Situating sadness: Women and depression in social context*. New York, NY: New York University Press.
- Tanne, J. H. (2008). U.S. psychiatrist steps down after questions about drug company payments. *BMJ*, 337, a2088.
- Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine*, 358(3), 252–260. doi:10.1056/NEJMsa065779
- Venarde, D. F. (1999). *Medication and meaning: Psychotherapy patients' subjective experiences of taking selective serotonin reuptake inhibitors (SSRIs)* (Unpublished doctoral dissertation). State University of New Jersey, Piscataway.
- Verbeek-Heida, P. M., & Mathot, E. F. (2006). Better safe than sorry—why patients prefer to stop using selective serotonin reuptake inhibitor (SSRI) antidepressants but are afraid to do so: Results of a qualitative study. *Chronic Illness*, 2(2), 133–142.
- Whitaker, R. (2010). *Anatomy of an epidemic: Magic bullets, psychiatric drugs, and the astonishing rise of mental illness in America*. New York: Crown.
- Williams, M., Teasdale, J., Zindel, S., & Kabat-Zinn, J. (2007). *The mindful way through depression: Freeing yourself from chronic unhappiness*. New York, NY: Guilford Press.
- Whitemore, E., Ford, M., & Sack, W. H. (2003). Effectiveness of day treatment with proctor care for young children: A four-year follow-up. *Journal of Community Psychology*, 31(5), 459–468.

Jacob Z. Hess, PhD, research director, Utah Youth Village. **Jeffrey R. Lacasse**, PhD, assistant professor, School of Social Work, College of Public Programs, Arizona State University. Correspondence: jhess@youthvillage.org; 5800 South Highland Drive, Salt Lake City, UT 84121.

Manuscript received: February 22, 2010

Revised: July 2, 2010

Accepted: July 6, 2010