

Where is the Evidence for “Evidence-Based” Therapy?*

Jonathan Shedler

“Evidence-based therapy” has become quite the catchphrase. The term “evidence-based” comes from medicine. It gained attention in the 1990s and was, at the time, a call for critical thinking. It reflected the recognition that “we’ve always done it this way” is not a good enough reason to keep doing something. Medical decisions should reflect clinical judgment, patients’ values and preferences, and relevant scientific research.

But “evidence-based” has come to mean something very different in the psychotherapy world. The term has been appropriated to promote a particular ideology and agenda. It has become a code word for *manualised* treatment—most often, brief, highly-structured cognitive behavioural therapy (CBT). “Manualised” means the therapy is literally conducted by following an instruction manual. The treatment may be pre-scripted in a way that leaves little room for understanding patients as individuals.

Behind the “evidence-based” therapy movement lies what I will call the “master narrative”, a narrative that increasingly dominates the mental health landscape. The master narrative goes something like this: “In the dark ages, therapists practiced untested, unscientific therapy. Science shows that evidence-based therapies are superior.” This narrative has become a justification for all-out attacks on traditional (i.e., psychodynamic) therapy—that is, psychotherapy that fosters self-understanding and insight in the context of a meaningful, ongoing therapy relationship.

* © 2015 by Jonathan Shedler, PhD.

Here is a small taste of what proponents of “evidence-based” therapy have been saying in public: “The empirically supportive psychotherapies are still not widely practiced. As a result, many patients *do not have access to adequate treatment*” (Hollon et al., 2002; emphasis added). Notice the linguistic sleight-of-hand: If it is not an “evidence-based” (i.e., manualised) treatment, it is inadequate. Walter Mischel of Columbia University wrote, “The disconnect between what clinicians do and what science has discovered is an *unconscionable embarrassment*” (Mischel, 2008; emphasis added).

When this master narrative gets into the media, things get worse. The venerable *Washington Post* ran an article titled, “Is your therapist a little behind the times?” (Baker et al., 2009). It likened traditional (read, psychodynamic) therapy to pre-scientific medicine when “healers commonly used ineffective and often injurious practices such as blistering, purging and bleeding”. *Newsweek* sounded a similar note in an article titled, “Ignoring the evidence: why do psychologists reject science?” (Begley, 2009).

Notice how a form of McCarthyism enters the picture. Because proponents of brief, manualised therapies have appropriated the term “evidence-based” for their own use, it becomes difficult to have an intelligent conversation about what constitutes good therapy—to question claims for “evidence-based” therapy is to risk the accusation of being “anti-science”.

You may be thinking that in light of the strong claims for “evidence-based” therapies—and the denigration of psychodynamic or insight-oriented therapies—there must be amazingly strong scientific evidence for their benefits. There is not. There is a yawning chasm between what we are told research shows and what research actually shows.

What empirical research really shows is that “evidence-based” therapies are *ineffective* for most people most of the time. In Part I, I discuss what empirical research really shows. In Part 2, I take a closer look at some troubling practices in “evidence-based” therapy research.

Part I: what research really shows

Research shows that “evidence-based” therapies are *weak* treatments. Their benefits are trivial. Most patients do not get well. Even the trivial benefits do not last.

This may be very different from what you have heard elsewhere. You may be thinking, who is this guy? And why should I believe him? I will revisit this question at the conclusion. I am not asking you to believe me. That is why I will be referencing primary sources.

The gold standard of evidence in “evidence-based” therapy research is the randomised controlled trial. Patients with a certain diagnosis are randomly assigned to either a treatment or control group, and the study compares the two groups.

The mother of all randomised controlled trials for psychotherapy is the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program. It was the first really large multi-site research project investigating what are now called “evidence-based” therapies. The study included three active treatments: manualised CBT, manualised interpersonal therapy (IPT), and antidepressant medication. The control group got a placebo pill and clinical management, but not psychotherapy. The study was initiated in the mid-1970s and the first major publications started coming out around 1990.

For the past twenty-five years, we have been told that this study showed that CBT, IPT, and antidepressant medication are “empirically validated” treatments for depression. We have been told that these treatments were scientifically proven to be powerful and effective. I will focus on CBT because that is what gets the most attention and, of course, it is the theme of this conference.

The claims for the benefits of CBT were based on the finding that CBT was “statistically significantly” more effective than the placebo control group. “Statistically significant” does not mean what most people think. Set aside preconceptions about the word *significant* and consider, instead, the actual difference in the NIMH study between the CBT group and the control group that got the sugar pill.

The primary outcome measure in the NIMH study was the 54-point Hamilton Depression Rating Scale. The difference between the CBT treatment group and the control group was 1.2 points.

The 1.2 point difference is trivial and clinically meaningless. It does not pass the “So what?” test. It does not pass the “Does it matter?” test. It does not pass the “Why should anyone care?” test.

How can there be such a mismatch between what we have been told versus what the study actually found? You may be wondering whether the original researchers did not present the data clearly. That is not the case. The first major research report from the NIMH study was published in 1989 in *Archives of General Psychiatry* (Elkin et al., 1989). The authors wrote: "There was limited evidence of the specific effectiveness of interpersonal psychotherapy and *none for cognitive behaviour therapy*" (emphasis added). That is what the original research report says.

In 1994, the principle investigator wrote a comprehensive review of what we had really learned from that study, titled "The NIMH Treatment of Depression Collaborative Research Program. Where we began and where we are" (Elkin, 1994).

Writing in careful academic language, the principle investigator stated: "What is most striking in the follow-up findings is the relatively small percentage of patients who remain in treatment, fully recover, and remain completely well throughout the 18-month follow-up period." The percentage was so small that it "raises questions about whether the potency of the short-term treatments for depression has been oversold" (Elkin, 1994, p. 131).

What was that percentage, actually? It turns out that only 24% of the patients in the study got well and stayed well. Another way of saying that is that about 75%—the overwhelming majority—did *not* get well. How can this be? We have been told the opposite for the last twenty-five years. We have been told that manualised CBT is powerful and effective.

We can now revisit the term *significant*. In the English language, the word *significant* is a synonym for important or meaningful. But that is not what the term means in statistics. In statistics, it is a technical term of art. It means that a research finding probably did not occur by chance. In the NIMH study, there was a 1.2 point difference between the CBT group and the control group. That is clinically meaningless—nobody would dispute that. But the difference was "statistically significant", meaning it probably did not occur by chance.

There are few other fields where people talk about "significance" instead of talking about actual benefits. When a researcher emphasises "statistical significance", something is being hidden. If there is a meaningful treatment benefit, one talks about that, not "significance". If we have a drug that is

effective in lowering blood pressure, we say the drug decreased blood pressure by so much. If we had an effective weight loss programme, we would say the average person in the programme lost twenty-five pounds, or thirty pounds, or whatever. If we had a drug that lowered cholesterol, we would talk about how much it lowered cholesterol.

We would not talk about “significant differences”. When researchers emphasise “statistical significance”, something is being hidden.

The NIMH findings were published more than twenty-five years ago. Surely, research findings for CBT must have improved over time. So let us jump ahead to the most recent state-of-the-art randomised controlled trial for depression (Driessen et al., 2013). This recent study included 341 depressed patients who were randomly assigned to sixteen sessions of manualised CBT or sixteen sessions of manualised psychodynamic therapy. This study was published in 2013 in the *American Journal of Psychiatry*.

The authors wrote, “One notable finding was that only 22.7% of the patients achieved remission” (Driessen et al., 2013, p. 1047). They continued, “Our findings indicate that a substantial proportion of patients . . . require more than time-limited therapy to achieve remission.” In other words, about 75% of patients did not get well. It is essentially the same finding reported in the NIMH study a quarter of a century ago.

The appropriate conclusion to be drawn from both of these major studies is that brief, manualised treatments are *ineffective* for most depressed patients most of the time.

So I have reviewed the earliest major study and the most recent. What about all the research in between? The results are largely the same. The research is summarised in a review paper by lead author Drew Westen (Westen et al., 2004). The paper is a detailed, comprehensive literature review of manualised CBT for depression and anxiety disorders.

The researchers found that the average patient who received manualised CBT for depression remained clinically depressed after treatment (with an average Beck Depression Inventory score of eleven). What about other conditions besides depression? How about panic disorder? Panic may be the one condition for which brief, manualised CBT work *best*. But the average

patients who received “evidence-based” treatment for panic disorder still had panic attacks almost weekly, and still endorsed four of out of seven symptoms listed in the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. These patients were not getting well either.

Another finding was that the treatment benefits of manualised “evidence-based” therapies evaporated quickly. Treatment outcome is usually measured the day treatment ends. But when patients are followed over time, the benefits evaporate. The majority of patients who receive an “evidence-based” treatment—more than 50%—seek treatment again within six to twelve months for the same condition. And it would be a mistake to conclude that those who do not seek further treatment are well. Some may have gotten well. Others may have concluded that psychotherapy is unhelpful and given up on it.

Part 2: a closer look at research practices

In this section, I will address some of the research practices behind claims for manualised, “evidence-based” therapies. I will address the following issues: First, most patients are never counted in the research studies. Second, the control groups are shams. Third, manualised, “evidence-based” therapy has *not* shown superiority to any other legitimate psychotherapy. Fourth, data are being suppressed.

Most patients never get counted

In the typical randomised controlled trial for “evidence-based” therapies, about two-thirds of the patients get excluded from the studies *a priori* (Westen et al., 2004). That is, they have the diagnosis and seek treatment, but because of the study’s inclusion and exclusion criteria, they are excluded from participation. Typically, the patients that get excluded are those who meet *DSM* criteria for more than one diagnosis, or have some form of personality pathology, or are considered unstable in some way, or who may be suicidal. In other words, the two-thirds that get excluded are the patients we treat in real-world practice.

So two-thirds of the patients who seek treatment get excluded before the study begins. Of the one-third that do get included,

about half show improvement. So we are now down to about 16% of the patients who initially sought treatment. But that is just patients who show “improvement”. If we consider the percentage of patients that actually *get well*, we are down to about 11% of those who originally sought treatment. If we consider the percentage that get well and *stay well*, we are down to roughly 5%.

In other words: scientific research demonstrates that “evidence-based” treatments are effective and have lasting benefits for approximately 5% of the patients who initially present for treatment. Here is another way to look at it: an iceberg represents all patients who seek treatment for a given condition—depression, generalised anxiety, whatever. The tip of the iceberg, above the water, represents the patients we hear about. All the rest—the huge part of the iceberg below the water—we do not hear about. They do not get counted. They are invisible.

Control groups are shams

Second point: the control group is usually a sham. What do I mean? I mean that “evidence-based” therapies are almost never compared to legitimate alternative therapies. The control group is usually a foil invented by researchers who are committed to demonstrating the benefits of CBT. In other words, the control group is a pseudo-treatment designed to fail.

A state-of-the-art, NIMH-funded study of post-traumatic stress disorder (PTSD) provides a good example of what I mean by a sham control group (Gilboa-Schechtman et al., 2010). The study focused on “single incident” PTSD. The patients were previously healthy. They developed PTSD after experiencing a specific, identifiable trauma.

The study claims to compare psychodynamic therapy with a form of CBT called prolonged-exposure therapy. It claims to show that CBT is superior to psychodynamic therapy. This is what it says in the discussion section: “[CBT] was superior to [psychodynamic therapy] in decreasing symptoms of PTSD and depression, enhancing functioning . . . and increasing overall improvement” (Gilboa-Schechtman et al., 2010, p. 1040).

That is what was communicated to the media, the public, and policy makers. If you read the fine print and do a little homework, things look very different.

Who were the therapists who provided the “psychodynamic” treatment in this study? Were they experienced, qualified, psychodynamic therapists? No. *It turns out they were graduate students.* They received exactly two days of training in psychodynamic therapy *from another graduate student*—a graduate student in a research laboratory committed to CBT. In contrast, the therapists who provided CBT were trained for five days by the developer of this form of therapy, world-famous clinician and researcher Edna Foa. That is not exactly a level playing field.

But that was the least of the problems with the study. The so-called psychodynamic therapists were also *forbidden to discuss the trauma that brought the patient to treatment.* Imagine that—you come to treatment for PTSD because you have experienced a traumatic event, and your therapist is forbidden from discussing it with you. When patients brought up the trauma, the therapists were instructed to change the topic.

If anyone practiced like that in the real world, it could be considered malpractice. In research, that is considered a control group, and a basis for claims that CBT is superior to psychodynamic therapy.

“Superiority” of evidence-based therapy is a myth

In case you are thinking that the PTSD study is unusual—perhaps cherry-picked to make a point—that is not so. There is a comprehensive review of the psychotherapy research literature that addresses this exact issue (Wampold et al., 2011). It focussed on randomised controlled trials for both anxiety and depression. The researchers examined studies that claimed to compare an “evidence-based” therapy with an alternative form of psychotherapy.

The researchers examined over 2,500 abstracts. After closer examination, they winnowed that down to 149 studies that looked like they might actually compare an “evidence-based” therapy to another legitimate form of therapy. But when they finished, there were only fourteen studies that compared “evidence-based” therapy to a control group that received anything approaching real psychotherapy.

Many of the studies claimed to use a control group that received “treatment as usual”. But “treatment as usual” turned

out to be “predominantly ‘treatments’ that *did not include any psychotherapy*” (Wampold et al., 2011, p. 1310; emphasis added). I am not interpreting or paraphrasing. This is a direct quotation from the article. In other words, the so-called evidence-based treatments were not compared to other forms of psychotherapy; they were compared to *doing nothing*.

Alternatively, they were compared to control groups that received sham psychotherapy, where the therapists had their hands tied—as in the PTSD study I described above.

This literature review was published in a conservative scholarly journal, so the authors had to state their conclusions in careful academic language. They concluded: “Currently, there is insufficient evidence to suggest that transporting an evidence-based therapy to routine care that already involves psychotherapy will improve the quality of services” (Wampold et al., 2011, p. 1311).

In somewhat plainer English: “evidence-based” therapies have *not* shown greater effectiveness than any other legitimate psychotherapy. That is what the scientific literature actually shows. It is not just my opinion. It is also the formal scientific policy conclusion of the American Psychological Association (American Psychological Association, 2013).

Data are being suppressed

“Publication bias” bias is a well-known phenomenon in research. Publication bias refers to the fact that studies with positive results—that show the outcome desired by the investigators—tend to get published. Studies that fail to show desired results tend not to get published.

For this reason, the published research literature can provide a biased or skewed picture of actual research findings. There is a name for this phenomenon, it is called the “file-drawer effect”. For every published study with positive results, how many studies with negative results are hidden in file drawers? How can you prove there are file drawers stuffed with negative results? It turns out there is a way to do this. There are statistical methods to estimate how many unpublished studies have negative results that are hidden from view.

A team of researchers tackled this question for research on CBT for depression (Cuijpers et al., 2010). They found that the

published benefits of CBT are exaggerated by 75% due to publication bias. In other words, the actual benefit a patient is likely to receive is only about a quarter of what the research literature would lead you to believe.

How do you find out something like this? How can you know what is hidden in file drawers? You know by examining what is called a funnel plot. It sounds complicated, but it is actually a simple idea. Suppose you are doing a survey—“*Are people in the UK for or against some policy?*”—and you have a very small sample of only three people. The results can be all over the place. Depending on the three people you happen to draw for your sample, it may look like 100% are in favour or 100% are against. So when you have small sample sizes, you get quite a scatter, quite a range, of results. As sample sizes get larger, the findings stabilise and converge.

If you graph the findings—in this case, the relationship between sample size and treatment benefit—you get a plot that looks like a funnel (Figure 1, left graph). Studies with smaller sample sizes show more variability in results, and studies with larger sample sizes tend to converge on more similar values. That is what it should look like if data are not being hidden.

In fact, what it looks like is something like the graph on the right (Figure 1, right graph). The data points that are supposed to be in the lower left area of the graph are missing.

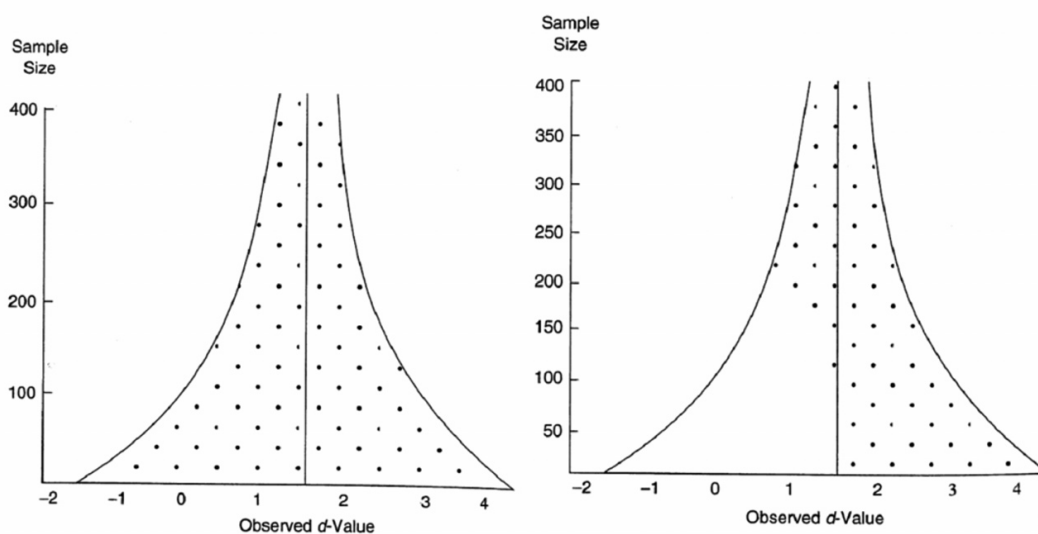
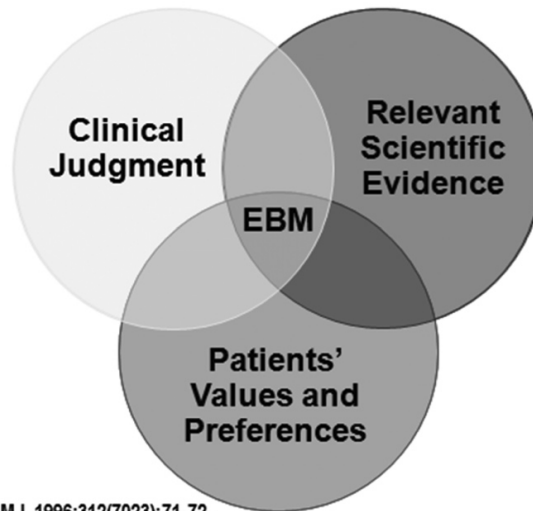


Figure 1

What Is Evidence-Based Medicine?



Sackett DL, et al. *BMJ*. 1996;312(7023):71-72.

Figure 2

What “evidence-based” is supposed to mean

What is “evidence-based” really supposed to mean? I noted earlier that the term originated in medicine. Evidence-based medicine (EBM) was supposed to represent the convergence or intersection of 1) relevant scientific evidence, 2) patients’ values and preferences, and 3) the experience and clinical judgment of the practitioner (Figure 2).

What has happened to these ideas in the field of psychotherapy? “Relevant scientific evidence” no longer matters, because proponents of so-called evidence-based therapies ignore evidence for therapy that is not pre-scripted, manualised therapy. In 2010, I published an article in *American Psychologist* titled, “The efficacy of psychodynamic psychotherapy” (Shedler, 2010). The article demonstrated that the benefits of psychodynamic therapy are at least as large as those of so-called evidence-based therapy—and moreover, the benefits of psychodynamic therapy *last*. Proponents of “evidence-based” therapy typically disregard this evidence.

“Patients’ values and preferences” also do not matter, because patients are not being informed and offered meaningful choices. They may be offered only brief manualised treatment and told it is the “gold standard” of care. “Clinical judgment” also no

longer matters, because clinicians are expected to follow manuals rather than exercise meaningful clinical judgment. They are being asked to function as technicians, not clinicians.

One could argue that “evidence-based”, as the term is now used with respect to psychotherapy, is a perversion of every founding principle on which the concept of evidence-based medicine rests.

Conclusion

Who is this Shedler guy? Why should I believe him? Everything I have read from every respectable scholarly source seems to be at odds with what he just told us.

Why should you believe me? You should not believe me. You should not take my word for any of this, or anyone else’s word. I will leave you with three simple things to do, starting today, that will help you sift truth from hyperbole. When somebody makes a claim for the benefit of a treatment, any treatment, follow these three simple steps:

Step 1: say, “Show me the study.” Give me a reference, a citation, a PDF. Put the study into my hands. Sometimes it does not exist.

Step 2: if the study does exist, read it—especially the fine print.

Step 3: draw your own conclusion. Ask yourself: do the actual methods and findings of this study justify the claim I just heard?

If you make a practice of following these steps, you may make some shocking discoveries.

References

- American Psychological Association (2013). Recognition of psychotherapy effectiveness. *Psychotherapy, 50*: 102–109. Available at: www.apa.org/about/policy/resolution-psychotherapy.aspx
- Baker, T., McFall, R., & Shoham, V. (2009). Is your therapist a little behind the times? *The Washington Post*, 15 November, 2009.
- Begley, S. (2009). Ignoring the evidence: why do psychologists reject science? *Newsweek, 154*(15): 30.
- Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). Efficacy of cognitive-behavioural therapy and other

- psychological treatments for adult depression: meta-analytic study of publication bias. *British Journal of Psychiatry*, 196: 173–178. doi: 10.1192/bjp.bp.109.066001.
- Driessen, E., Van, H. L., Don, F. J., Peen, J., Kool, S., Westra, D., Hendriksen, M., Schoevers, R. A., Cuijpers, P., Twisk, J. W. R., & Dekker, J. J. M. (2013). The efficacy of cognitive-behavioral therapy and psychodynamic therapy in the outpatient treatment of major depression: a randomized clinical trial. *American Journal of Psychiatry*, 170: 1041–1050.
- Elkin, I. (1994). The NIMH Treatment of Depression Collaborative Research Program. Where we began and where we are. In: A. E. Bergin & S. L. Garfield (Eds.), *Handbook of Psychotherapy and Behavior Change* (4th edn.) (pp. 114–139). New York: Wiley.
- Elkin, I., Shea, M., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., Glass, D. R., Pilkonis, P. A., Leber, W. R., Docherty, J. P., Fiester, S. J., & Parloff, M. B. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry*, 46(11): 971–982. doi:10.1001/archpsyc.1989.01810110013002.
- Gilboa-Schechtman, E., Foa, E. B., Shafran, N., Aderka, I. M., Powers, M. B., Rachamim, L., Rosenbach, L., Yadin, E., & Apter, A. (2010). Prolonged exposure versus dynamic therapy for adolescent PTSD: a pilot randomized controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 49: 1034–1042.
- Hollon, S. D., Thase, M. E., Markowitz, J. C. (2002). Treatment and prevention of depression. *Psychological Science in the Public Interest*, 3: 39–77.
- Mischel, W. (2008). Editorial: connecting clinical practice to scientific progress. *Psychological Science in the Public Interest*, 9(2): i–ii.
- Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *American Psychologist*, 65: 98–109. doi: 10.1037/a0018378.
- Wampold, B. E., Budge, S. L., Laska, K. M., Del Re, A. C., Baardseth, T. P., Fluckiger, C., Minami, T., Kivlighan, D. M., & Gunn, W. (2011). Evidence-based treatments for depression and anxiety versus treatment-as-usual: a meta-analysis of direct comparisons. *Clinical Psychology Review*, 31: 1304–1312.
- Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin*, 130: 631–663. doi:10.1037/0033-2909.130.4.631.