49 A MAJOR ERROR IN THE EVALUATION OF PSYCHOLOGICAL TREATMENTS FOR ANXIETY

David H. Barlow
Boston University

A number of years ago, with the help of a large grant from the National Institute of Mental Health, my clinical research team and I were evaluating new treatments for generalized anxiety disorder (GAD), which consists of a debilitating and almost continuous experience of anxiety characterized by intense worry about any number of different situations and a state of very unpleasant arousal and muscle tension (Barlow, Rapee, & Brown, 1992). These individuals are always anticipating the next disaster that might be around the corner and so are in a continual state of anxious readiness that often results in substantial fatigue by the end of the day. The problem is, the worry and anxiety never really dissipate but switch from one topic to another. We treated one group with what at the time was a new application of cognitive therapy directed at their attributions and appraisals of the severity of the threat so that they could better cope with whatever stressful situations they encountered. In a second group, we evaluated the effects of deep muscle relaxation, which was focused more on the high levels of arousal and tension these individuals were feeling. Finally, we combined those two interventions, a combination we hypothesized would be the most effective since it addressed both facets of GAD. We randomly assigned 65 carefully diagnosed patients to one of these three treatment conditions and compared the outcomes to those from a waitlist control group who were assessed but did not receive treatment. Patients in the waitlist control group were all treated after the study was over. This type of experiment is called a "randomized clinical trial," which is the gold standard for evaluating the effects of drug or psychological treatments.

It is very important to establish in these clinical trials not only whether the treatment is better than no therapy at all but also whether is it better than an alternative therapy. These kinds of comparisons help us isolate the treatment's active ingredients or what is unique about the treatment as the important factor responsible for improvement, rather than alternative sources of explanation for any improvement, such as placebo effects, patient expectancies, or the natural healing process that occurs with the passage of time. In other words, is it valid to say that the treatment itself is effective? We call this type of validity "internal validity" because it refers

to ruling out factors that may influence improvement other than the treatment itself.

But there is another kind of validity that is important in these types of trials, which we call "external validity." This refers to the extent to which an internally valid intervention is effective in different settings or under different circumstances from those where it was tested and how easily can it be disseminated and implemented in those settings. To take an extreme example, if doctors develop a treatment that is very expensive, and painful as well, many patients would be reluctant to undergo this treatment, particularly if an alternative treatment that was cheaper and less painful was nearly as good. There are other factors that determine if a treatment is externally valid in addition to acceptability and cost-effectiveness, including how generally effective the treatment is with different types of patients. So, if a treatment is not acceptable to patients, is too expensive, or works only with a few people, we would conclude it's not externally valid.

As you can probably guess, treating this many patients in a clinical trial over a period of three or four years in a research setting is expensive, since all patients receive free treatment, and also we must pay for the cost for research and personnel carrying out the roles as therapists and evaluators, and so on, so it's important to get it right. But it wasn't until the trial was over that we realized our mistake. Basically, the study was internally valid in that it was capable of ruling out many alternative sources of explanation for change, other than the treatment itself, but we had failed to recognize the substantial threats to external validity. This was because we thought it very important from the point of view of internal validity to make sure that all patients got exactly the same "dose" of treatment. So they each saw the therapist for same amount of time to make sure there weren't any differences in time spent with the therapist that could have introduced therapeutic healing factors not directly associated with the treatment procedures. For this reason, all patients received 15 sessions of treatment no matter how long the treatment actually was supposed to take. Thus, the treatment consisting of relaxing exercises alone might typically take five or six sessions, but we continued work with these patients on relaxation for 15 sessions. The combined treatment, on the other hand, covering both relaxation exercises and cognitive therapy, really should have taken a bit longer to be delivered in an intelligible and reasonable manner, but we insisted on fitting everything into 15 sessions.

What happened was that patients receiving relaxation exercises alone became bored with the treatment after five or six sessions and dropped out to a significantly greater extent than from the other treatments, somewhat confounding our ability to evaluate the results. In the condition where people received the combined treatment, on the other hand, only about half as many patients responded positively to treatment as in the other two conditions, and this was the group we expected to do the best! When we looked closely at what we had done, the problem seemed to be that we were attempting to cram so much information into each session that the patient never had a chance to really absorb the treatment sufficiently.

Since that time, the community of clinical scientists doing research on psychological and drug treatments has learned a great deal about the proper ways to conduct

these clinical trials. We are always attempting to learn from each other's work and from each other's mistakes, so that the quality of the science we undertake improves, and we can make better determinations of which treatments are truly effective for patients or are "evidenced based." What we and others learned from this rather serious mistake is that, while it's extremely important to determine that the treatment is working because of the specific components of the treatment and not some unrecognized other factors (internal validity), it is just as important to evaluate treatments in the way they are likely to be delivered in frontline clinical settings such that they would feasible and acceptable to patients, as well as cost-effective (external validity.) Only by being sensitive to both types of validity can we conduct successful clinical trials.

REFERENCE

Barlow, D. H., Rapee, R. M., & Brown, T. A. (1992). Behavioral treatment of generalized anxiety disorders. *Behavior Therapy, 23, 551–570.*

CRITICAL THINKING QUESTIONS

1. Why could a clinical trial prove that a treatment is efficacious but still not have it adopted by clinicians and healthcare policymakers?
2. It has been established that deep relaxation exercises work for many anxiety disorders, but why didn't it work in this clinical trial?
3. Why is external validity as important as internal validity?