Behaviour Research and Therapy 90 (2017) 147–158

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

Applying new evidence standards to youth cognitive behavioral therapies – A review

Leslie R. Rith-Najarian*, Alayna L. Park, Tina Wang, Ana I. Etchison, Denise A. Chavira, Bruce F. Chorpita

Department of Psychology, University of California, Los Angeles 1285 Franz Hall, Los Angeles, CA, 90095, USA

ARTICLE INFO

Article history:
Received 2 February 2016
Received in revised form
11 November 2016
Accepted 16 December 2016
Available online 23 December 2016

Keywords:
Youth
CBT
Evidence-based
Standards
Anxiety
Treatment selection

ABSTRACT

This review included 136 published randomized controlled trials (RCTs) of youth cognitive behavioral therapy (CBT) treatments. We aimed to test the premise that evidence-based youth treatments can be better differentiated from each other by applying more nuanced standards of evidence. Accordingly, we applied three standards to this article sample to determine how many treatments produced significant results: (a) on multiple target symptom measures, (b) at follow-up, and/or (c) against an active comparison group. We identified how many trials met standards individually and in combination. Although 87 of the 136 articles produced at least one significant treatment result at post-assessment, the subsets of "passing" articles were smaller and varied for any one of our three standards, with only 11 articles (8%) meeting all three standards simultaneously. Implications are discussed regarding the definition of "evidence-based," the need for multi-parameter filtering in treatment selection and clinical decision making, and future directions for research. We ultimately argue the value in assessing youth treatments for different types of evidence, which is better achieved through dynamic sets of standards, rather than a single approach to assessing general strength of evidence.

1. Introduction

Given that more than 1 in 10 youth globally are estimated to experience mental health disorders (Polanczyk, Salum, Sugaya, Caye, & Rohde, 2015), research on child and adolescent mental health is of great importance. Youth mental health disorders have a persisting effect on psychological, economic, and social outcomes into adulthood (Patel, Flisher, Hetrick, & McGorry, 2007), indicating the need for early intervention. In response to this need, more than 800 randomized trials have developed and tested youth psychotherapy treatments (PracticeWise, 2016). Faced with this rich and ever-growing knowledge base, clinical scientists and practitioners are provided with more empirically supported interventions, but correspondingly must choose among these multiple options (Chorpita, Rotheram-Borus et al., 2011).

To better differentiate among treatments, one approach is to apply new, higher standards that can identify those treatments that produce relatively more benefits. Traditionally, a common standard for a treatment to achieve "evidence-based" or "empirically supported" status involves a randomized controlled trial (RCT): the treatment must outperform a comparison group at post-treatment on some outcome measure at a statistically significant level (cf. Chambless & Hollon, 1998; Silverman & Hinshaw, 2008). However, it has been argued that because the majority of treatments can demonstrate relatively more symptom reduction than an inactive control condition, this current standard has provided diminishing usefulness (Wampold, Imel, & Miller, 2009). Accordingly, we need to consider and evaluate additional criteria that could be used to select among many treatments with empirical support (Becker, Chorpita, & Daleiden, 2011). It is our contention that we might better differentiate youth treatment options if new standards specify particular *types* of evidence-based support. A single definition for "evidence-based" may no longer be a sufficient or informative enough metric for categorizing treatments.

There have been many proposed criteria and standards (some are already in widespread use) that further specify how treatments can demonstrate efficacy or strength of evidence. Example suggestions have included: following RCTs with effectiveness research in clinical settings with various populations, conducting cost-effectiveness research, using rigorous and appropriate data

* Corresponding author.
E-mail address: leslierrn@ucla.edu (L.R. Rith-Najarian).

http://dx.doi.org/10.1016/j.brat.2016.12.011
0005-7967/

analyses in two separate good quality group-design experiments, or conducting studies by independent research teams in different settings (Chambless & Hollon, 1998; Silverman & Hinshaw, 2008). From many potential parameters, we selected three candidate standards for our review: (a) significant results on multiple measures, (b) significant results at follow-up, and (c) significant results against an active comparison group.

Using multiple measures has long been a recommendation for psychological research, as results from multiple measures on the same construct represent not only reliable effects but also convergent validity (Campbell & Fiske, 1959; Hunsley & Mash, 2007). One of the problems with interpreting RCT results as demonstrating "empirical support" is that statistically significant effects are typically observed for only some of the measures used (Kazdin, 2008). Accordingly, a trial with multiple significant results may be more convincingly evidence-based relative to a trial with only one significant result.

Follow-up outcomes can demonstrate the durability of treatment effects, which has already been suggested as a candidate standard for strength of evidence (Becker et al., 2011; Chorpita, Daleiden, et al., 2011). Effects that last through follow-up can convey that changes made by a youth during treatment are internalized and maintained over a significant period of time (Weisz, McCarty, & Valeri, 2006). If a treatment produces durable effects, then the likelihood of a youth needing to reenter treatment may be reduced.

Use of active control groups (i.e., other treatment groups or at least minimally treated groups) can address external validity concerns by controlling for non-study care or by demonstrating effects relative to existing treatments (Freedland, Mohr, Davidson, & Schwartz, 2011). Problematically, a majority of youth psychotherapy RCTs have been found to use inactive control groups, and therefore they only control for time but not for attention or nonspecific common therapeutic factors (Weisz, Jensen-Doss, & Hawley, 2005). Comparing the effects of a treatment group to those of an another active group can increase confidence that any observed between-group superiority was likely due to the specifics of the treatment.

Given that these three criteria – multiple measures, follow-up, and active comparison groups – each convey valuable information, they serve as good example standards of different types of evidence. Ultimately, our intention in applying these three candidate standards is not to set a new bar for "evidence-based treatment", as different criteria may be of greater importance on a case-by-case basis (for researchers, clinicians, or clients). There are many other example standards – as we have partially reviewed – that could matter more for certain purposes. For this reason, we have only briefly overviewed the justification for our selection of standards, to refrain from overemphasizing their importance over other potential standards. However, these three standards are a meaningful starting point, and other standards could be tested in the future.

We selected cognitive behavior therapy (CBT) as our evidence base because CBT is often considered the treatment of choice for youth with psychological disorders (Benjamin et al., 2011; Hayes, Villatte, Levin, & Hildebrandt, 2011), and because it is a prime example of a proliferation of research, with at least 250 randomized trials examining CBT for youth (PracticeWise, 2016). Based on current "evidence-based" standards, youth CBT treatments have produced robust post-treatment effects across more diagnostic domains than has any other treatment modality for youth (Chorpita, Daleiden, et al., 2011). New standards for evaluating the evidence base of youth CBT can help make sense of the literature in aggregate, ultimately informing the selection process among many CBT treatment protocols. This approach may demonstrate that not all youth CBTs produce uniform types of evidence, consequently improving our understanding of the variable strengths of current youth CBT interventions.

We reviewed 136 published youth RCT studies that targeted a spectrum of diagnoses and included at least one CBT group. First, we examined how many articles: (a) used the necessary methods for each of our three proposed standards, (b) sufficiently reported results for each standard, and (c) produced significant results for each standard. Second, we examined the significant results to compare how many articles met the individual standards and how many met combinations of the standards. The aims of this review are to: 1) compare the effectiveness of applying a single "evidence-based" standard versus applying more specific standards in order to narrow down a pool of CBT options, 2) examine whether youth CBTs may fail to meet standards due to methodology/reporting or due to their ability to produce significant results, and 3) test three specific standards to determine their usefulness as parameters. Given the comprehensiveness of this review of youth CBTs, the findings will demonstrate how different evidence standards yield different sets of treatment protocols. Such findings could provide some insight into how we might more effectively use "evidence-based" standards for treatment evaluation and selection processes.

2. Method

2.1. Article selection and sample

2.1.1. Original article identification

Potential articles were identified from PracticeWise Evidence-Based Services (PWEBS) Literature Database, which we accessed in April 2014 and again in May 2016. PWEBS is a database that summarizes youth mental health service RCT publications (from 1965 through 2015), which are collected from: (1) electronic databases (e.g., PsychINFO); (2) reference lists of relevant literature reviews; and (3) personal communications/nominations from national scholars in treatment outcome research, members of Hawaii's Evidence Based Services Committee, members of the Minnesota Department of Human Services, and other individuals in the professional community.

2.1.2. Criteria for study inclusion

To identify the qualifying CBT treatment groups, 884 articles coded in PWEBS were screened. To be included in this review, study articles must have: (1) tested an active CBT treatment (with or without medication); (2) tested at least one comparison group (e.g., waitlist, medication only, placebo, another CBT group, another active treatment, treatment as usual); (3) used random assignment; (4) explicitly targeted the intervention to treat one of the following: anxiety, depression, disruptive behavior, eating disorder, obsessive-compulsive disorder, substance use, or traumatic stress; and (5) provided results for a target symptom measure. See Fig. 1 for flow chart of article exclusion.

2.1.3. Literature search for separate follow-up articles

Because PWEBS included only the originally published RCT articles, a literature search was conducted in July 2014 and again in May 2016 to identify additional follow-up data published in separate papers. Identifying and integrating data from these additional articles gave our dataset more complete coverage of follow-up data for the RCTs in our sample. Follow-up studies were obtained through internet searches on PsychINFO and Web of Science, refined by: (a) publications citing the original study article; (b) publications that were authored by at least one of the original study authors; and (c) the terms *long-term*, *follow-up*, *longitudinal*, and *outcomes*. Abstracts were examined to confirm if an article was a