Article

Clinical Psychological Science
1–16

DOI: 10.1177/2167702619858424
www.psychologicalscience.org/CPS

An Upper Limit to Youth Psychotherapy Benefit? A Meta-Analytic Copula Approach to Psychotherapy Outcomes

Payton J. Jones[1], Patrick Mair[1], Sofie Kuppens[2,3], and John R. Weisz[1]
[1]Department of Psychology, Harvard University; [2]Department of Psychology and Educational Sciences, KU Leuven; and [3]Working Group in Research and Innovation, KdG University College

Abstract
Across 50 years of research, extensive efforts have been made to improve the effectiveness of psychotherapies for children and adolescents. Yet recent evidence shows no significant improvement in youth psychotherapy outcomes. In other words, efforts to improve the general quality of therapy models do not appear to have translated directly into improved outcomes. We used multilevel meta-analytic data from 502 randomized controlled trials to generate a bivariate copula model predicting effect size as therapy quality approaches infinity. Our results suggest that even with a therapy of perfect quality, achieved effect sizes may be modest. If therapy quality and therapy outcome share a correlation of .20 (a somewhat optimistic assumption given the evidence we review), a therapy of perfect quality would produce an effect size of Hedges's $g = 0.83$. We suggest that youth psychotherapy researchers complement their efforts to improve psychotherapy quality by investigating additional strategies for improving outcomes.

Keywords
meta-analysis, psychotherapy, youth psychotherapy, psychotherapy research, copula, open data, open materials

Received 10/10/18; Revision accepted 3/15/19

Does well-designed, well-documented, psychologically principled, and carefully implemented psychotherapy lead to better outcomes than therapy of lower quality? Empirical evidence on the association between therapy quality and therapy outcome is more mixed than one might expect.

The literature reveals varying opinions on what constitutes a therapy of high quality. One view suggests that an important dimension of therapy quality is the presence of advantageous "specific factors" or "theory-specified factors" in psychotherapies (Castonguay & Grosse, 2005; Webb, DeRubeis, & Barber, 2010). These researchers emphasize the idea that the content of therapy is an important dimension of therapy quality. They stress the frequency with which certain theoretically driven approaches involving specific content, such as exposure and response prevention for OCD, outperform other psychotherapies (DeRubeis, Brotman, & Gibbons, 2005). The theory-specified factors perspective implies that randomized controlled trials comparing different types of therapy are of great importance to the scientific literature because this approach is likely to reveal the types of therapeutic content that are most effective in reducing psychopathology.

Whereas differing specific factors and treatment types have often dominated the discussion of therapy, some influential theorists and researchers have discounted the importance of these factors. Since 1936, some influential figures in the field have argued that the specific steps followed in therapy may have little impact relative to the influence of certain common factors (Messer & Wampold, 2002; Rosenzweig, 1936). One prominent version of this perspective has been labeled

Corresponding Author:
Payton J. Jones, Harvard University, 1240 William James Hall, 33 Kirkland St., Cambridge, MA 02138
E-mail: payton_jones@g.harvard.edu

the "Dodo Bird" conjecture, in reference to the character in Lewis Carol's book, *Alice in Wonderland*, who proclaims: "Everybody has won, and all must have prizes." The Dodo Bird conjecture proposes that diverse types of therapies are equally effective provided that they possess certain common factors. One aspect of this perspective is the notion that across a broad range of bona fide therapies, the specific factors associated with therapy quality bear little relation to therapy outcome (Wampold et al., 1997). Importantly, the common factors approach does not discount the notion that therapy quality matters—it simply emphasizes different *types* of therapy quality that are independent of therapeutic content, such as adequate therapist training in facilitative interpersonal skills (Anderson, McClintock, Himawan, Song, & Patterson, 2016).

The Dodo Bird hypothesis has generated both controversy and data synthesis, with various meta-analyses and reviews cited in support of each position in the debate. Some meta-analyses failed to identify therapy type as a significant moderator of outcome (e.g., Baardseth et al., 2013; Miller, Wampold, & Verhely, 2008; Wampold et al., 1997) and have been cited as support for the Dodo Bird conjecture. Findings of other meta-analyses and some reviews indicated that the specific procedures performed in therapy matter. For instance, some of these findings identified significant between-therapy type differences in magnitude of effect size for various treated problems (e.g., Chambless & Ollendick, 2001; Hunsley & Di Guilio, 2002; Weiss & Weisz, 1995; Weisz, Weiss, Han, Granger, & Morton, 1995), and others have identified therapies for which evidence shows adverse effects (Lilienfeld, 2007). The existence of harmful therapies suggests that the dimension of therapy quality is related to therapy outcome, at least on the extreme low end of therapy quality (e.g., in which therapy is designed in opposition to psychological principles). In other relevant work, researchers have shown that type of therapy can have a marked impact when symptoms are especially severe (e.g., Lorenzo-Luaces, DeRubeis, van Straten, & Tiemens, 2017). As these examples indicate, the various syntheses of evidence have suggested that therapy content may matter in some cases but not in all. Both theoretical camps promote the idea that psychotherapy varies in its quality and its quality can be improved, but the camps differ as to how this should be done. Researchers in the specific factors camp have focused on improving therapy content, whereas those in the common factors camp have focused on maximizing factors that exist independent of therapy content, such as therapist skills in interacting with clients.

Both perspectives are relevant to the present article, in which we focus on the relation between psychotherapy quality and psychotherapy outcome. To define psychotherapy quality in a way that encompasses both perspectives, we have tried to synthesize points from both sides of the debate. The specific factors view suggests that high quality in therapy will include the use of procedures that have a basis in sound psychological principles and the accumulation of evidence from empirical studies together with training of therapists in the specific procedures involved and ensuring adherence to the specified protocols (e.g., Chambless & Ollendick, 2001). The common factors view suggests that high quality in therapy will include an array of therapist characteristics, such as skill in the interpersonal aspects of working with clients. For purposes of the present article, we include both perspectives, operationally defining *quality of therapy* to include both (a) the specific contents of therapy protocols and the procedures (e.g., therapist training) used to ensure faithful delivery of those contents and (b) common factors (e.g., therapist interpersonal skills) that may influence the conduct of therapy independently of specific treatment content, provided that the procedures or elements of (a) and (b) are intended to improve client outcome and do not depend on client factors.

To better understand how quality is defined and operationalized throughout this article, one can use the metaphor of a psychological scale that measures therapy quality. Our metaphorical scale for quality would include a list of items that derive from both the specific factors and common factors approach. When we refer to quality as a general principle, we refer to the metaphorical sum score of all items on the scale (or perhaps more precisely as an extracted principal component from all items). That is, we aim to represent quality as an abstract dimension comprising all relevant aspects of high-quality therapies.

How Much Does Quality Matter?

As noted, researchers have argued over what constitutes a high-quality therapy. But to what degree do these different aspects of therapy quality predict outcome? We reviewed the literature on psychotherapy quality in an exploratory search. Our aim was to identify empirical articles that estimated the relationship between therapy outcome and some aspect of psychotherapy quality. Because research in this area is scarce, we broadened our review of this area to include both adult- and youth-focused therapies.

To make sure we had adequately represented diverse views on this issue, we contacted prominent psychotherapy researchers and asked them to recommend studies. To identify researchers, we searched PsycINFO for the period from January 1990 to January 2018 using the terms "common factors in psychotherapy," "empirically supported psychotherapy," and "psychotherapy quality." We identified authors of the identified publications who were frequently cited for work related to these topics. In addition, we identified current and past editors of journals in which psychotherapy research is frequently published. The resulting list of authors included 19 prominent researchers with diverse theoretical perspectives: David Barlow, Larry Beutler, Ronald Brown, Dianne Chambless, David Clark, Michelle Craske, Joanne Davila, Robert DeRubeis, Judy Garber, Mark Hilsenroth, Steven Hollon, Alan Kazdin, Philip Kendall, Michael Lambert, John Norcross, Francheska Perepletchikova, Dan Strunk, and Bruce Wampold. We sent an e-mail[1] to each author requesting that they identify the most scientifically sound study in which the relationship between quality and outcome was assessed.

Twelve of the 19 authors responded to our request. Some of the authors declined to provide a study, noting theoretical concerns with the idea of therapy quality or concerns related to unfamiliarity with more recent literature on therapy outcomes. Other authors provided more than one study. All provided studies, including results from our own initial literature review, were initially considered as part of the exploratory analysis.

After reviewing the full pool of nominated studies, we excluded several studies from further analyses because of (a) failure to report effect sizes, (b) failure to include a discernable measure of therapy quality, or (c) use of therapy quality measures that depended, in full or in part, on client factors (e.g., therapeutic alliance between therapist and client). A list of excluded studies and reasons for exclusion can be found at Open Science Framework (https://osf.io/dhu7y/).

The results of our exploratory search are presented in Table 1. We measured a variety of different types of therapy quality, ranging from treatment type to therapist competence to therapist facilitative interpersonal skills. Our review included both single studies of therapy quality as well as meta-analyses of evidence across many studies. Aside from comparing different types of psychotherapy, none of the studies in our search included experimental manipulations of therapy quality, indicating an important area of research that may be neglected.

A histogram of the pooled effect sizes is presented in Figure 1. A high bar in this histogram indicates that many effect sizes in the literature fell within the range of effect indicated on the *x*-axis. For instance, the first bar in the graph indicates that 17 effect sizes included in our review fell in the range between .00 and .02. Effect sizes are given as r^2 type, which reflects the proportion of variance accounted for by therapy quality on therapy outcome (see Fritz, Morris, & Richler, 2012). In summary, most effect sizes were close to zero, indicating that higher therapy quality did not relate to better therapy outcome. Meta-analyses and more recent studies in general reported smaller effects compared with individual studies and older studies. All meta-analytic effects fell between 0 and .005. In general, these exploratory analyses showed an association between therapy quality and therapy outcome that was quite modest, at best.

How Good Can Therapy Be? A Focus on Youth Psychotherapy

After synthesizing findings of the studies listed in Table 1, we were interested in applying what could be learned from that synthesis to estimate the extent to which improving therapy quality might improve psychotherapy outcome. For that purpose, we needed a large pool of psychotherapy outcome studies. Because therapy procedures and protocols as well as required therapist skills are quite different for treatment of children and adolescents (herein youths) than for treatment of adults, we thought it best to focus on one or the other age group. Although both age groups are important, our past research on youth psychopathology and psychotherapy and the fact that we had access to data from 502 randomized controlled trials of youth psychotherapy led to our focus on therapy with young people. This work illustrates a procedure that could, of course, be applied in the future to any group, defined by age or any other factor. Using this large youth psychotherapy data set, we sought to determine the efficacy of an *optimal quality* therapy—that is, a therapy in which all beneficial clinician factors were maximized. In other words, we were not interested in answering the question of "How good is therapy?" but rather in answering the question of "How good could therapy be?" Answering this question is akin to an optimization problem in mathematics. First, a function must be specified that describes relevant inputs (therapy quality) and outputs (therapy outcome). The function is then analyzed to identify the point at which a maximal output is achieved on the basis of the inputs. We sought to answer this question on the basis of relevant knowledge regarding youth psychotherapy quality and outcome.

To address these questions using empirical data, we first generated a bivariate distribution function, known as a copula, between therapy quality and treatment outcome drawing on an extensive meta-analysis of randomized controlled trials (RCTs) of youth psychotherapy. We then utilized our simulated distribution to predict the upper limit of effect size as therapy quality approaches infinity. In other words, we posed the question: "If we could design a youth psychotherapy of perfect quality, how effective would it be?"

Table 1. Past Studies: Relationship of Indicators of Therapy Quality to Therapy Outcome

Authors	Year	Journal	Operationalization of quality	r^2
		Meta-analyses		
Weisz et al.	2017	*American Psychologist*	Treatment type (posttreatment outcome)	.00118
			Treatment type (follow-up outcome)	.00092
Webb, DeRubeis, & Barber	2010	*Journal of Consulting and Clinical Psychology*	Therapist adherence	.00400
			Therapist competence	.00490
		Single studies		
Anderson, McClintock, Himawan, Song, & Patterson	2016	*Journal of Consulting and Clinical Psychology*	Therapist facilitative interpersonal skills (last session)	.04000
			Therapist social skills	.00250
Goldberg et al.	2016	*Journal of Counseling Psychology*	Therapist experience (time)	.00000
			Therapist experience (cases)[a]	.00280
Branson, Shafran, & Myles	2015	*Behavior Research and Therapy*	Therapist competence (PHQ-9)	.02132
			Therapist competence (GAD-7)	.06605
			Therapist competence (reliable improvement)	.08009
			Therapist competence (PHQ-9, T2)	.04752
			Therapist competence (GAD-7, T2)	.05063
			Therapist competence (reliable improvement, T2)[a]	.00490
			Therapist experience (PHQ-9)	.06864
			Therapist experience (GAD-7)	.00137
			Therapist experience (reliable improvement)[a]	.04537
			Therapist experience (PHQ-9, T2)	.00624
			Therapist experience (GAD-7, T2)	.03960
			Therapist experience (reliable improvement, T2)[a]	.00023
Boswell et al.	2013	*Journal of Consulting and Clinical Psychology*	Therapist adherence[a]	.00640
Strunk, Brotman, DeRubeis, & Hollon	2010	*Journal of Consulting and Clinical Psychology*	Therapist competence (by session)	.07840
			Therapist competence (posttreatment, Outcome 1)	.10890
			Therapist competence (posttreatment, Outcome 2)	.05760
Boswell, Castonguay, & Wasserman	2010	*Journal of Consulting and Clinical Psychology*	Therapist training	n.s.
Anderson, Ogles, Patterson, Lambert, & Vermeersch	2009	*Journal of Clinical Psychology*	Therapist facilitative interpersonal skills (correlation)	.22090
			Therapist facilitative interpersonal skills (model)	.08570
			Social skills	.00000
			Percentage of work time conducting treatment	.00000
			Theoretical orientation	.00010
Huppert, Bufka, Barlow, Gorman, Shear, & Woods	2001	*Journal of Consulting and Clinical Psychology*	Therapist experience with psychotherapy	.40967
			Therapist experience with CBT	.07311
Elkin et al.	1989	*Archives of General Psychiatry*	Treatment type (CBT vs. IPT)	n.s.

Note: n.s. = not significant; PHQ-9 = Patient Health Questionnaire-9 (Kroenke, Spitzer, & Williams, 2001); GAD-7 = Generalized Anxiety Disorder–7 (Spitzer, Kroenke, Williams, & Löwe, 2006); T2 = Time 2; CBT = cognitive behavioral therapy; IPT = interpersonal psychotherapy.
[a]Direction of the effect was opposite that of the theoretical expectation.

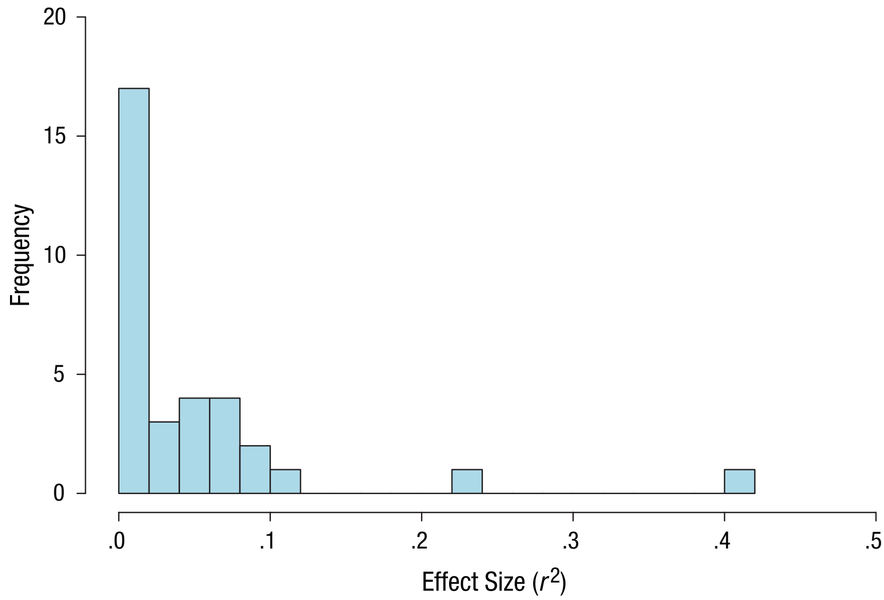

Fig. 1. Relationship between quality and outcome: effect sizes in the literature. The histogram shows pooled effect sizes for the strength of the relationship between therapy quality and therapy outcome from the reviewed literature. The height of each bar represents the frequency with which the effect sizes in the reviewed literature fell within the range indicated on the *x*-axis. Effect sizes are reported as r^2 values.

Method

Database

We used meta-analytic data on the effectiveness of youth psychotherapy compared with control conditions. The database included peer-reviewed RCTs found through a search of the PsycINFO and PubMed databases from January 1960 through May 2017; this updated a search that had previously ended with December 2013, as reported in a meta-analysis by Weisz et al. (2017). We searched for RCTs that had tested psychological therapies for youth depression, anxiety (including problems relating to obsessive compulsive disorder and posttraumatic stress disorder), conduct problems, and attention-deficit/hyperactivity disorder (ADHD, including problems related to inattention and overactivity). These problems account for the majority of youth mental health referrals and treatment. The PsycINFO search used 21 key terms related to psychological therapy (e.g., *psychother-*, *counseling*) that had been used in previous youth therapy meta-analyses, crossed with outcome-assessment topic and age-group constraints. PubMed's indexing system (MeSH) searches publishers who may use different keywords for the same concepts; we used *mental disorders*, with the following search limits: clinical trial, child, published in English, and human subjects. In addition, we searched relevant reviews and meta-analyses, followed reference trails in the reports we identified, and obtained additional studies identified through correspondence with youth-therapy researchers.

We used the following criteria for study inclusion: (a) participants selected and treated for psychopathology; (b) youths randomly assigned to treatment and control conditions, in which at least one of the treatment conditions was psychological therapy (we excluded treatment conditions involving pharmacotherapy or pharmacotherapy combined with psychotherapy); (c) mean participant age 4 to 18 years; (d) outcome measures administered to both treatment and control participants after treatment; and (e) the study was written in English. Psychopathology was defined as either a disorder in a formal diagnostic system or elevated symptoms (e.g., clinical range on standardized measures of psychopathology, treatment referral by parents) because (a) both definitions of psychopathology are common in the treatment-research literature (Weisz, 2004; Weisz & Kazdin, 2017), (b) youths with elevated symptoms experience significant impairment (Costello, Angold, & Keeler, 1999; Silverman & Hinshaw, 2008), (c) such youths frequently receive psychotherapy (Jensen & Weisz, 2002; Weisz, Ugueto, Cheron, & Herren, 2013), and (d) diagnostic categories and their definitions within formal systems have varied markedly across the decades, ruling out sole reliance on diagnosis.

Of the 4,592 studies retrieved and screened, 502 met inclusion criteria (see flowchart in Fig. S1 in the Supplemental Material available online). Studies spanned 1963

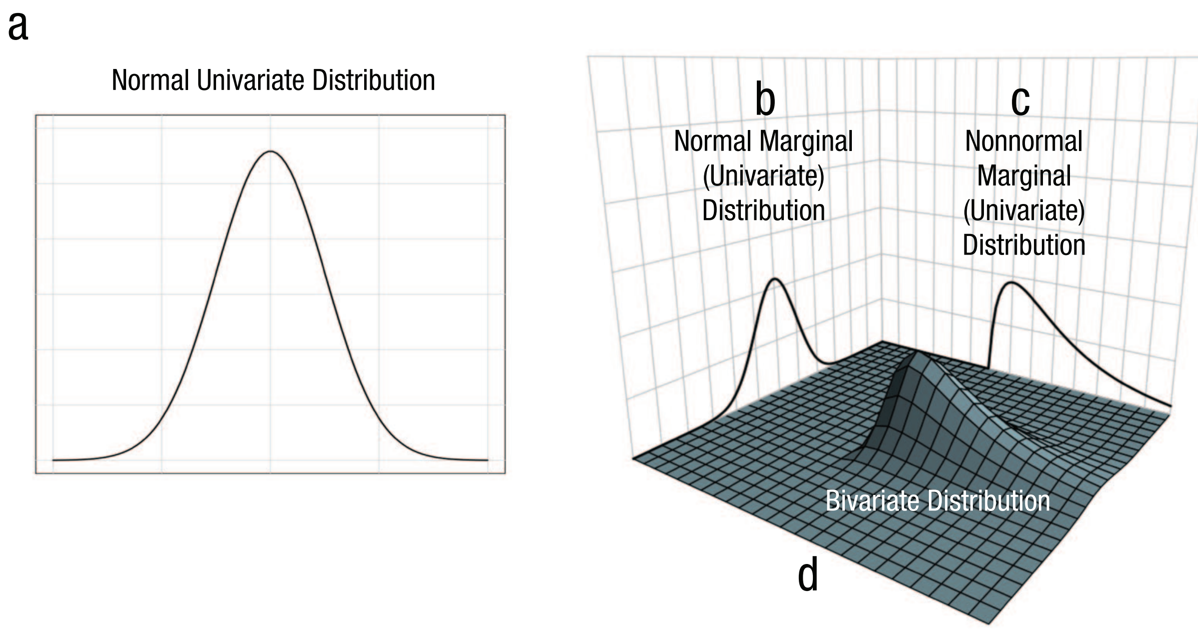

Fig. 2. Explanatory figure for understanding bivariate distributions. (a) A normally distributed univariate distribution. The height of the line indicates the likelihood that a value randomly drawn from the distribution equals the given value on the *x*-axis (i.e., the density). (b) The same normally distributed univariate distribution but represented as a marginal distribution of the bivariate distribution shown in (d). (c) A univariate distribution that follows a nonnormal shape. This distribution is also a marginal distribution of the bivariate distribution. (d) The bivariate distribution created by a combination of the distributions shown in (b) and (c). The height of the curve at any given point indicates the density value for the joint distribution.

to 2017. These studies included a total of 38,055 participants, for a total of 6,241 dependent effect sizes of treatment versus control conditions. Studies included participant samples with a mean age of 10.47 years (*SD* = 3.77) and a majority of males (61.63%; *SD* = 24.80). Most samples (64.10%) were majority White.

Effect sizes

Effect sizes were initially calculated as Cohen's *d* (Cohen, 1988), representing the mean difference between treatment and control conditions divided by the pooled standard deviation. Subsequently, all effect size values were adjusted using Hedges's *g* small-sample correction (Hedges & Olkin, 1985). In subsequent mentions of effect size, we refer to this unbiased effect-size estimate of the population standardized mean difference (*g*). Studies reporting only *p* values or significant effects (1.7% of cases) were assigned the minimum *g* that would produce the significance level given the sample size. Studies reporting only a nonsignificant effect (11.8% of cases) were assigned $g = 0$. We present a sensitivity analysis at a later point to account for the effect of these imputed values on our results.

Coupling two distributions: an introduction to bivariate models

A single variable can be described by a univariate distribution (see Fig. 2a). A univariate distribution is typically represented as a line on a two-dimensional plot in which the height of the line indicates the likelihood that a value randomly drawn from the distribution equals the given value on the *x*-axis (more formally called a *density*). Probability values for a given range can be calculated by measuring the area underneath the density curve. Perhaps the most accessible distribution function to psychologists is the normal distribution, which takes a bell-curve shape, and values in the middle are of highest density. However, different variables follow a variety of different distributions.

Sometimes, we are interested in specifying a joint distribution involving two variables (a bivariate distribution; see Fig. 2d). A bivariate distribution cannot be easily represented as a two-dimensional line because it involves combining two different univariate distributions. Instead, it can be represented as a three-dimensional "hill." The jointly formed bivariate distribution is particularly useful because at any given point on the hill we can assess the density of both of the univariate distributions simultaneously. If we view the hill from the perspective of one of the axes, we can see the outline of one of the univariate distributions in the background—this is called a *marginal distribution* (see Figs. 2b and 2c). A marginal distribution is nothing more than a univariate distribution that is also part of a higher dimensional distribution function.

The bivariate distribution shown in Figure 2d is actually only one of many possible combinations of the two marginal distributions shown in Figures 2b and 2c because bivariate distributions depend not only on the shape of the two marginal distributions but also on the dependence structure, or the association, between them. In our case, a strong dependence structure means a large correlation between therapy quality and therapy outcome. If a strong dependence structure existed in the univariate distributions shown in Figure 2, we might be able to make out a "ridge" along the hill that slants diagonally. The change of a bivariate distribution on the basis of changes in the dependence structures can be viewed as animations (see the Supplemental Material).

Classical bivariate distributions require both univariate distributions to be of the same type. For instance, if two variables are both normally distributed, their combined distribution is known as the *bivariate normal distribution*, which is a special case of a wide range of possible bivariate distributions. The dependence structure in a bivariate normal distribution is described by linear dependence, expressed through a variance–covariance matrix.

Classical distributions unfortunately are limited because they allow us to examine only variables that follow the same distribution. In some circumstances, the two distributions we wish to examine are not of the same type, as shown by the bivariate distribution in Figure 2. In such cases we can use copulas (e.g., Joe, 1997; Nelsen, 2007; Sklar, 1959) to create a more flexible model. Mathematically, this bivariate concept also can be easily generalized to multivariate settings, although any distribution beyond a bivariate distribution becomes impossible to visualize. Copulas are a well-established and validated statistical approach to modeling distributions (Joe, 1997; Nelsen, 2007) and are frequently used in the fields of economics (Patton, 2012), civil engineering (Dupuis, 2007), finance (Cherubini, Luciano, & Vecchiato, 2004), and climate research (Schoelzel & Friederichs, 2008). In psychology, copulas mostly have been used for methodological developments in psychometrics (Braeken, Kuppens, De Boeck, & Tuerlinckx, 2013; Braeken, Tuerlinckx, & De Boeck, 2007; Mair, Satorra, & Bentler, 2012; Nikoloulopoulos & Joe, 2015).

To put it simply, copulas can *couple* distributions of different types in potentially more complex dependency structures. Copulas are useful for modeling associations between variables when the univariate distributions do not both follow the same distribution. For example, in the field of resource management, Shiau (2006) used copulas to examine the relationship between the duration of droughts (which follows an exponential distribution) and the severity of said droughts (which follows a gamma distribution). Copulas are also a useful approach to modeling associations when a specific domain cannot be easily measured but the probability density function is known or can be reasonably assumed. Even if a specific domain can be measured, utilizing common distribution functions can be useful for generalization. There are many different types of copulas; in this analysis, we used normal copulas, a popular type that specifies dependency structure by means of a correlation. Normal copulas should not be confused with normal distributions because normal copulas can use a variety of possible distribution functions for each marginal distribution.

Copulas were an essential approach to answering the central question of the present study for three reasons. First, the effect sizes in our meta-analytic database were not normally distributed, which precluded the use of a bivariate normal distribution. Second, a copula approach allowed us to produce simulations using an imputed distribution function of therapy quality according to our "best-case" model. We compared across many possible distributions of therapy quality to ensure the robustness of our results. Finally, copulas were essential because the distributions of therapy outcome and therapy quality were not of the same type.

For the effect-size marginal distribution, we used the empirical effect-size values (Hedges's g) from our meta-analytic data to estimate an appropriate density function. We conducted an exploratory search strategy to find a distribution that best fit our data using the fitdistr function from the *fitdistrplus* package (Version 1.0-4; Delignette-Muller & Dutang, 2015) for the R software environment (Version 3.5.1; R Core Team, 2018). The goodness of fit was evaluated using Akaike information criterion (AIC) and Bayesian information criterion (BIC). We evaluated the normal distribution, gamma distribution, power-normal distribution, Cauchy distribution, log-normal distribution, and the double-exponential distribution, with the double-exponential distribution (also known as the Skew-Laplace distribution) demonstrating the highest goodness of fit (Aryal & Nadarajah, 2004; see Fig. S2 in the Supplemental Material). The double-exponential is a highly flexible distribution that essentially places two different exponential functions back to back, allowing for a high middle peak in the distribution and flexibility in terms of skewness.

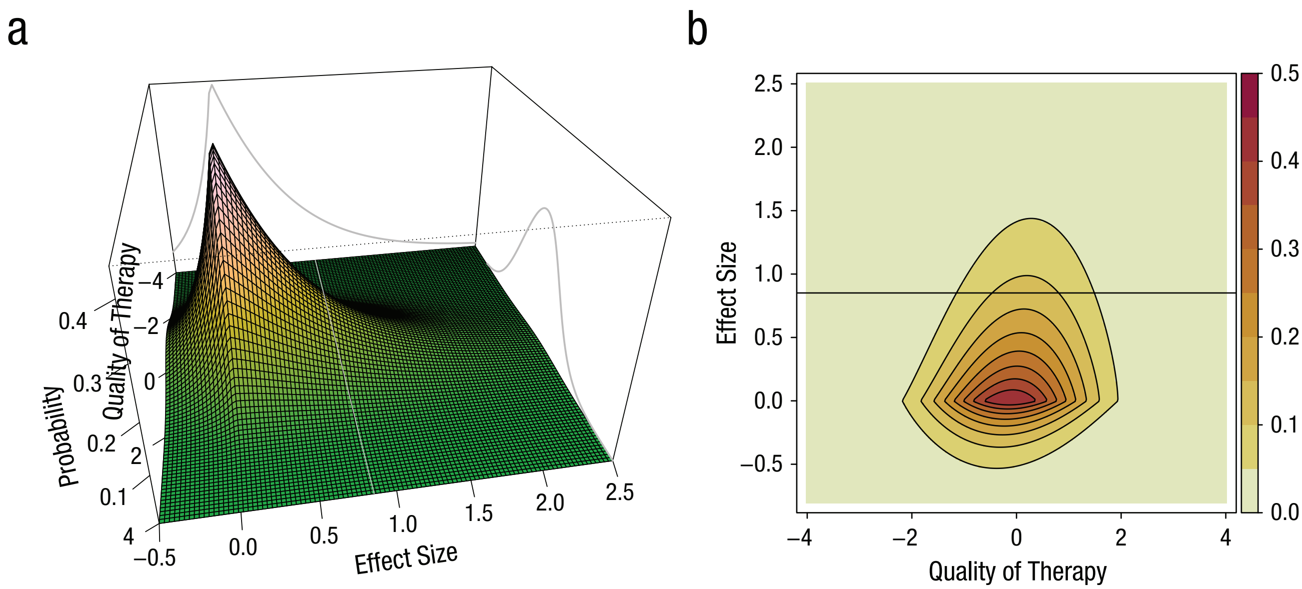

Fig. 3. Representative copula model based on meta-analytic data. (a) A three-dimensional representation of the estimated copula model. The *y*-axis represents the density of a given combination of quality of therapy (*z*-axis) and effect size of therapy outcome (*x*-axis). Most therapies are associated with modest effect sizes even when the quality of therapy is high. (b) The same copula model rendered in two dimensions. This plot is identical to a top-down view of the figure on the left; higher densities are indicated by warmer colors. For an introduction to this figure, see the section Coupling Two Distributions: An Introduction to Bivariate Models.

For the therapy quality marginal distribution, we did not directly model on the basis of empirical findings but instead simulated a hypothetical dimension. Although it may likely seem strange to those unfamiliar with copula modeling, the *range* of therapy quality is irrelevant to obtaining results for the other marginal distributions. That is, it does not matter whether quality ranges from 0 to 5, from −0.2 to 0.3, or from 1 to 10,000—the results would be equivalent in each case because we always model a distribution function that allows the range to extend from negative infinity to positive infinity. The only factor related to therapy quality that is relevant to creating accurate predictions in terms of effect size is the shape of the therapy quality dimension—that is, the distribution function that it follows. Because our empirical set of data rarely included assessments of therapy quality and therefore did not provide useable evidence regarding how quality is shaped across various studies, we started by using a normal distribution (see Fig. 2a), which is the most common type of distribution in psychology and was deemed appropriate by those of us experienced with the psychological treatment literature. Again, we emphasize that the range of the therapy quality dimension does not affect results but only the shape of the distribution.

Because we cannot be completely confident that quality falls under a normal distribution, we also used five different types of distributions (i.e., different shapes of distributions) and checked across the results of each to examine the robustness of our results. For the purposes of this article, we present the results below using the normal distribution for therapy quality (i.e., a bivariate normal copula with a double-exponential and a normal marginal distribution) and include the results from other distributions elsewhere (see Open Science Framework; https://osf.io/dhu7y/). The results were consistent across each distribution, with the normal distribution representing a liberal estimate of the upper limit (i.e., a relatively high upper limit compared with the other distributions tested). Figure 3 displays an example copula with a double-exponential and a normal marginal distribution.

In a first set of simulations, we systematically varied the correlation from 0 to 1 in steps of .01 to demonstrate how the upper limit changes according to the relationship between therapy quality and therapy outcome. One can think of the analysis in terms of a point cloud with a line extending through it. Although the actual observed points in the cloud are limited in scope (i.e., the observed effect sizes from randomized controlled trials), the line (i.e., the model) extends to infinity in each direction. This allows us to make predictions about unobserved points that are far beyond the point cloud (i.e., the model) itself. As the relationship between quality and outcome increases, so does the expected upper limit of psychotherapy outcome.

Simulations and estimation of upper limits

We constructed copulas using the normalCopula function from the *copula* package (Hofort, Kojadinovic, Mächler, & Yan, 2018; Kojadinovic & Yan, 2010) for R.

Table 2. Estimated Upper Limit by Correlation Between Quality and Effect Size

r	Hedges's *g*	*r*	Hedges's *g*	*r*	Hedges's *g*
.00	0.40	.40	1.27	.80	2.14
.01	0.43	.41	1.27	.81	2.15
.02	0.45	.42	1.30	.82	2.17
.03	0.45	.43	1.34	.83	2.18
.04	0.49	.44	1.35	.84	2.21
.05	0.51	.45	1.38	.85	2.23
.06	0.51	.46	1.41	.86	2.26
.07	0.55	.47	1.41	.87	2.27
.08	0.59	.48	1.44	.88	2.31
.09	0.61	.49	1.46	.89	2.31
.10	0.62	.50	1.50	.90	2.35
.11	0.63	.51	1.51	.91	2.38
.12	0.66	.52	1.52	.92	2.39
.13	0.70	.53	1.56	.93	2.42
.14	0.72	.54	1.56	.94	2.44
.15	0.72	.55	1.59	.95	2.46
.16	0.75	.56	1.60	.96	2.49
.17	0.76	.57	1.62	.97	2.51
.18	0.79	.58	1.66	.98	2.52
.19	0.82	.59	1.67	.99	2.54
.20	0.83	.60	1.70	1.00	2.55
.21	0.84	.61	1.72		
.22	0.87	.62	1.75		
.23	0.89	.63	1.77		
.24	0.92	.64	1.77		
.25	0.94	.65	1.81		
.26	0.97	.66	1.82		
.27	0.99	.67	1.85		
.28	0.99	.68	1.87		
.29	1.02	.69	1.89		
.30	1.05	.70	1.91		
.31	1.07	.71	1.94		
.32	1.08	.72	1.95		
.33	1.11	.73	1.99		
.34	1.13	.74	2.00		
.35	1.17	.75	2.01		
.36	1.19	.76	2.03		
.37	1.21	.77	2.06		
.38	1.22	.78	2.08		
.39	1.24	.79	2.13		

Note: *r* = simulated correlation between quality of therapy and effect size; Hedges's *g* = upper limit of Hedges's *g* based on 100 samples of 1,000 observations each.

We systematically varied the correlation from 0 to 1 in steps of .01 to examine the upper limit at each possible dependency value. Each case represents an estimated bivariate distribution function given a predetermined degree of dependence between therapy quality and therapy effect size as an indicator of treatment outcome.

To estimate the upper limit of treatment outcome given a "perfect" therapy quality, we simulated 100 data sets of 1,000 observations each from each of the dependency specifications using the mvdc function in the *copula* package. For each simulated data set, we then fitted a linear regression depicting the association between the two marginal distributions. Afterward, we estimated the value of effect size at the 99.9th percentile of therapeutic quality on the basis of the fitted linear regression. For each of the dependency specifications, we created a point estimate by using the median value of the predicted upper limits from the 100 data sets. This process can be conceptualized as bootstrapping the upper limit of effect size across each of the dependency values between 0 and 1 as therapeutic quality approaches infinity. This initial analysis yielded feasible estimates for the upper limit across all possible dependencies.

Results

We modeled the upper limit of treatment outcome in numerous situations, systematically varying the correlation between treatment quality and treatment outcome from 0 to 1 in steps of .01. We estimated a range for the upper limit of therapy effect size because therapy quality approached infinity by fitting linear distributions to simulations from each copula generated and approximating effect size at the 99.9th percentile of therapy quality. This analysis was done with the intent to help the reader understand how the upper limit varies as a function of the dependence between quality and therapy. As presented in Table 2, each upper limit estimate was constructed from the median of 100 samples drawn from the copula matching the given dependence level.

An example of a representative bivariate distribution can be seen in Figure 2. Additional examples of bivariate distributions and animated versions showing the continuous change in the distributions on the basis of different dependencies can be seen in the Supplemental Animations in the Supplemental Material. Animations are useful to show how the bivariate distribution changes dynamically as a function of the dependence and to rotate figures to see the full three-dimensional perspective.

Table 2 provides insight into the upper limit of therapy effect size at various levels of dependence between therapy quality and treatment outcome. With a perfect correlation between therapy quality and outcome (r = 1.0) and maximized quality (99.9th percentile), the expected effect size (g) is 2.55—representing the highest possible value for the upper limit of therapy efficacy. If therapy quality and therapy outcome share a small to medium correlation of .2 (a somewhat optimistic assumption given the evidence we have reviewed) and therapy quality is maximized (99.9th percentile), the expected effect size is 0.83.

Discussion

How good can youth psychotherapy be? We explored this question via a series of steps, starting with an exploratory literature search and ending with mathematical simulations based on meta-analytic data from youth treatment research. Psychotherapy researchers often have made the optimistic assumption that improving the quality of therapy will result in improved treatment outcome as reflected by higher effect sizes. Although this is sometimes the case, improved therapy quality does not always correspond to an increase in the efficacy of treatment. We conducted an exploratory search of articles that assessed a wide variety of treatment quality indicators with treatment outcome. We contacted prominent researchers in this field and received suggested articles to ensure that our search encompassed adequate breadth. This exploratory search yielded small effects across many different types of treatment quality with few exceptions (e.g., Anderson et al., 2009; Huppert et al., 2001).

We next turned to generating a mathematical model to describe the relationship between therapy quality, defined broadly, and therapy outcomes as measured in 502 RCTs of youth psychotherapy. In simulations, we modeled the outcome of therapy across several parameterizations, indicating that in a best-case scenario, youth psychotherapy has a large but not unprecedented effect size.

If we assume an optimistic small to medium correlation ($r = .2$) between quality of therapy and effect size, our model predicted that as quality of therapy approaches perfection, the effect size of youth psychotherapy is estimated to reach $g = 0.83$. Although 0.83 represents a large effect size, this may seem like a low estimate to psychotherapy researchers who have hoped to make large improvements to psychotherapy in the future. Researchers may have assumed that the effects of youth psychotherapies—and thus the mean effects identified in meta-analyses—will increase over the years with advances in treatment development and research. However, recent data (Weisz et al., 2017, 2019) have not shown a significant increase over the years. Our present analyses suggest the possibility that this pattern may reflect, to some degree, an upper limit to the growth of youth-psychotherapy effect size.

Astute readers may note that many RCTs in the past (including many RCTs in our own data set) have produced effect sizes larger than 0.83 and may wonder how such data can fit with our conclusions. There are several possibilities that help explain this observation. First, 0.83 is a point estimate, but observed effects would be expected to be distributed across a range, with some effect sizes markedly higher than 0.83. The model we produced generalizes across all client populations—it is highly likely that effect sizes greater than 0.83 exist for certain subsets of client populations (e.g., youths with anxiety problems), with lower effect sizes for other subsets (e.g., youths with depression). In addition, our model generalizes across differing levels of severity in the client population, which also may have an impact on effect size. Future research is needed to examine highly effective therapies and clarify the causal factors that lead to larger effects. Novel developments in research and statistics such as machine learning may hold promise for better understanding these effects. Finally, when many RCTs are conducted, especially when conducted with small sample sizes, it is inevitable that some effect sizes will be artificially inflated because of random error, which may be the case for some RCTs with very large effect sizes.

It is certainly possible that psychotherapy quality indicators we failed to identify are more strongly correlated with outcome than the variables our search did identify and thus that the upper limit extends upward beyond our estimates. In that event, the picture may be more optimistic, but that would not necessarily contradict the principle that a youth psychotherapy benefit ceiling exists, whatever that ceiling is ultimately found to be. If there is, in fact, such a ceiling, it may be useful to consider why this might be the case.

One explanation may lie in the basic truth that the outcome of youth psychotherapy is highly overdetermined, with substantial variance accounted for by an array of additional factors, outside therapy, that impact the lives of young people (Weisz et al., 2019). In addition to psychotherapy, factors encompassing genetic endowment, biological makeup, family context, and the broader social environment may exert strong influence, in some cases eclipsing the influence exerted by psychotherapy. This may be especially true for young people, whose ability to control life events and living conditions is more constrained than is the case with adults; youth outcomes may be impacted by family financial resources, parents' behavior, sibling relationships, peer influence, neighborhood conditions, events at school, and a variety of other forces the young person may have little or no capacity to alter. Ultimately, an hour of psychotherapy per week is in a kind of competition with all that happens during the other 110+ waking hours, and many of the forces that can contribute to psychological distress and dysfunction during those hours may not be readily altered by therapy. From this perspective, it may make sense to construe youth psychotherapy as but one of many forces that can impact youth mental health and functioning and in many cases not the most powerful of those forces. One logical implication of this view is that there must be a natural upper limit to the influence psychotherapy alone can exert.

Implications for clinical researchers

Although our model used a large sample of youth psychotherapy RCTs, findings only reflect psychotherapy as it has been structured and tested to date. These studies reflect only the models of intervention and assessment that treatment developers and researchers have thought of thus far. It is possible, in principle, that significant changes in the ways therapies are designed and implemented and assessment is done could change the picture substantially, including both the strength of association between quality and outcome and the estimated upper limit of treatment benefit. Small, incremental changes to current approaches may not be sufficient, but more dramatic changes—including altered therapy models and shifts in the ways therapy quality is assessed—may have potential (see Weisz et al., 2019).

A great deal of time, money, and research expertise has gone into creating high-quality therapies. Much was well spent—we now have empirically supported treatments for a variety of mental illnesses. However, it now may be time to change our priorities. Our analyses indicate that expending additional effort to improve the quality of therapy as currently structured may offer diminishing returns in terms of the efficacy we are able to produce. Expanding effort on other priority areas may provide benefits that have not yet been maximized. Of course, there is no guarantee that alternative areas of exploration will yield greater benefits; they may have equally prohibitive limitations. Still, they likely merit further exploration. Examples might include (a) dissemination of existing empirically supported psychotherapies to expand the number of lives impacted, (b) alternative modes of psychotherapy delivery that may outperform current models, (c) greater focus on understudied indicators of psychotherapy quality that have shown relatively stronger association with outcome than most indicators (e.g., facilitative interpersonal skills training for therapists), and (d) complementing therapy with strategies that address personal and environmental factors that influence outcome (e.g., expert case management to help clients address real-life events and challenges that stand in the way of good adjustment and functioning).

Although many effective psychological treatments have now been developed (APA Presidential Task Force for Evidence-Based Practice, 2006), a substantial gap in care exists, with many who need help never accessing these treatments. The National Survey on Drug Use and health estimated that only 43% of Americans with mental illness receive any treatment at all (Park-Lee, Lipari, Hedden, Copello, & Kroutil, 2016), and among those who do receive care, the clear majority do not receive empirically supported treatments (Shafran et al., 2009). This problem is even more pronounced among ethnic minorities: African Americans were less likely to access services than European Americans (12.5% vs. 25.4%), and Hispanic Americans were less likely to receive adequate care than European Americans (10.7% vs. 22.7%; Wells, Klap, Koike, & Sherbourne, 2001). A case can be made that the emphasis among psychotherapy researchers on incremental improvements in the best therapies may be misplaced given that 60% of those with mental illness receive no care at all and most of those who do receive care do not receive evidence-based treatment.

The traditional office-visit psychotherapy model carries certain limitations. Given the modest results of our estimated limits of psychotherapy outcome, psychotherapy cannot be considered to represent a complete solution to mental illness. Moreover, the psychotherapy model is unsustainable on a large scale; there are simply not enough therapists to do the job or funds to compensate them for all the care that may be needed (Kazdin & Blase, 2011). Scalable mental health promotion and prevention efforts may hold promise to alleviate the burden of mental illness (Kazdin, 2019; Kazdin & Blase, 2011; Schleider & Weisz, 2017). In addition, much of the variance in mental illness can be explained by social and occupational factors, such as socioeconomic status, job stress, academic stress, and lack of education (e.g., Hudson, 2005; Jones, Park, & Lefevor, 2018; Wadsworth & Achenbach, 2005). Rather than working separately, psychotherapists could collaborate within teams of general practitioners, psychiatrists, social workers, sociologists, and others to address factors that are often unaddressed by psychotherapy.

We stress that our findings should not be interpreted as a recommendation against psychotherapy. Our analysis does not suggest that psychotherapy is ineffective, nor does it suggest that alternative intervention strategies (e.g., medication, self-help) are more effective than psychotherapy. In a great number of cases, psychotherapy is regarded as the most effective treatment for a given mental illness (Birmaher, Brent, & Benson, 1998; Butler, Chapman, Forman, & Beck, 2006; Cheung et al., 2007; Cuijpers et al., 2013), produces the lowest rate of relapse (Dobson, Hollon, Schmaling, Kohlenberg, & Gallop, 2008; Hollon, Stewart, & Strunk, 2006), is the most cost-effective option (Antonuccio, Thomas, &

Danton, 1997), and has the fewest side effects (Thase et al., 2007). What our analysis does suggest is that psychotherapy may have limited room for improvement beyond what is currently achieved by the best evidence-based psychotherapies—or at least, less room than we might have assumed.

Limitations

Our analyses tested the upper limit across multiple theoretically indicated distributions to estimate therapy quality. The results showed that the upper limit of treatment outcome was relatively robust across these distributions. The results presented in this article utilize the normal distribution, perhaps the most theoretically appropriate choice. More importantly, it is highly unlikely that assuming a normal distribution in this case would lead us to *underestimate* the upper limit of effect size. For example, publication bias might result in some type of negative skew of the quality distribution; but an underestimation of effect size would only occur if the distribution of quality of therapy was heavily positively skewed. In other words, our model would underestimate the true upper limit only if most RCTs included therapies of low quality and only a few RCTs included therapies of high quality (the reverse of publication bias). We have little reason to suspect this trend. Moreover, the distribution of effect sizes in our sample was nonnormal, being much more "peaked" (e.g., leptokurtic) than a normal distribution. If the quality of therapy were similarly leptokurtic, this also would have led us to overestimate the upper limit of effect size (see Laplace distribution at Open Science Framework; https://osf.io/dhu7y/). In other words, the normal distribution is not only a theoretically guided choice; it is also a choice that has a much greater chance of overestimating the upper limit rather than underestimating it. For the distribution of effect sizes, we chose a distribution that most closely fit the empirical data. We also do not expect an impactful publication bias in terms of effect size (Weisz et al., 2019). We also tested alternative fits to the empirical data, with similar results (see Open Science Framework; https://osf.io/dhu7y/). If researchers have other hypotheses about the distribution of quality of therapy, we encourage them to replicate our analysis using their preferred marginal distributions.

Including all types of quality in a single figure has the downside of overrepresentation of more frequently studied types of quality indicators at the expense of less studied indicators. Moreover, we only included measures of quality that were not dependent on the client. There may be certain types of quality indicators (e.g., cultural adaptation; Smith, Rodriguez, & Bernal, 2011) that may be helpful for a subset of clients. In addition, we limited quality to a single dimension. It is possible that therapy quality is not appropriately represented by a single dimension but instead is better described by multiple dimensions (e.g., a dimension for specific factors and a dimension for common factors or multiple dimensions within each). From a mathematical standpoint, such multivariate extensions to our bivariate model are possible and represent a potential area that could be productively explored in the future.

In the original collection of effect sizes, studies that reported a nonsignificant effect but did not report the exact effect size were inputted as an effect size of 0 (13.4% of cases). To assess the extent to which this affected the outcomes of our study, we conducted a sensitivity analysis by recomputing our core analysis while excluding all effect sizes of 0 (e.g., excluding all studies that reported nonsignificant effects). This change increased the estimated upper limit of psychotherapy (i.e., median = 0.95, 95% confidence interval = [0.69, 1.20]). This increase of 0.12 raises some concern that our conservative method of estimation may have led to a somewhat underestimated upper limit.

Our model does not address client-specific factors in psychotherapy or the effect of individual psychotherapists. Effect sizes of randomized controlled trials show the effect of a certain type of therapy (including training and implementation models) rather than a certain type of therapist (or client). Thus, our data do not allow us to model the potential relationship between the quality of the individual therapist and the outcome of treatment (Wampold & Bolt, 2006). It is possible that effect sizes could be further bolstered through appropriate training of therapist-specific factors (Kim, Wampold, & Bolt, 2006) or personalizing treatments (Lorenzo-Luaces et al., 2017; Ng & Weisz, 2016). Our model also collapses across multiple mental disorders, severities, and other factors; thus, the upper limit of youth psychotherapy for anxiety is likely higher than our estimate, and the upper limit for youth psychotherapy for depression or ADHD is likely lower (Weisz et al., 2017).

Conclusion

We constructed a bivariate model to encapsulate the potential relationship between quality of youth psychotherapy and treatment outcome. Therapy quality refers to the degree to which an intervention is optimally designed and implemented according to what is known or assumed to be known about psychotherapy. Treatment outcome, in contrast, reflects the actual change in symptoms over the course of treatment compared with control. With reference to previous literature, we estimate that most optimistically, therapy quality has a low to moderate relationship with treatment outcome. Our model suggests that the effect size of a therapy with "perfect" quality may be disappointingly low. Specifically, if a modest correlation between quality and outcome is assumed, the model estimates a maximum effect size of 0.83. Although our modeling approach comes with important limitations, the results suggest that expensive efforts to improve psychotherapy quality may have diminishing returns, especially when focusing on aspects of therapy quality that share only a modest relationship with therapy outcome.

Action Editor

Stefan G. Hofmann served as action editor for this article.

Author Contributions

P. J. Jones and J. R. Weisz jointly developed the study concept. P. J. Jones and P. Mair jointly created the methodological design and conducted analyses. S. Kuppens coded and adapted effect sizes in the meta-analytic data. P. J. Jones conducted the initial exploratory literature search, and P. J. Jones and J. R. Weisz jointly contacted researchers for further exploration of the literature. P. J. Jones wrote the initial draft of the manuscript. All of the authors approved the final manuscript for submission.

ORCID iD

Payton J. Jones https://orcid.org/0000-0001-6513-8498

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/2167702619858424

Open Practices

All data and materials have been made publicly available via Open Science Framework and can be accessed at https://osf.io/myfg7. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/2167702619858424. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at https://www.psychologicalscience.org/publications/badges.

Notes

1. Dear Professor ________:

> As part of a new study examining the relation between psychotherapy quality and psychotherapy outcome, we are contacting you and a small number of other prominent psychotherapy researchers. We ask you to please identify what you regard as the most scientifically sound study in the literature in which investigators assessed the relation between a credible measure of psychotherapy quality, on the one hand, and therapy outcome, on the other hand.
>
> If you can provide either a PDF or an exact reference for the study you identify, we will be very grateful. Thanks very much for considering our request.

References

Anderson, T., McClintock, A. S., Himawan, L., Song, X., & Patterson, C. L. (2016). A prospective study of therapist facilitative interpersonal skills as a predictor of treatment outcome. *Journal of Consulting and Clinical Psychology*, *84*, 57–66. doi:10.1037/ccp0000060

Anderson, T., Ogles, B. M., Patterson, C. L., Lambert, M. J., & Vermeersch, D. A. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist success. *Journal of Clinical Psychology*, *65*, 755–768. doi:10.1002/jclp.20583.

Antonuccio, D. O., Thomas, M., & Danton, W. G. (1997). A cost-effectiveness analysis of cognitive behavior therapy and fluoxetine (Prozac) in the treatment of depression. *Behavior Therapy*, *28*, 187–210. doi:10.1016/S0005-7894(97)80043-3

APA Presidential Task Force for Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, *61*, 271–285. doi:10.1037/0003-066X.61.4.271

Aryal, G., & Nadarajah, S. (2004). On the skew Laplace distribution. *Journal of Information and Optimization Sciences*, *26*, 205–217. doi:10.1080/02522667.2005.10699644

Baardseth, T. P., Goldberg, S. B., Pace, B. T., Wislocki, A. P., Frost, N. D., Siddiqui, J. R., . . . Minami, T. (2013). Cognitive-behavioral therapy versus other therapies: Redux. *Clinical Psychology Review*, *33*, 395–405. doi:10.1016/j.cpr.2013.01.004

Birmaher, B., Brent, D. A., & Benson, R. S. (1998). Summary of the practice parameters for the assessment and treatment of children and adolescents with depressive disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *37*, 1234–1238. doi:10.1097/00004583-199811000-00029

Boswell, J. F., Castonguay, L. G., & Wasserman, R. H. (2010). Effects of psychotherapy training and intervention use on session outcome. *Journal of Consulting and Clinical Psychology*, *78*, 717–723. doi:10.1037/a0020088

Boswell, J. F., Gallagher, M. W., Sauer-Zavala, S. E., Bullis, J., Gorman, J. M., Shear, M. K., . . . Barlow, D. H. (2013). Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for panic disorder. *Journal of Consulting and Clinical Psychology*, *81*, 443–454. doi:10.1037/a0031437

Branson, A., Shafran, R., & Myles, P. (2015). Investigating the relationship between competence and patient outcome with CBT. *Behaviour Research and Therapy*, *68*, 19–26. doi:10.1016/j.brat.2015.03.002

Braeken, J., Kuppens, P., De Boeck, P., & Tuerlinckx, F. (2013). Contextualized personality questionnaires: A case for copulas in structural equation models for categorical data. *Multivariate Behavioral Research*, *48*, 845–870. doi:10.1080/00273171.2013.827965

Braeken, J., Tuerlinckx, F., & De Boeck, P. (2007). Copula functions for residual dependency. *Psychometrika*, *72*, 393–411. doi:10.1007/s11336-007-9005-4

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review*, *26*, 17–31. doi:10.1016/j.cpr.2005.07.003

Castonguay, L. G., & Grosse, M. (2005). Change in psychotherapy: A plea for no more "nonspecific" and false dichotomies. *Clinical Psychology: Science and Practice*, *12*, 198–201. doi:10.1093/clipsy.bpi026

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, *52*, 685–716. doi:10.1146/annurev.psych.52.1.685

Cherubini, U., Luciano, E., & Vecchiato, W. (2004). *Copula methods in finance*. New York, NY: Wiley.

Cheung, A. H., Zuckerbrot, R. A., Jensen, P. S., Ghalib, K., Laraque, D., & Stein, R. E. K. (2007). Guidelines for adolescent depression in Primary Care (GLAD-PC): II. Treatment and ongoing management. *Pediatrics*, *120*, e1313–e1326. doi:10.1542/peds.2006-1395

Cohen, J. (1988). *Statistical power analyses for the social sciences*. Hillsdale, NJ: Erlbaum.

Costello, E. J., Angold, A., & Keeler, G. P. (1999). Adolescent outcomes of childhood disorders: The consequences of severity and impairment. *Journal of the American Academy of Child and Adolescent Psychiatry*, *38*, 121–128. doi:10.1097/00004583-199902000-00010

Cuijpers, P., Sijbrandij, M., Koole, S. L., Andersson, G., Beekman, A. T., & Reynolds, C. F. (2013). The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: A meta-analysis of direct comparisons. *World Psychiatry*, *12*, 137–148. doi:10.1002/wps.20038

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, *64*, 1–34. doi:10.18637/jss.v064.i04

DeRubeis, R., Brotman, M., & Gibbons, C. (2005). A conceptual and methodological analysis of the nonspecifics argument. *Clinical Psychology: Science and Practice*, *12*, 174–183. doi:10.1093/clipsy.bpi022

Dobson, K. S., Hollon, S. D., Schmaling, K. B., Kohlenberg, R. J., & Gallop, R. (2008). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the prevention of relapse and recurrence in major depression. *Journal of Consulting and Clinical Psychology*, *76*, 468–477. doi:10.1037/0022-006X.76.3.468.

Dupuis, D. J. (2007). Using copulas in hydrology: Benefits, cautions, and issues. *Journal of Hydrologic Engineering*, *12*, 381–393. doi:10.1061/(ASCE)1084-0699(2007)12:4(381)

Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., Glass, . . . Parloff, M. B. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments. *Archives of General Psychiatry*, *46*, 971–983.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*, 2–18. doi:10.1037/a0024338

Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., & Wampold, B. E. (2016). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology*, *63*, 1–11. doi:10.1037/cou0000131

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hofort, M., Kojadinovic, I., Mächler, M., & Yan, J. (2018). *Elements of copula modeling with R*. New York, NY: Springer.

Hollon, S. D., Stewart, M. O., & Strunk, D. (2006). Enduring effects for cognitive behavior therapy in the treatment of depression and anxiety. *Annual Review of Psychology*, *57*, 285–315. doi:10.1146/annurev.psych.57.102904.190044

Hudson, C. G. (2005). Socioeconomic status and mental illness: Tests of the social causation and selection hypotheses. *American Journal of Orthopsychiatry*, *75*, 3–18. doi:10.1037/0002-9432.75.1.3

Hunsley, J., & Di Guilio, G. (2002). Dodo bird, phoenix, or urban legend? *Scientific Review of Mental Health Practice*, *1*, 11–22. Retrieved from https://www.srmhp.org/0101/psychotherapy-equivalence.html

Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2001). Therapists, therapist variables, and cognitive-behavioral therapy outcome in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology*, *69*, 747–755. doi:10.1037/0022-006X.69.5.747

Jensen, A. L., & Weisz, J. R. (2002). Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnoses for clinic-referred children and adolescents. *Journal of Consulting and Clinical Psychology*, *70*, 158–168. doi:10.1037//0022-006X.70.1.158

Joe, H. (1997). *Dependence modeling with copulas*. Boca Raton, FL: CRC Press.

Jones, P. J., Park, S. Y., & Lefevor, G. T. (2018). Contemporary college student anxiety: The role of academic distress, financial stress, and support. *Journal of College Counseling*, *21*, 252–264.

Kazdin, A. E. (2019). Annual Research Review: Expanding mental health services through novel models of intervention delivery. *Journal of Child Psychology and Psychiatry*, *60*, 455–472. doi:10.1111/jcpp.12937

Kazdin, A. E., & Blase, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on Psychological Science*, *6*, 21–37. doi:10.1177/1745691610393527

Kim, D. M., Wampold, B. E., & Bolt, D. M. (2006). Therapist effects in psychotherapy: A random-effects modeling of the National Institute of Mental Health Treatment of Depression

Collaborative Research Program data. *Psychotherapy Research*, *16*, 161–172. doi:10.1080/10503300500264911

Kojadinovic, I., & Yan, J. (2010). Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software*, *34*, 1–20. doi:10.18637/jss.v016.i09

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*, 606–613

Park-Lee, E., Lipari, R. N., Hedden, S. L., Copello, E. A. P., & Kroutil, L. A. (2016). Receipt of services for substance use and mental health issues among adults: Results from the 2015 National Survey on Drug Use and Health. *Receipt of services for substance use and mental health issues among adults: Results from the 2015 national survey on drug use and health*. Retrieved from https://www.samhsa.gov/data/sites/default/files/NSDUH-ServiceUseAdult-2015/NSDUH-ServiceUseAdult-2015/NSDUH-ServiceUseAdult-2015.htm

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, *2*, 53–70. doi:10.1111/j.1745-6916.2007.00029.x

Lorenzo-Luaces, L., DeRubeis, R. J., van Straten, A., & Tiemens, B. (2017). A prognostic index (PI) as a moderator of outcomes in the treatment of depression: A proof of concept combining multiple variables to inform risk-stratified stepped care models. *Journal of Affective Disorders*, *213*, 78–85. doi:10.1016/j.jad.2017.02.010

Mair, P., Satorra, A., & Bentler, P. M. (2012). Generating nonnormal multivariate data using copulas: Applications to SEM. *Multivariate Behavioral Research*, *47*, 1–19. doi:10.1080/00273171.2012.692629

Messer, S. B., & Wampold, B. E. (2002). Let's face facts: Common factors are more potent than specific therapy ingredients. *Clinical Psychology: Science and Practice*, *9*, 21–25. doi:10.1093/clipsy.9.1.21

Miller, S., Wampold, B., & Verhely, K. (2008). Direct comparisons of treatment modalities for youth disorders: A meta-analysis. *Psychotherapy Research*, *18*, 5–14. doi:10.1080/10503300701472131

Nelsen, R. (2007). *An introduction to copulas*. New York, NY: Springer

Ng, M. Y., & Weisz, J. R. (2016). Building a science of personalized intervention for youth mental health. *Journal of Child Psychology and Psychiatry*, *57*, 216–236. doi:10.1111/jcpp.12470

Nikoloulopoulos, A. K., & Joe, H. (2015). Factor copula models for item response data. *Psychometrika*, *80*, 126–150. doi:10.1007/s11336-013-9387-4

Patton, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, *110*, 4–18. doi:10.1016/j.jmva.2012.02.021

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry*, *6*, 412–415. doi:10.1111/j.1939-0025.1936.tb05248.x

Schleider, J. L., & Weisz, J. R. (2017). Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*, 107–115. doi:10.1016/j.jaac.2016.11.007

Schoelzel, C., & Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research–Introduction to the copula approach. *Nonlinear Processes in Geophysics*, *15*, 761–772.

Shafran, R., Clark, D. M., Fairburn, C. G., Arntz, A., Barlow, D. H., Ehlers, A., . . . Wilson, G. T. (2009). Mind the gap: Improving the dissemination of CBT. *Behaviour Research and Therapy*, *47*, 902–909. doi:10.1016/j.brat.2009.07.003

Shiau, J. T. (2006). Fitting drought duration and severity with two-dimensional copulas. *Water Resources Management*, *20*, 795–815. doi:10.1007/s11269-005-9008-9

Silverman, W. K., & Hinshaw, S. P. (Eds.). (2008). Evidence-based psychosocial treatments for children and adolescents: A 10-year update [Special Issue]. *Journal of Clinical Child & Adolescent Psychology*, *37*(1). Retrieved from https://www.tandfonline.com/toc/hcap20/37/1

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges [Distribution functions of n dimensions and their origins]. *Publications de l'Institut de Statistique de l'Université de Paris*, *8*, 229–231.

Smith, T. B., Rodríguez, M. B., & Bernal, G. (2011). Culture. *Journal of Clinical Psychology*, *67*, 166–175. doi:10.1002/jclp.20757

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*, 1092–1097.

Strunk, D. R., Brotman, M. A., DeRubeis, R. J., & Hollon, S. D. (2010). Therapist competence in cognitive therapy for depression: Predicting subsequent symptom change. *Journal of Consulting and Clinical Psychology*, *78*, 429–437. doi:10.1037/a0019631

Thase, M. E., Friedman, E. S., Biggs, M. M., Wisniewski, S. R., Trivedi, M. H., Luther, J. F., . . . Rush, A. J. (2007). Cognitive therapy versus medication in augmentation and switch strategies as second-step treatments: A STAR*D report. *American Journal of Psychiatry*, *164*, 739–752. doi:10.1176/appi.ajp.164.5.739

Wadsworth, M. E., & Achenbach, T. M. (2005). Explaining the link between low socioeconomic status and psychopathology: Testing two mechanisms of the social causation hypothesis. *Journal of Consulting and Clinical Psychology*, *73*, 1146–1153. doi:10.1037/0022-006X.73.6.1146

Wampold, B. E., & Bolt, D. (2006). Therapist effects: Clever ways to make them (and everything else) disappear. *Psychotherapy Research*, *16*, 184–187. doi:10.1080/10503300500265181

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes." *Psychological Bulletin*, *122*, 203–215. doi:10.1037/0033-2909.122.3.203

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, *78*, 200–211. doi:10.1037/a0018912

Weiss, B., & Weisz, J. R. (1995). Relative effectiveness of behavioral and nonbehavioral child psychotherapy. *Journal of Consulting and Clinical Psychology*, *63*, 317–320. doi:10.1037/0022-006X.63.2.317

Weisz, J. R. (2004). *Psychotherapy for children and adolescents: Evidence-based treatments and case examples*. New York, NY: Cambridge University Press.

Weisz, J. R., & Kazdin, A. E. (Eds.). (2017). *Evidence-based psychotherapies for children and adolescents* (3rd ed.). New York, NY: Guilford.

Weisz, J. R., Kuppens, S., Ng, M. Y., Eckshtain, D., Ugueto, A. M., Vaughn-Coaxum, R., . . . Fordwood, S. R. (2017). What five decades of research tells us about the effects of youth psychological therapy: A multilevel meta-analysis and implications for science and practice. *American Psychologist*, *72*, 79–117. doi:10.1037/a0040360

Weisz, J. R., Kuppens, S., Ng, M. Y., Vaughn-Coaxum, R. A., Ugueto, A. M., Eckshtain, D., & Corteselli, K. A. (2019). Are psychotherapies for young people growing stronger? Tracking trends over time for youth anxiety, depression, ADHD, and conduct problems. *Perspectives on Psychological Science*, *14*, 216–237. doi:10.1177/1745691618805436

Weisz, J. R., Ugueto, A. M., Cheron, D. M., & Herren, J. (2013). Evidence-based youth psychotherapy in the mental health ecosystem. *Journal of Clinical Child and Adolescent Psychology*, *42*, 274-286. doi:10.1080/15374416.2013.764824

Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology*, *55*, 542–549. doi:10.1037/0022-006X.55.4.542

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: a meta-analysis of treatment outcome studies. *Psychological Bulletin*, *117*, 450–468. doi:10.1037/0033-2909.117.3.450

Wells, K., Klap, R., Koike, A., & Sherbourne, C. (2001). Ethnic disparities in unmet need for alcoholism, drug abuse, and mental health care. *The American Journal of Psychiatry*, *158*, 2027–2032. doi:10.1176/appi.ajp.158.12.2027