Clinical Psychology Review 73 (2019) 101774

Contents lists available at ScienceDirect

Clinical Psychology Review

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

Review

Is there a sleeper effect of exposure-based vs. cognitive-only intervention for anxiety disorders? A longitudinal multilevel meta-analysis

Ioana R. Podina[a,*], Andreea Vîslă[b], Liviu A. Fodor[c], Christoph Flückiger[b]

[a] Laboratory of Cognitive Clinical Sciences, Department of Psychology, University of Bucharest, Bucharest, Romania
[b] Department of Psychology, University of Zurich, Zurich, Switzerland
[c] Doctoral School 'Evidence-based assessment and psychological interventions' and International Institute for the Advanced Studies of Psychotherapy and Applied Mental Health, Babes-Bolyai University, Cluj-Napoca, Romania

HIGHLIGHTS

- No evidence to support a sleeper effect in psychotherapy for anxiety disorders
- No evidence to support that CT's effect on anxiety was more enduring than ET's or vice versa.
- Results favor a uniform efficacy hypothesis regarding ET and CT's enduring efficacy in time.

ARTICLE INFO

Keywords:
Longitudinal meta-analysis
Sleeper effect
Enduring efficacy
Anxiety disorders
Exposure treatment
Cognitive therapy

ABSTRACT

There is a longstanding debate in the cognitive behavioral literature whether exposure-based methods produce more sustainable outcomes relative to cognitive methods or vice versa. This debate concerns particularly the time after treatment termination (at follow-up assessments), also referred to as the *sleeper effect*. Therefore, the aim of the current meta-analysis was to examine the enduring efficacy of Exposure Therapy (ET) in comparison to Cognitive Therapy (CT) from treatment termination to follow-up in anxiety disorders. Available literature also allowed for the assessment of their long-term additive benefits relative to ET only. Traditional random effects analyses with restricted maximum likelihood estimators and multilevel longitudinal analyses were conducted on 39 randomized controlled trials ($N = 1878$). Traditional analyses revealed no differential efficacy at post-treatment or follow-up. Similarly, the multilevel longitudinal analyses identified no differential growth in efficacy from treatment termination to follow-up. The majority of the variables investigated did not moderate the results. However, there was evidence suggesting that CT was superior to ET when treatment was delivered individually, while ET was superior to CT when delivered as group therapy. Overall, the findings did not validate a number of assumptions, such as the existence of a sleeper effect. Several strengths and limitations are further discussed in the paper.

1. Introduction

Anxiety disorders are among the most common mental health problems. The current global prevalence of anxiety disorders ranges from 5.3% in African cultures to 10.4% in Anglo-European cultures (Baxter, Scott, Vos, & Whiteford, 2013). In other words, 1 in 14 people worldwide meet the diagnostic criteria for an anxiety disorder at one point in life (Baxter et al., 2013). Moreover, anxiety disorders are characterized by a considerable rate of relapse, as defined by a return of anxiety symptoms after treatment termination, with estimates of relapse ranging from 19% to 62% (Craske & Mystkowski, 2006; Fava et al., 2001; Newman, Llera, Erickson, Przeworski, & Castonguay, 2013).

It has been argued that considering the relapse rate of anxiety disorders proposed treatments should ideally have lasting effects not limited to post-treatment measurements (Gibby, Casline, & Ginsburg, 2017). There is an ongoing debate about whether particular psychotherapy interventions produce more sustainable outcomes in the months or years following therapy as opposed to other psychotherapies (e.g., Flückiger, Del Re, Munder, Heer, & Wampold, 2014; Flückiger, Del Re, & Wampold, 2015; Kivlighan III et al., 2015). This growth in effect from treatment termination to follow-up is referred to as the sleeper effect (Bell, Marcus, & Goodlad, 2013; Flückiger et al., 2015);

* Corresponding author at: Department of Psychology, University of Bucharest, 90 Panduri Street, 050657 Bucharest, Romania.
E-mail address: ioana.podina@fpse.unibuc.ro (I.R. Podina).

https://doi.org/10.1016/j.cpr.2019.101774
Received 13 November 2018; Received in revised form 31 May 2019; Accepted 19 August 2019
Available online 30 October 2019
0272-7358/

and it remains relatively understudied in the field of psychotherapy research (Flückiger & Del Re, 2017; Wampold & Imel, 2015). The sleeper effect posits that symptom alleviation as a result of intervention may require additional time to materialize.

The sleeper effect focus may have in part contributed to the split within the psychotherapy community where some treatment techniques are believed to ensure more benefits in time relative to others (e.g., Shedler, 2010; Stopa & Clark, 1993; Tolin, 2010); such is the case of Cognitive-Behavioral Therapy (CBT). In CBT, the standalone treatment success of Exposure therapy (ET) and Cognitive therapy (CT) - two of the most widely used complementary CBT methods - raises the issue of whether (a) challenging maladaptive thoughts rather than (b) using standalone behavioral interventions ensures a more durable relief from anxiety symptoms (Hofmann, 2008; Kazantzis et al., 2018; Longmore & Worrell, 2007). This issue is all the more important as it is central to any further inquiry into specific mediators of CBT effects.

Currently, there are only two other notable traditional meta-analyses that investigated the relative efficacy of ET against CT in the treatment of anxiety disorders (Norton & Price, 2007; Ougrin, 2011). Both meta-analyses generally concluded that there appears to be no evidence of differential efficacy between CT and ET with regard to the treatment of anxiety disorders with the exception of social phobia where patients benefited more from CT. Aside from reaching common conclusions, these meta-analyses also shared common limitations that are enlisted below.

First, prior works were traditional meta-analyses pooling effect sizes separately for post-treatment and follow-up. However, change effects in time, such as the sleeper effect, can only be tested through a longitudinal meta-analysis that can effectively investigate treatment by time interactions as argued by (Flückiger & Del Re, 2017; Maas, Hox, & Lensvelt-Mulders, 2004). In this respect, there are only a handful of longitudinal meta-analyses of this kind in psychotherapy research (Flückiger et al., 2014, 2015; Flückiger, Del Re, Wampold, Symonds, & Horvath, 2012; Kivlighan III et al., 2015). In a nutshell, the meta-analysis by Kivlighan III et al. (2015) compared psychodynamic treatment with a broad range of alternative treatments in regard to anxiety and depression; while Flückiger et al. (2014) used longitudinal analyses to investigate the broad spectrum of evidence-based psychotherapies vs. treatment as usual in regard to anxiety and depression. The other two longitudinal meta-analyses by Flückiger et al. (2012, 2015) focused on the relationship between alliance and outcome efficacy across a wide spectrum of treatments, respectively, compared psychotherapies with and without additional components in terms of long-term efficacy. In contrast to prior work that focused on the generalizability of results across various psychotherapy trials and frameworks, the current meta-analysis is specifically aimed at investigating the enduring efficacy of CT and ET for anxiety disorders. This issue should be addressed given the still standing controversies within the CBT framework (e.g., Hofmann, 2008; Kazantzis et al., 2018; Longmore & Worrell, 2007).

Second, although some of the studies investigated combined treatments (i.e., CT plus ET) relative to exposure only, none of the previous meta-analyses explored these potential additive effects at treatment termination or in time. A combined treatment approach would be all the more important as in the CBT literature it is assumed that both behavioral and cognitive components are necessary for an enduring treatment effect and that using both methods would provide better symptom relieve relative to a standalone cognitive or behavioral treatment. However, this idea requires additional support as there are currently no meta-analyses investigating the additive advantage of a combined treatment relative to standalone cognitive or behavioral therapy (e.g., Norton & Price, 2007; Ougrin, 2011).

Third, heterogeneity was assessed only in the Ougrin (2011) meta-analysis and, although moderate to high heterogeneity was found in some cases, the sources of heterogeneity were not explored by subsequent subgroup analyses. A thorough assessment of heterogeneity is required to pinpoint potential moderator variables. Moreover, both meta-analyses require updating since they were published more than six years ago.

Based on the shortage of studies in the literature, we argue that longitudinal meta-analyses that focus on growth of effect sizes from post-treatment to follow-up are long due as many competing approaches claim maximized effects over time with limited evidence supporting these claims (e.g. Flückiger et al., 2014; Flückiger & Del Re, 2017; Kivlighan III et al., 2015).

1.1. The current meta-analysis

In order to address the limitations of the previous literature, we aim to conduct several traditional and multilevel longitudinal meta-analyses: (a) A traditional random effects meta-analysis was employed in order to update prior work in light of more recent publications with regard to the relative efficacy of ET relative to CT at separate points in time (i.e., post-treatment and follow-up). (b) Due to the limited research properly examining enduring treatment efficacy, a multilevel longitudinal meta-analysis examined the sustainability of the effect of CT relative to ET from treatment termination to follow-up. (c) In light of the available literature, we were also able to investigate whether a combined treatment format (ET + CT) provided an additive advantage in time relative to standalone ET following treatment termination. For all these purposes, we considered anxiety-specific measures as outcomes (Wampold et al., 2011).

The study also examined a number of 7 variables that we expected to have a moderating effect on the results, as listed below. Their selection was guided by both theoretical considerations and previous research.

1.2. Potential allegiance effects

Therapist and researcher allegiance refers to the therapist's and, respectively, the researcher's belief that one treatment is preferable over another (e.g., Gaffan, Tsaousis, & Kemp-Wheeler, 1995; Luborsky et al., 1999; Munder, Brütsch, Leonhart, Gerger, & Barth, 2013; Wampold & Imel, 2015). Allegiance might be assessed through the therapist's and/or researcher's advocacy of the preferred therapy at the expense of others or even through the therapist's familiarity with a certain type of therapy. Additionally, direct contribution to the development of a certain form of therapy or to its empirical testing can increase the likelihood of allegiance bias (Falkenström Allegiance Control for Therapists measure; FACT; Falkenström, Markowitz, Jonker, Philips, & Holmqvist, 2013). The two concepts are different, albeit interrelated: while therapist allegiance might manifest itself through treatment delivery and implementation, researcher allegiance may go as far as to affect study design, preferential reporting, and the interpretation of the results. Both therapist and researcher allegiance have been shown to affect the outcomes of clinical trials in the sense that the lack of proper control for allegiance may result in artificially increased effect sizes (Falkenström et al., 2013; Luborsky et al., 1999; Munder, Flückiger, Gerger, Wampold, & Barth, 2012; Munder et al., 2012). Not accounting for allegiance effects might compromise the validity of the conclusions drawn in comparative psychotherapy studies (Munder et al., 2013; Spielmans & Flückiger, 2018; Wampold & Imel, 2015).

1.3. Treatment delivery format

The treatment delivery format, namely individual or group sessions, might be a potential moderator based on previous literature. For example, in the case of generalized anxiety disorder or obsessive-compulsive disorder, meta-analytic findings suggest that individual therapies outperform group therapies (Covin, Ouimet, Seeds, & Dozois, 2008; Eddy, Dutra, Bradley, & Westen, 2004), while in the case of social anxiety no such moderation effect was found (Acarturk, Cuijpers, Van Straten, & De Graaf, 2009). Hence, given that there is no consensus over

the optimal treatment delivery format in the treatment of anxiety, we decided to include this categorical variable in the current meta-analysis.

1.4. Type of anxiety

The type of anxiety disorder is another relevant variable that may impact the results of clinical trials and meta-analyses. From a theoretical standpoint, each disorder presents different symptomatology and, thus, may be differentially impacted by therapy. This theoretical perspective also becomes apparent in the empirical literature. For example, certain trials hinted at the superiority of CT over ET and/or other behavioral techniques in the treatment of social anxiety (Clark et al., 2006; Hofmann, 2004), while for other anxiety disorders (e.g. specific phobia), meta-analyses as well as clinical guidelines highlight exposure as the treatment of choice (Moriana, Gálvez-Lara, & Corpas, 2017; Wolitzky-Taylor, Horowitz, Powers, & Telch, 2008). Hence, the type of anxiety disorder might be a relevant moderator for the purposes of the current meta-analysis.

1.5. Comorbidities

Anxiety disorders record a high and well known comorbidity rate both with depression and other anxiety disorders. Indeed, a report from the *Netherlands Study of Depression and Anxiety* (Klein Hofmeijer-Sevink et al., 2012). The presence of comorbid disorders has been known to affect outcomes of clinical trials for specific anxiety disorders (e.g., Olatunji, Cisler, & Tolin, 2010), which is why most trials tend to control for the presence of comorbidities, either by means of exclusion criteria or otherwise. For the scope of this meta-analysis, the presence of comorbid conditions alongside the primary anxiety diagnostic will be taken into consideration as a dichotomous moderator (i.e., present or absent).

1.6. Treatment dose and number of sessions

Treatment dose mainly refers to whether the number of sessions was matched or not across treatment conditions. The required dose for efficient psychotherapy interventions is subject to ample debates (e.g. Baldwin, Berkeljon, Atkins, Olsen, & Nielsen, 2009). Some meta-analyses identified a positive dose-response relationship, while others found no such connection regarding the treatment of anxiety (e.g., Norton & Price, 2007). In light of these inconclusive results, most comparative efficacy meta-analyses account for this potential moderator in their analyses. As per recommendations by Spielmans and Flückiger (2018), the number of psychotherapy sessions will be treated in the current research as a continuous moderator.

1.7. Publication year

The year of publication might also be a relevant moderator for the current meta-analysis in light of the changing reporting standards of individual trials potentially impacting the efficacy of remission rates (Johnsen & Friborg, 2015) or longitudinal efficacy more generally (Flückiger et al., 2014). With the addition of novel reporting standards, such as intent-to-treat samples, more recent trials might have higher quality reporting standards which may differentially impact in comparison with older trials the end results of the meta-analysis. As the studies included span over a larger period of time (1988–2015), the year of publication is a variable that needs to be considered in the current meta-analysis.

Based on the above-mentioned literature, we expect that (a) the traditional analyses will render no differential results in efficacy between ET and CT at post-assessment and follow-up similar to prior meta-analytic evidence and that (b) multilevel longitudinal analyses will further extend these results in the sense that there will be no differences regarding the enduring effect in time and no evidence favoring a sleeper effect.

2. Methods

2.1. Literature search and selection criteria

Potentially relevant studies were identified following a systematic search of the Web of Science, Scopus, PsychInfo, PubMed, Ebsco, Proquest, Cochrane Library and Google Academic databases, conducted in June 2018. We also searched the references within the most recent systematic reviews and meta-analyses investigating the separate effect of cognitive therapy and exposure in the treatment of anxiety disorders (Norton & Price, 2007; Ougrin, 2011). Combinations of the following keywords were used when researching the databases: “anxiety”, “panic”, “phobia”, “obsessive”, “compulsive”, “*trauma*”, “PTSD”, “cognitive”, “behavior*”, “cognitive therapy”, “CT”, “cbt”, “exposure”, “treatment”, “therapy”, “random*”. The search string that was used is specified in Supplementary Material 1.

Studies were included if they were a) randomized controlled trials (RCTs) in which b) in-person ET was compared with a treatment consisting of CT only and/or ET + CT for c) adults that were d) diagnosed with any type of anxiety disorder, including post-traumatic stress disorder, obsessive-compulsive disorder. The latter two were included as DSM-IV formerly identified them as diagnosed anxiety disorders. All studies were required to include treatment conditions that could be defined as ET, CT or a combination of both. Only anxiety-specific outcomes were relevant for the purposes of the current meta-analysis measured at post-treatment and/or follow-up.

When multiple comparisons were available, we selected only those that met our criteria outlined previously. Studies that presented no comparisons of interest (e.g., relaxation, technology-assisted interventions) or that combined medication with the treatment conditions described above, were excluded. L.A.F. screened initially all records and full-texts were retrieved for further examination. I.R.P. and L.A.F. examined full-texts and selected the eligible RCTs.

2.2. Studies included

A number of 1250 records were identified through database searching and screening of other sources (867 after duplicates were removed). We excluded 807 records based on title/abstract inspection and examined the full-texts for 60 articles. The flowchart of the inclusion process along with specific criteria for study exclusion is presented in Fig. 1, in accordance with the PRISMA statement (Moher, Liberati, Tetzlaff, & Altman, 2009). The selection process resulted in the inclusion of 39 studies that met the inclusion criteria (see the Supplementary Material 2 for a list of the included studies).

2.3. Quality assessment

In order to estimate the quality of the included studies, we used the four criteria of the Cochrane Collaboration Risk of Bias Tool (Higgins et al., 2011). This instrument was developed in order to assess the possible source of bias in RCTs, and the following criteria were rated: a) *random sequence generation* - the adequate generation of the allocation sequence, b) *allocation concealment* - the concealment of allocation to conditions, c) *blinding of outcome assessment* - blinding of participants, personnel and outcome assessors and d) *incomplete outcome data* - the manner in which incomplete data were addressed. The Risk of Bias assessment was performed by two independent raters and disagreements were discussed and resolved by reaching consensus. Ratings varied from unclear Risk of Bias to low Risk of Bias or high Risk of Bias.

There was a substantial interrater agreement with regard to three of the four criteria, respectively Cohen's kappa $k = 0.87$, (95% *CI*: 0.70 to 1.04) for the sequence generation criteria; $k = 0.40$ (95% *CI*: 0.007 to

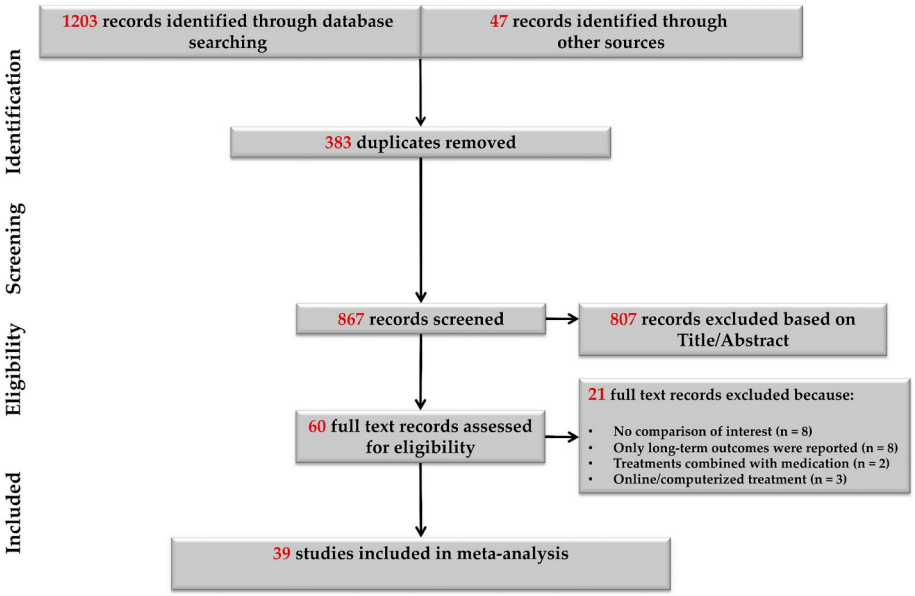

Fig. 1. PRISMA flow-diagram of the study selection process.

0.79) for the allocation concealment criteria; $k = 0.79$ (95% *CI*: 0.60 to 0.98) for the blinding of outcome assessors' criteria; and $k = 0.95$ (95% *CI*: 0.86 to 1.04) for the incomplete outcome data criteria.

2.4. Moderators and extracted data

We extracted a series of variables from the studies included, detailed in Supplementary Table S1, for further use in moderator analyses or study description purposes. Moderator coding was performed by two independent raters and disagreements were discussed and resolved by reaching consensus.

2.4.1. Therapist allegiance

The therapist allegiance variable was assessed with the *Falkenström Allegiance Control for Therapists* instrument (FACT; Falkenström et al., 2013) which is a single item measuring therapist allegiance on a scale from 0 to 3: 0 - The article does not mention the issue of potentially differing therapist allegiances; 1 - The issue of potentially different allegiances is mentioned but not assessed or controlled for in any way; 2 - The authors report making some attempt to control for therapist allegiance. However, they make no attempt to measure therapist allegiance or to statistically control for this variable; 3 - The authors measure therapist allegiance and, if necessary, apply relevant statistical control for this variable. The interrater agreement with regard to this moderator was fair, respectively $k = 0.29$ (95% *CI*: 0.07 to 0.50).

2.4.2. Researcher allegiance

The researcher allegiance variable was coded according to the guidelines provided by Weisz, Jensen-Doss, and Hawley (2006), respectively, Spielmans, Gatlin, and McFall (2010), which recommend using clear *yes* or *no* dichotomous metrics when coding for researcher allegiance effects. Researcher allegiance was rated based on a single item; it was present if (a) the author of the manuscript was a developer of one of the treatments investigated in the paper and/or (b) was a thesis/dissertational committee member if the paper was a dissertation or a PhD thesis. Interrater agreement for this moderator was $k = 0.75$, (95% *CI*: 0.54 to 0.94).

Continuous variables were collected as well. The length of the follow-up treatment was reported in weeks (Supplementary Table S1). Another dimensional variable was the number of sessions used for treatment delivery. Moreover, several subgroup variables were coded in the current work. A yes/no rating system was used to code for the treatment dose, where "yes" signaled an equivalent number of sessions across conditions and "no" indicated a non-equivalent number of treatment sessions. The anxiety subtype was extracted, as well as the treatment delivery format (i.e, individual or group based) and the presence of comorbidities was marked by "yes" (with comorbidities) or "no" (without comorbidities) rating system.

In order to evaluate if drop-out rates were associated with the treatment that was received by the participants, we conducted an additional analysis with the drop-out variable. Drop-outs were defined as all randomized participants not finishing treatment, regardless of the reasons. Odds ratios (ORs) indicated the odds of participants dropping out from the ET versus the comparison groups, with statistically significant ORs below 1 indicating smaller odds for drop-out in the ET group and statistically significant ORs above 1 indicating smaller odds for drop-out in the CT or ET + CT groups. For a similar procedure please see Fodor et al. (2018).

2.5. Meta-analysis

To determine the impact of exposure (ET) when compared to cognitive restructuring (CT) or exposure plus cognitive restructuring (ET + CT), we considered means, standard deviations, and the number of participants for every study at every assessment time. If the study addressed more than one comparison of interest, every direct comparison of interest was coded (e.g., ET vs. CT, ET vs. ET + CT). If reported, we used intent-to-treat information from studies (k = 7; i.e., Arntz, 2002; Ressick, 2002; Bryant et al., 2003, 2008; Foa et al., 2005; Vaccaro et al., 2014; Weck et al., 2015). If multiple measures were reported, the effect sizes (ESs) were aggregated by accounting for the sample size and within-study correlation among outcome measures of 0.50 (Del Re & Hoyt, 2010; Wampold et al., 1997). All within-study ESs were converted to unbiased Hedges' g (Del Re & Hoyt, 2010). A positive g indicates that CT/ET + CT was superior to ET (measures indicating better functioning were reverse scored).

The targeted outcome was labeled *anxiety-specific* and it included trait or state measures that were centered around the primary diagnosis of a particular anxiety disorder, clinician-administered and/or self-reported. This was the case of instruments such as the Liebowitz Social Anxiety Scale for social anxiety, the Specific Phobia Questionnaire for specific phobias, or the Yale-Brown Obsessive-Compulsive Scale for

OCD. Moreover, there was a substantial interrater agreement for this outcome, $k = 0.95$.

Two different within-study ES were considered and therefore two different meta-analyses for each of the two comparisons (i.e. ET vs. CT, ET vs. ET + CT) were conducted. To standardize the follow-up assessments across studies, three different assessments times were coded: (0) Post-treatment assessments, (1) follow-up assessments between 1 and 11 months (i.e., FU1), and (2) follow-up assessments between 12 and 18 months (i.e., FU2). This modality of splitting the time points followed the procedure set by Flückiger et al. (2014). For the purposes of the present study, time was analyzed grand-mean centered (Raudenbush & Bryk, 2002).

For each assessment time, we first conducted separate analyses based on traditional random effects restricted maximum likelihood estimators with known variance (Hedges & Olkin, 1985; Viechtbauer, 2005). These analyses were conducted using the *R* statistical software packages for meta-analysis "MAd" (Del Re & Hoyt, 2014) and "metafor" (Viechtbauer, 2010). The "*MAd*" package was used for extracting information for effect size calculation (when other indicators than means were reported), for computing individual studies effect size(s), within-study aggregation of effect sizes, omnibus analyses, and for categorical moderator analyses. The "*metafor*" package was used for computing the heterogeneity, asymmetry, and publication bias testing, as well as for continuous moderator analyses (i.e., mixed effect models with moderators). We also used the R statistical software package "ggplot2" (Wickham, 2016) for generating graphics.

For the multilevel longitudinal analyses, we used hierarchical linear modeling (HLM; Raudenbush, Bryk, Cheong, Congdon, & Du Toit, 2011) with 2 levels where the assessment times at level 1 (i.e. termination, FU1, and FU2) were nested within direct comparisons in the studies at level 2. More specifically, a mixed model $g = y_{00} + y_{10} * Time + u_0 + u_1 * Time$ was used, where y_{00} represented the grand-mean centered intercept (i.e. time was centered to the overall mean from termination to FU2), y_{10} represented the growth/slope over the various assessment times (i.e., test of the time by group interaction), and u_0 and u_1 represented the error terms at Levels 1 and 2 (random effects model). The growth/slope estimates are based on 17 contrasts for the ET vs. CT comparison and 18 contrasts for the ET vs. ET + CT comparison, with more than one reported assessment time (Raudenbush & Bryk, 2002).

There were three overlap situations where the same sample was used in more than one study (i.e., Cottraux et al., 2001 & Olatunji et al., 2010; Marks et al., 1998 & Lovell et al., 2001; Van Balkom et al., 1998 & Van Oppen et al., 1995). Hence, we grouped these studies under the same identification number in order to properly address this issue in the subsequent analyses. Furthermore, the study of Van Oppen and collaborators (1995) contained the exact analyses performed on a subsample included by Van Balkom et al. (1998). Therefore, we decided to only include in our analyses the data reported in Van Balkom et al. in order to avoid double reporting of the same results.

Heterogeneity was assessed using the *Q* and I^2 statistics (Higgins & Thompson, 2002). If the *Q* statistic is significant, it can be assumed that the effects aggregated in the analysis are heterogeneous and a moderator analysis is justified. I^2 is an index of heterogeneity that is computed as a percentage and reflects the proportion of variability in effects size that is due to true differences among the studies. Heterogeneity assessed with the I^2 statistic can be interpreted in the following manner: a value of 0% indicates no observed heterogeneity, ≤25% indicates low heterogeneity, ≤50% indicates moderate heterogeneity, and 75% and above indicates high heterogeneity. To identify publication bias, asymmetry was tested based on rank correlation (Begg & Mazumdar, 1994) and regression tests (Egger, Smith, Schneider, & Minder, 1997). Furthermore, if the asymmetry tests were significant, the funnel plot was examined via the "*trim and fill*" procedure (Duval & Tweedie, 2000). To detect possible factors that moderate the differences in ES between ET and CT/ET + CT, moderator analyses were conducted.

For the drop-out analysis, because we expected rather small trials - with some reporting few or zero drop-outs - we employed Peto's method (Yusuf, Peto, Lewis, Collins, & Sleight, 1985). This method is recommended both by simulation studies and systematic review methodology standards (Cheng, Pullenayegum, Marshall, Iorio, & Thabane, 2016; Higgins et al., 2011). Trials with zero drop-outs in both arms were excluded, given their potential for inflating bias, especially in the case of small trials (Cheng et al., 2016). In order to calculate Peto odds ratios, we extracted the number of drop-outs and the number of remaining participants in each treatment arm, from each study. Finally, a pooled overall drop-out statistic was derived, by conducting an odds ratio meta-analysis, using the "metan" package (Harris et al., 2007) available for the STATA software (Stata SE, version 15).

2.6. Power analysis

Retrospective power calculations were undergone in R, following the standard method described in Valentine, Pigott, and Rothstein (2010), as well as in Jackson and Turner (2017). Power was computed taking into account the number of effect sizes included in the analyses, as well as the average magnitude of the observed effect size, average group size across studies, and heterogeneity. Power estimates varied from 0.15 to 0.40, which is characteristic of comparative psychotherapy meta-analyses (e.g., Flückiger et al., 2015).

As the power and the sample size of the studies included is a critical and explanatory element in a meta-analysis (Kazdin & Whitley, 2003), we also conducted power analyses for every single study using the G*Power software (Faul, Erdfelder, Lang, & Buchner, 2007) and assuming a power of 0.80 and an alpha of 0.05. The threshold for a clinically relevant treatment effect in mood disorders has been estimated to be around g = 0.24 (Cuijpers, Turner, Koole, Van Dijke, & Smit, 2014). Because comparative psychotherapy studies should be expected to result in small effect sizes, we calculated in G*Power how many participants would be required to detect an effect size of at least 0.24. We found that the required number of participants to detect a clinically significant change in such a trial would be 114. Only two of the included studies had sufficient power sample wise to detect an effect size of at least 0.24 (i.e., Foa et al., 2005; Resick et al., 2002; see Supplementary Table S1).

3. Results

3.1. Characteristics of included studies

The characteristics of the included studies are presented in Supplementary Table S1. The 39 RCTs included 38 relevant comparisons (i.e. 18 ET vs. CT comparisons and 20 ET vs. ET + CT comparisons[1]). The most frequent anxiety subtypes were OCD and PTSD followed by social phobia, panic disorder, and specific phobia. Most of the treatments were administered individually and included between 1 and 20 sessions, with the majority of the studies reporting an equal number of sessions across treatment conditions. The types of exposure that were employed consisted of imaginal exposure, in vivo exposure, and a combination of imaginal and in vivo exposure. More than half of the studies included patients presenting comorbidites with other anxiety disorders without providing further details on this aspect. All in all, the length of follow-up varied between 1 month to 18 months.

[1] Multilevel longitudinal analysis could be conducted on 17 out of 18 ET – CT comparisons as one study reported only post-treatment data. Similarly, multilevel longitudinal analysis could be conducted on 18 out of the 20 ET vs. ET + CT comparisons as two studies had only post-treatment outcomes (i.e. Foa et al., 2005 and Öst et al., 2004).

3.2. Risk of bias

Most trials had uncertain Risk of Bias for two of the four domains. There were only two studies that met all four criteria, while five studies met three out of four criteria. Seventeen studies were rated as having a low Risk of Bias on only one of the four criteria, while eight studies did not meet any of the four criteria having a low Risk of Bias. Supplementary Fig. S1 illustrates the percentage of studies with a low, unclear (i.e., authors did not provide enough information) and high Risk of Bias on the four Cochrane quality criteria. Eleven of the 39 studies had a low Risk of Bias on the sequence generation criteria. Twenty-four of the 39 studies had a low Risk of Bias on the blinding of outcome assessors criteria. Eighteen of the 39 studies had a low Risk of Bias on the incomplete outcome data criteria, while only two of the 39 studies were classified as having a low Risk of Bias on the allocation concealment criteria.

3.3. ET vs. CT

The results from the traditional and multilevel longitudinal analyses for the ET – CT contrast across assessments points are summarized in Table 1. Moreover, Supplementary Fig. S2 depicts the aggregated ES for each study and the 95% CI around the ES.

The pooled ES for anxiety was constant across assessment points (i.e., 0.04). None of these effects were significant (see Table 1). Significant heterogeneity was found at post-treatment (I^2 = 49% [*CI* = 10%, 78%]) and FU1 (I^2 = 73% [44%, 91%]). Significance tests of asymmetry indicated that publication bias was not present in the studies included at any assessment point (p > .29). Publication bias was also visually inspected by means of funnel plots (i.e., sample funnel plot – Supplementary Fig. S3). As asymmetry tests were not significant and therefore no indication of publication bias was present, the "trim and fill procedure" (Duval & Tweedie, 2000), which would estimate that the number of missing studies needed to attain complete symmetry, was not computed.

The slope of the multilevel longitudinal analysis (i.e., that included only those 17 studies with repeated assessments) indicated no linear increase or decrease of effects over time (slope: −0.03, p = .49). The slopes for anxiety of every single study are presented in Fig. 2a.

Regarding drop-out rates, fifteen trials reported non-zero drop-outs in at least one group and six trials reported zero drop-outs in both groups (Supplementary Fig. S4). Drop-out rates did not significantly differ between the groups, (OR = 1.19, 95% CI: 0.77 to 1.84, χ^2 = 7.13, p = .93).

3.4. ET vs. ET + CT

The results from the traditional and multilevel longitudinal analyses for the ET – ET + CT contrast across assessments points are summarized in Table 2. Furthermore, Supplementary Fig. S5 depicts the ES for each study and the 95% CI around the ES.

The pooled ES for anxiety ranged from 0.02 to 0.13 overall assessment points. None of these effects were significant (see Table 2). No significant heterogeneity was found at either time point, p > .18. Significance tests of asymmetry indicated that publication bias was not present in the included studies, p > .28. Publication bias was also visually inspected by means of funnel plots (i.e. sample funnel plot – Supplementary Fig. S6). As asymmetry tests were not significant and therefore no indication of publication bias was present, the "trim and fill" procedure (Duval & Tweedie, 2000) was not needed.

The slope of the multilevel longitudinal analysis indicated no linear increase or decrease of effects over time (slope: 0.01, p = .69). The slopes for anxiety of every single study are presented in Fig. 2b.

For drop-out rates, sixteen trials reported non-zero drop-outs in at least one group and five trials reported zero drop-outs in both groups (Supplementary Fig. S7). Drop-out rates did not significantly differ between the groups, (OR = 0.87, 95% CI: 0.62 to 1.20, χ^2 = 11.87, p = .68).

3.5. Moderator analyses

In order to test for moderators, we first ran separate analyses based on traditional random effects models to detect possible factors that moderate the effect sizes of ET versus CT for anxiety at the assessment points where significant heterogeneity was found (i.e., post-treatment and FU1). Moderators that were significant in the traditional random effect models were then tested using multilevel models.

Out of all the tested moderators using the traditional random effects model, only the treatment delivery format was a significant moderator at post-treatment (p = .049; Table 3). More specifically, CT was superior to ET when treatment was delivered individually (g = 0.09, k = 16), while ET was superior to CT when delivered as group therapy (g = −0.45, *k* = 2). Multilevel longitudinal models supported the moderating effect of individually delivered treatment vs. group therapy of the traditional random effects models (γ_{01} = − 0.52, SE = 0.32, *t* (15) = 1.6, see Table 3). However, this result should be interpreted with caution as the group therapy condition included only 2 comparisons. Furthermore, treatment delivery format was not a significant moderator at follow-up (p = .52; Table 3).

4. Discussion

The current study investigated the relative efficacy of ET in comparison and in addition to CT for anxiety disorders at post-treatment and follow-up. In contrast to prior meta-analyses, the focus of the current study was on the differential growth in efficacy in the time elapsed from treatment termination to follow-up. The main findings are discussed below.

Table 1
Traditional and multilevel longitudinal analyses for CT vs. ET.

Longitudinal model					
	Post	FU1 (1–11 months)	FU2 (12–18 months)	Intercept	Slope
k	18	12	7	17[a]	
g	0.04	0.04	0.04	0.05	−0.03
g (CI)	−0.11, 0.19	−0.23, 0.3	−0.17, 0.26	−0.09, 0.19	−0.07, 0.04
Q	33.7[b]	36.8[c]	9.5	50.7[c]	15.5
I^2	49%	73%	35%	68%	0%
I^2 (CI)	10%, 78%	44%, 91%	0%, 88%	26%, 81%	0%, 23%

[a] Longitudinal model: 36 effect sizes at Level 1 nested within 17 contrasts at level 2.
[b] p < .01.
[c] p < .001.

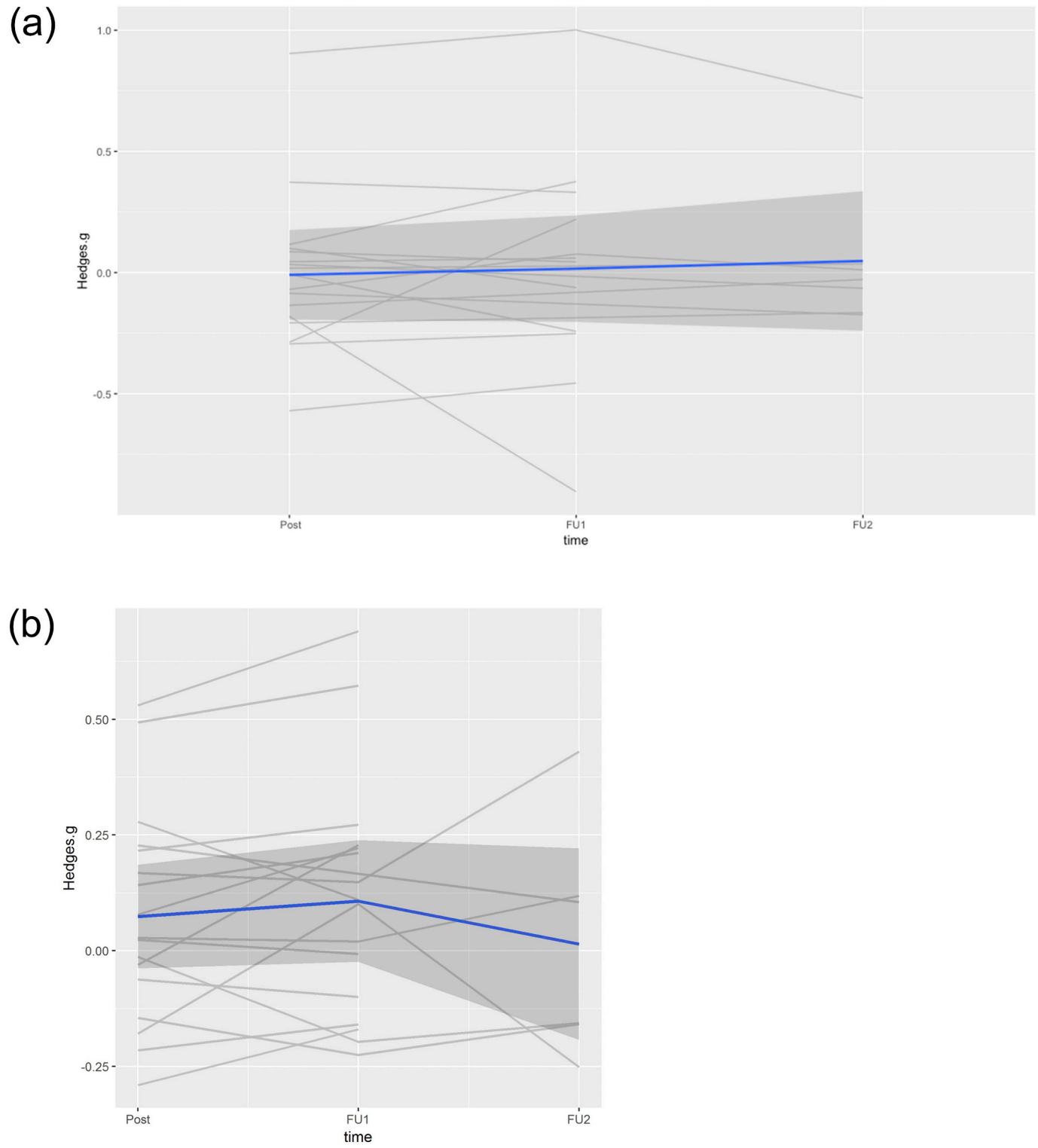

Fig. 2. (a) & (b). The slope of the longitudinal analysis, that included only those studies with repeated assessments, from treatment termination to FU1 and FU2 in the case of the ET vs. CT contrast (2a) and the ET vs. ET + CT contrast (2b). The blue line represents the overall growth over time, while the individual lines represent growth over time within each study. The light grey shading around the overall line represents its standard error. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.1. Main analyses

The traditional random effects analyses with restricted maximum likelihood estimators indicated that there were no significant differences in relative efficacy between ET and CT and, most importantly, no additive benefits at either post-treatment or follow-up. The lack of differential efficacy does not in any way imply that either ET or CT are ineffective. Moreover, the results are in line with prior traditional meta-analyses that found small and non-significant differences between the two interventions (i.e., Norton & Price, 2007; Ougrin, 2011). The results join the growing body of research that cues to *uniform efficacy* between different methods of therapy (e.g., Elliott, Barker, & Hunsley,

Table 2
Traditional and multilevel longitudinal analyses for ET + CT vs. ET.

Longitudinal model					
	Post	FU1 (1–11 months)	FU2 (12–18 months)	Intercept	Slope
k	20	17	6	18[a]	
g	0.13	0.13	0.02	0.09	0.01
g (CI)	0.001, 0.25	−0.01, 0.27	−0.22, 0.25	−0.05, 0.23	−0.12, 0.13
Q	24.42	17.51	3.22	15.5	1.96
I^2	29.56%	18.81%	0%	0%	0%
I^2 (CI)	0%, 58%	0%, 57%	0%, 77%	0% 28%	0% 0%

[a] Longitudinal model: 40 effect sizes at Level 1 nested within 18 contrasts at level 2.

Table 3
Moderation analysis for the ET vs. CT contrast.

	Traditional random effect model			Longitudinal model	
Moderators	Post	FU1			
	$Q_b(df)$ *kg*	$Q_b(df)$ *kg*		*y (SE)*	*t*
TF	3.88 (1)*	34 (1)	Intercept y_{00}	03 (.09)	3
			TF y_{01}	−52 (.32)	−1.6 +
Individual	16.09	10.07	Slope y_{10}	−06 (.31)	−01
Group	2.45	2.14	*TF* y_{11}	31 (.31)	1.0

Notes. TF = treatment format.
+ $p < .10$.
* $p < .05$.

2014; Wampold et al., 1997).

None of the multilevel longitudinal analyses of change from termination to follow-up produced statistically significant effects. There is, therefore, no evidence to favor a sleeper effect in the current work. In fact, there is marginal empirical evidence to support a sleeper effect in comparative psychotherapy research in general (Flückiger et al., 2012; Flückiger et al., 2014, 2015; Flückiger & Del Re, 2017; Kivlighan III et al., 2015). As in the case of the current work, none of the previous longitudinal meta-analyses found a growth in relative efficacy at follow-up irrespective of what form of psychotherapy was investigated. Moreover, some researchers argue that there is no reason whatsoever to assume that these effects would appear especially at follow-up (Flückiger & Del Re, 2017).

There are several explanations for the recurrent lack of differential efficacy in the CBT literature and in the current results, presented as it follows.

(1) Prior meta-analyses chose to explain the non-differential results through a process-based approach (e.g., Ougrin, 2011). According to this approach, change in maladaptive cognitions takes place in both CT and ET, hence the overlap in results. However, more process-based assessments are needed in support of this assumption. In fact, a recent meta-review concluded that the evidence for process-outcome relations in CBT is just emerging (Kazantzis et al., 2018).
(2) Other explanations include the possibility that patients may not have actually improved due to specific behavioral or cognitive mechanisms of change, but rather because of general factors common to both conditions (e.g. Wampold & Imel, 2015; Norcross & Lambert, in press) such as the client's attributions regarding therapeutic change (Powers, Smits, Whitley, Bystritsky, & Telch, 2008), clients' expectations for improvement (Constantino, Vîslă, Coyne, & Boswell, 2019), clients' perception of treatment credibility (Constantino, Boswell, Iles, Coyne, & Vîslă, 2019), and therapeutic alliance (Flückiger, Del Re, Horvath, & Wampold, 2018).
(3) A third methodological explanation argues that potential confounders such as dropout rates, external events over time, and additional treatment may make it difficult to detect differences in relative efficacy at follow-up (Durham, Higgins, Chambers, Swan, & Dow, 2012; Flückiger & Del Re, 2017).
(4) Assuming the non-significant differences are a reflection of a uniform efficacy, the results are seemingly in line with the Dodo bird interpretation, which favors no meaningful differences among evidence-based psychotherapies (e.g., Luborsky, Singer, & Luborsky, 1975; Wampold & Imel, 2015). However, the present meta-analysis was not conceptualized to test non-inferiority trials (Sedgwick, 2013). As such, the present results indicate no superiority of one approach over another in the sense of small effect size differences.

Finally, this is the first meta-analysis to investigate the additive effects of ET plus CT, yet no differential efficacy using traditional or multilevel longitudinal analyses was found. This result could also have alternative explanations. A highly likely alternative could be that in an effort to match treatment-dose across conditions, the intensity of the dose of each component in the ET + CT condition was split in half which could have reduced any additive advantage on efficacy. Another explanation could be that ET and CT may have various substantial overlaps in their theoretical assumptions and their collaborative decision-making process, making it difficult to separate one from another (e.g., Mulder, Murray, & Rucklidge, 2017).

4.2. Moderator analyses

Treatment delivery format was a significant moderator of the results in both traditional and multilevel longitudinal analyses. Even more, CT was superior to ET when treatment was delivered individually, while ET was superior to CT when delivered as group therapy. A potential explanation of this result could be that CT for anxiety is more efficient when delivered individually than in a group format as the challenge of maladaptive thoughts should be tailored depending on each patient. However, ET could benefit from a group format and may challenge the patient to surpass his/her fear following the example of another patient. Nevertheless, this result should be interpreted with caution as the group therapy condition included only two comparisons.

None of the remaining variables produced statistically significant effects in both traditional and multilevel longitudinal analyses. Interestingly, the type of anxiety disorder was not a significant moderator as well. This result contradicts previous assumptions and some trials indicating that CT is more efficacious for anxiety disorders with a theorized cognitive mechanism, such as social anxiety (Clark et al., 2006; Hofmann, 2004), or that ET is more efficacious for specific phobia where clinical guidelines highlight exposure as the treatment of choice (Moriana et al., 2017; Wolitzky-Taylor et al., 2008).

4.3. Strengths and limitations

The present study adds to the narrow list of meta-analyses that employed a multilevel longitudinal model in comparative psychotherapy research. An important attribute of longitudinal models is that it enables the computation of the growth (slope) of repeated

assessment points, ultimately delivering a more accurate overview of the trajectory of change in time. Though a few notable longitudinal meta-analyses have been conducted, this is the first research study of its kind to examine the growth in effect sizes beyond ET's and CT's treatment completion. Other existing meta-analyses on the topic (Norton & Price, 2007; Ougrin, 2011) have only computed and reported effect sizes at pre-established moments in time (i.e., post-treatment, follow-up), a method that does not allow an estimate of the effect size over time.

Nonetheless, the current work also has several limitations described as it follows. First, it was not clear whether booster sessions were used in post-treatment due to a lack of reporting on this matter. Second, the vast majority of RCTs had an unclear Risk of Bias across domains. Only two studies could be rated as low Risk of Bias on all domains preventing us from reliably assessing the relationship between overall Risk of Bias and outcome. The only Risk of Bias domain where most trials reported information was incomplete outcome data. Half of the studies rated high risk on the attrition bias category, again questioning the reliability of the ES estimations, as the exclusion of participants from RCT analyses was shown to distort outcomes (Abraha et al., 2015). Now that CONSORT (Montgomery et al., 2018) for social and psychological interventions (CONSORT-SPI) is available, future trials should seek to better address and report the Risk of Bias in their trials.

Third, the current longitudinal meta-analysis is underpowered to detect close to 0 effect sizes and more longitudinal publications would be required for highly reliable results. However, there is some reassurance in our results in the sense (a) that traditional analyses match prior work and (b) that the non-significant longitudinal analyses are fully in line with the scarcity of longitudinal meta-analyses on comparative psychotherapy. Unfortunately, this seems to be the case of psychotherapy in general, where out of hundreds of studies that contrast the results of different interventions only a handful have enough power to detect clinically significant small differences (e.g., Cuijpers, van Straten, Bohlmeijer, Hollon, & Andersson, 2010). The studies included in the current meta-analysis are no exception, only two out of the 39 studies had enough sample size to detect small significant differences. Future comparative psychotherapy trials should definitely improve on this aspect and have enough power to identify significant small effect sizes as there is no reasoning to assume otherwise.

Fourth, there was considerable heterogeneity of the pooled ESs. Though we did include a comprehensive list of moderators, we were not able to fully account for the heterogeneity of the ET-CT contrast suggesting the presence of undetermined variables related to the heterogeneity of ESs. This is usually an indicator that studies differ one from another in some manner. But it may be that they do not differ on the traditional aspects reported in these studies, but on variables that are usually not accounted for. Prior meta-analyses on the topic have addressed several of the moderators that we addressed here and were not able to fully account for the heterogeneity in their results either (e.g., Ougrin, 2011). There are several details that are almost never reported in these studies, such as details regarding common factors that are known to be relevant for the success of an intervention (Horvath, Del Re, Flückiger, & Symonds, 2011).

Finally, though our work is the first ET/CT meta-analytical research to account for the less commonly studied moderator of therapist allegiance, the interrater agreement for this variable was low. Evidently, this casts doubt on the reliability of the subgroup analysis resulted in the case of this moderator, and could also suggest that the rating instrument that we used could improve in clarity. There are other limitations pertaining to this moderator that reflect the current state of affairs in the literature. In their recent work, Spielmans and Flückiger (2018) made an overview of potential moderators that are highly relevant to test the generalizability of effects across psychotherapy trials, these are (a) the structural equivalence of interventions, (b) preferences/allegiances, (c) therapist effects, and (d) sample representativeness. Their analysis of 15 meta-analyses published in 2016 indicated that these non-specific variables are often overlooked as potential moderators in psychotherapy research. Similar to Spielmans and Flückiger (2018), in the current research, most of the studies did not acknowledge therapist allegiance as a variable. However, lack of reporting does not entail the absence of bias and additional studies could shed some light on this aspect in the future.

5. Conclusions

The findings of the current work did not validate a number of assumptions in the literature, (a) such as the existence of a sleeper effect favoring either one of the investigated interventions, (b) that CT is more efficacious for anxiety disorders with a theorized cognitive mechanism (Hanrahan, Field, Jones, & Davey, 2013; Hofmann, 2004) while ET is more efficacious for anxiety disorders relying on conditioning, or (c) that ET should be preferably used in combination with a cognitive intervention for more enduring results (Kaczkurkin & Foa, 2015). Moreover, we posit that the lack of a differential efficacy between CT and ET does not whatsoever suggest that either of these interventions are inefficient. Rather, they may be equivalently efficacious when contrasted one to another at the end of the treatment and from that point onward. However, for clear-cut results, we need studies that are considerably larger in sample size than the ones that have been conducted so far and ideally more longitudinal trials in order to adequately power a meta-analysis to detect close to 0 effect sizes that are characteristic of comparative psychotherapy research.

Overall, judging by the longitudinal results of the current meta-analysis and prior work, there seems to be no reasoning to support a substantial and clinically meaningful sleeper effect in comparative psychotherapy research in general, and specifically not in the CT/ET distinction. Clearly, therapists and patients can continue to individually prefer an approach over another. However, overall, the necessity to choose a cognitive only intervention over a behavioral only approach or vice versa seems to be questionable in light of non-differential omnibus results. Hence, in the future, the literature on CT/ET can continue in either of the three ensuing modalities but should channel its efforts in one direction for clear cut results, which we are lacking so far. First, it can pursue to seek differences on other criteria than treatment efficacy, such as pan-theoretical criteria (e.g. therapist allegiance, therapist effects, patient characteristics). Second, it can continue to seek differences at a process-based level (e.g., change in maladaptive thoughts, patient-therapist relationship), while managing to separate through design the seemingly shared active ingredients which currently seems to be a difficult mission. Third, it can commit to an integrative approach with both cognitive and behavioral methods within it; this suggestion is in line with the multimodal and seemingly integrative CBT efforts towards a meta/generic practice that prevents against the division of CBT in numerous schools.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cpr.2019.101774.

Role of funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Contributors

IRP, AV, LAF and CF designed the study and wrote the meta-analytic protocol. IRP and LAF conducted literature searches, provided summaries of previous research, and coded the study characteristics. AV conducted the analyses in consultation with CF and IRP. IRP wrote the first draft of the manuscript and critical feedback was provided by AV, LAF, and CF. All authors have approved the final manuscript.

Declaration of Competing Interest

All authors declare that they have no conflicts of interest.

Acknowledgements

The authors thank Ana Cosmoiu, Claudiu Papasteri for assistance with the literature search and input on analyses, as well as Mirela Mohan for proofreading the manuscript.

References

Abraha, I., Cherubini, A., Cozzolino, F., De Florio, R., Luchetta, M. L., Rimland, J. M., ... Eusebi, P. (2015). Deviation from intention to treat analysis in randomised trials and treatment effect estimates: Meta-epidemiological study. *Bmj, 350*, h2445. https://doi.org/10.1136/bmj.h2445.

Acarturk, C., Cuijpers, P., Van Straten, A., & De Graaf, R. (2009). Psychological treatment of social anxiety disorder: A meta-analysis. *Psychological Medicine, 39*(2), 241–254. https://doi.org/10.1017/S0033291708003590.

Baldwin, S. A., Berkeljon, A., Atkins, D. C., Olsen, J. A., & Nielsen, S. L. (2009). Rates of change in naturalistic psychotherapy: Contrasting dose–effect and good-enough level models of change. *Journal of Consulting and Clinical Psychology, 77*(2), 203–211. https://doi.org/10.1037/a0015235.

Baxter, A. J., Scott, K. M., Vos, T., & Whiteford, H. A. (2013). Global prevalence of anxiety disorders: A systematic review and meta-regression. *Psychological Medicine, 43*(5), 897–910. https://doi.org/10.1017/S003329171200147X.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics,* 1088–1101. https://doi.org/10.2307/2533446.

Bell, E. C., Marcus, D. K., & Goodlad, J. K. (2013). Are the parts as good as the whole? A meta-analysis of component treatment studies. *Journal of Consulting and Clinical Psychology, 81*(4), 722–736. https://doi.org/10.1037/a0033004.

Cheng, J., Pullenayegum, E., Marshall, J. K., Iorio, A., & Thabane, L. (2016). Impact of including or excluding both-armed zero-event studies on using standard meta-analysis methods for rare event outcome: A simulation study. *BMJ Open, 6*(8), e010983. https://doi.org/10.1136/bmjopen-2015-010983.

Clark, D. M., Ehlers, A., Hackmann, A., McManus, F., Fennell, M., Grey, N., ... Wild, J. (2006). Cognitive therapy versus exposure and applied relaxation in social phobia: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 74*(3), 568–578. https://doi.org/10.1037/0022-006X.74.3.568.

Constantino, M. J., Boswell, J. F., Iles, B. R., Coyne, A. E., & Vîslă, A. (2019). A meta-analysis of the association between patients' early perception of treatment credibility and their posttreatment outcomes. *Psychotherapy, 55*(4), 486–495.

Constantino, M. J., Vîslă, A., Coyne, A. E., & Boswell, J. F. (2019). A meta-analysis of the association between patients' early-treatment outcome expectation and their post-treatment outcomes. *Psychotherapy, 55*(4), 473–485.

Covin, R., Ouimet, A. J., Seeds, P. M., & Dozois, D. J. (2008). A meta-analysis of CBT for pathological worry among clients with GAD. *Journal of Anxiety Disorders, 22*(1), 108–116. https://doi.org/10.1016/j.janxdis.2007.01.002.

Craske, M. G., & Mystkowski, J. L. (2006). Exposure therapy and extinction: Clinical studies. In M. G. Craske, D. Hermans, & D. Vansteenwegen (Eds.). *Fear and learning: From basic processes to clinical implications* (pp. 217–233). Washington, DC, US: American Psychological Association. https://doi.org/10.1037/11474-011.

Cuijpers, P., Turner, E. H., Koole, S. L., Van Dijke, A., & Smit, F. (2014). What is the threshold for a clinically relevant effect? The case of major depressive disorders. *Depression and Anxiety, 31*(5), 374–378.

Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychological Medicine, 40*(2), 211–223.

Del Re, A. C., & Hoyt W.T., MAd (2010). Meta-analysis with mean differences. *R package version 0.8*. http://CRAN.R-project.org/package=MAd.

Del Re, A. C., & Hoyt W.T., MAd (2014). Meta-analysis with mean differences. *R package version 0.8-2*. Retrieved from http://CRAN.R-project.org/package=MAd.

Durham, R. C., Higgins, C., Chambers, J. A., Swan, J. S., & Dow, M. G. (2012). Long-term outcome of eight clinical trials of CBT for anxiety disorders: Symptom profile of sustained recovery and treatment-resistant groups. *Journal of Affective Disorders, 136*(3), 875–881. https://doi.org/10.1016/j.jad.2011.09.017.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x.

Eddy, K. T., Dutra, L., Bradley, R., & Westen, D. (2004). A multidimensional meta-analysis of psychotherapy and pharmacotherapy for obsessive-compulsive disorder. *Clinical Psychology Review, 24*(8), 1011–1030. https://doi.org/10.1016/j.cpr.2004.08.004.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629.

Elliott, K., Barker, K. K., & Hunsley, J. (2014). Dodo bird verdict in psychotherapy. *The Encyclopedia of Clinical Psychology,* 1–5. https://doi.org/10.1002/9781118625392.wbecp0053.

Falkenström, F., Markowitz, J. C., Jonker, H., Philips, B., & Holmqvist, R. (2013). Can psychotherapists function as their own controls? Meta-analysis of the "crossed therapist" design in comparative psychotherapy trials. *The Journal of Clinical Psychiatry, 74*(5), 482–491. https://doi.org/10.4088/JCP.12r07848.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Fava, G. A., Rafanelli, C., Grandi, S., Conti, S., Ruini, C., Mangelli, L., & Belluardo, P. (2001). Long-term outcome of panic disorder with agoraphobia treated by exposure. *Psychological Medicine, 31*(5), 891–898. https://doi.org/10.1017/s0033291701003592.

Flückiger, C., & Del Re, A. C. (2017). The sleeper effect between psychotherapy orientations: A strategic argument of sustainability of treatment effects at follow-up. *Epidemiology and Psychiatric Sciences, 26*(4), 442–444. https://doi.org/10.1017/S2045796016000780.

Flückiger, C., Del Re, A. C., Horvath, A. O., & Wampold, B. E. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis [APA task force]. *Psychotherapy,* 1–61. https://doi.org/10.1037/pst0000172 free access.

Flückiger, C., Del Re, A. C., Munder, T., Heer, S., & Wampold, B. E. (2014). Enduring effects of evidence-based psychotherapies in acute depression and anxiety disorders versus treatment as usual at follow-up—A longitudinal meta-analysis. *Clinical Psychology Review, 34*(5), 367–375. https://doi.org/10.1016/j.cpr.2014.05.001.

Flückiger, C., Del Re, A. C., & Wampold, B. E. (2015). The sleeper effect: Artifact or phenomenon—a brief comment on Bell et al.(2013). *Journal of Consulting and Clinical Psychology, 83*(2), 438–442 (doi:10.1037%2Fa0037220).

Flückiger, C., Del Re, A. C., Wampold, B. E., Symonds, D., & Horvath, A. O. (2012). How central is the alliance in psychotherapy? A multilevel longitudinal meta-analysis. *Journal of Counseling Psychology, 59*(1), 10. https://doi.org/10.1037/a0025749.

Fodor, L. A., Coteț, C. D., Cuijpers, P., Szamoskozi, Ş., David, D., & Cristea, I. A. (2018). The effectiveness of virtual reality based interventions for symptoms of anxiety and depression: A meta-analysis. *Scientific Reports, 8*(1), 10323. https://doi.org/10.1038/s41598-018-28113-6.

Gaffan, E. A., Tsaousis, J., & Kemp-Wheeler, S. M. (1995). Researcher allegiance and meta-analysis: The case of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 63*(6), 966–980. https://doi.org/10.1037//0022-006X.63.6.966.

Gibby, B. A., Casline, E. P., & Ginsburg, G. S. (2017). Long-term outcomes of youth treated for an anxiety disorder: A critical review. Clinical child and family psychology adding an additional component to an existing treatment vis-à-vis the existing treatment. *Review, 20*(2), 201–225. https://doi.org/10.1007/s10567-017-0222-9.

Hanrahan, F., Field, A. P., Jones, F. W., & Davey, G. C. L. (2013). A meta-analysis of cognitive therapy for worry in generalized anxiety disorder. *Clinical Psychology Review, 33*(1), 120–132. https://doi.org/10.1016/j.cpr.2012.10.008.

Harris, R., Bradburn, M., Deeks, J., Harbord, R., Altman, D., Steichen, T., & Sterne, J. (2007). METAN: Stata module for fixed and random effects meta-analysis. *Stata Journal, 8*, 3–28.

Hedges, L., & Olkin, I. (1985). *Statistical models for meta-analysis.* San Diego, CA: Academic Press.

Higgins, J. P., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., ... Sterne, J. A. C., Cochrane Bias Methods Group, & Cochrane Statistical Methods Group (2011). The cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ, 343*, d5928. https://doi.org/10.1136/bmj.d5928.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558. https://doi.org/10.1002/sim.1186.

Hofmann, S. G. (2004). Cognitive mediation of treatment change in social phobia. *Journal of Consulting and Clinical Psychology, 72*(3), 392–399. https://doi.org/10.1037/0022-006x.72.3.393.

Hofmann, S. G. (2008). Common misconceptions about cognitive mediation of treatment change: A commentary to longmore and Worrell (2007). *Clinical Psychology Review, 28*(1), 67–70. discussion 71–74 https://doi.org/10.1016/j.cpr.2007.03.003.

Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48*(1), 9–16. https://doi.org/10.1037/a0022186.

Jackson, D., & Turner, R. (2017). Power analysis for random-effects meta-analysis. *Research Synthesis Methods, 8*(3), 290–302.

Johnsen, T. J., & Friborg, O. (2015). The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. *Psychological Bulletin, 141*(4), 747–768. https://doi.org/10.1037/bul0000015.

Kaczkurkin, A. N., & Foa, E. B. (2015). Cognitive-behavioral therapy for anxiety disorders: An update on the empirical evidence. *Dialogues in Clinical Neuroscience, 17*(3), 337–346.

Kazantzis, N., Luong, H. K., Usatoff, A. S., Impala, T., Yew, R. Y., & Hofmann, S. G. (2018). The processes of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research, 42*(4), 349–357. https://doi.org/10.1007/s10608-018-9920-y.

Kazdin, A. E., & Whitley, M. K. (2003). Treatment of parental stress to enhance therapeutic change among children referred for aggressive and antisocial behavior. *Journal of Consulting and Clinical Psychology, 71*(3), 504.

Kivlighan, D. M., III, Goldberg, S. B., Abbas, M., Pace, B. T., Yulish, N. E., Thomas, J. G., ... Wampold, B. E. (2015). The enduring effects of psychodynamic treatments Vis-à-Vis alternative treatments: A multilevel longitudinal meta-analysis. *Clinical Psychology Review, 40*, 1–14. https://doi.org/10.1016/j.cpr.2015.05.003.

Klein Hofmeijer-Sevink, M., Batelaan, N. M., van Megen, H. J. G. M., Penninx, B. W., Cath, D. C., van den Hout, M. A., & van Balkom, A. J. L. M. (2012). Clinical relevance of comorbidity in anxiety disorders: A report from the Netherlands study of depression and anxiety (NESDA). *Journal of Affective Disorders, 137*(1–3), 106–112. https://doi.org/10.1016/j.jad.2011.12.008.

Longmore, R. J., & Worrell, M. (2007). Do we need to challenge thoughts in cognitive behavior therapy? *Clinical Psychology Review, 27*(2), 173–187. https://doi.org/10.1016/j.cpr.2006.08.001.

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., ... Schweizer, E. (1999). The researcher's own therapy allegiances: A "wild card" in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6*(1),

95–106. https://doi.org/10.1093/clipsy.6.1.95.
Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies: Is it true that everyone has won and all must have prizes? *Archives of General Psychiatry, 32*(8), 995–1008.
Maas, C. J. M., Hox, J. J., & Lensvelt-Mulders, G. J. L. M. (2004). Longitudinal meta-analysis. *Quality & Quantity: International Journal of Methodology, 38*, 381–389.
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*(4), 264–269. https://doi.org/10.1016/j.jclinepi.2009.06.005.
Montgomery, P., Grant, S., Mayo-Wilson, E., Macdonald, G., Michie, S., Hopewell, S., ... Yaffe, J. (2018). Reporting randomized trials of social and psychological interventions: The CONSORT-SPI 2018 extension. *Trials, 19*, 407. https://doi.org/10.1186/s13063-018-2733-1.
Moriana, J. A., Gálvez-Lara, M., & Corpas, J. (2017). Psychological treatments for mental disorders in adults: A review of the evidence of leading international organizations. *Clinical Psychology Review, 54*, 29–43. https://doi.org/10.1016/j.cpr.2017.03.008.
Mulder, R., Murray, G., & Rucklidge, J. (2017). Common versus specific factors in psychotherapy: Opening the black box. *The Lancet Psychiatry, 4*(12), 953–962. https://doi.org/10.1016/S2215-0366(17)30100-1.
Munder, T., Brütsch, O., Leonhart, R., Gerger, H., & Barth, J. (2013). Researcher allegiance in psychotherapy outcome research: An overview of reviews. *Clinical Psychology Review, 33*(4), 501–511. https://doi.org/10.1016/j.cpr.2013.02.002.
Munder, T., Flückiger, C., Gerger, H., Wampold, B. E., & Barth, J. (2012). Is the allegiance effect an epiphenomenon of true efficacy differences between treatments? A meta-analysis. *Journal of Counseling Psychology, 59*(4), 631–637. https://doi.org/10.1037/a0029571.
Newman, M. G., Llera, S. J., Erickson, T. M., Przeworski, A., & Castonguay, L. G. (2013). Worry and generalized anxiety disorder: A review and theoretical synthesis of evidence on nature, etiology, mechanisms, and treatment. *Annual Review of Clinical Psychology, 9*(1), 275–297. https://doi.org/10.1146/annurev-clinpsy-050212-185544.
Norton, P. J., & Price, E. C. (2007). A meta-analytic review of adult cognitive-behavioral treatment outcome across the anxiety disorders. *The Journal of Nervous and Mental Disease, 195*(6), 521–531. https://doi.org/10.1097/01.nmd.0000253843.70149.9a.
Olatunji, B. O., Cisler, J. M., & Tolin, D. F. (2010). A meta-analysis of the influence of comorbidity on treatment outcome in the anxiety disorders. *Clinical Psychology Review, 30*(6), 642–654. https://doi.org/10.1016/j.cpr.2010.04.008.
Ougrin, D. (2011). Efficacy of exposure versus cognitive therapy in anxiety disorders: Systematic review and meta-analysis. *BMC Psychiatry, 11*(1), https://doi.org/10.1186/1471-244x-11-200.
Powers, M. B., Smits, J. A., Whitley, D., Bystritsky, A., & Telch, M. J. (2008). The effect of attributional processes concerning medication taking on return of fear. *Journal of Consulting and Clinical Psychology, 76*(3), 478–490. https://doi.org/10.1037/0022-006x.76.3.478.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods. Vol. 1*. Sage.
Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & Du Toit, M. (2011). *Hierarchical linear and nonlinear modeling (HLM7)*. Lincolnwood, IL: Scientific Software International1112.
Sedgwick, P. (2013). What is a non-inferiority trial? *BMJ, 347*. https://doi.org/10.1136/bmj.f6853.
Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *American Psychologist, 65*(2), 98–109. https://doi.org/10.1037/a0018378.
Spielmans, G. I., & Flückiger, C. (2018). Moderators in psychotherapy meta-analysis. *Psychotherapy Research, 28*(3), 333–346. https://doi.org/10.1080/10503307.2017.1422214.
Spielmans, G. I., Gatlin, E. T., & McFall, J. P. (2010). The efficacy of evidence-based psychotherapies versus usual care for youths: Controlling confounds in a meta-reanalysis. *Psychotherapy Research, 20*(2), 234–246. https://doi.org/10.1080/10503300903311293.
Stopa, L., & Clark, D. M. (1993). Cognitive processes in social phobia. *Behaviour Research and Therapy, 31*(3), 255–267. https://doi.org/10.1016/0005-7967(93)90024-o.
Tolin, D. F. (2010). Is cognitive–behavioral therapy more effective than other therapies? A meta-analytic review. *Clinical Psychology Review, 30*(6), 710–720. https://doi.org/10.1016/j.cpr.2010.05.003.
Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics, 35*(2), 215–247.
Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics, 30*(3), 261–293. https://doi.org/10.3102/10769986030003261.
Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), https://doi.org/10.18637/jss.v036.i03.
Wampold, B. E., Budge, S. L., Laska, K. M., Del Re, A. C., Baardseth, T. P., Flűckiger, C., ... Gunn, W. (2011). Evidence-based treatments for depression and anxiety versus treatment-as-usual: A meta-analysis of direct comparisons. *Clinical Psychology Review, 31*(8), 1304–1312. https://doi.org/10.1016/j.cpr.2011.07.012.
Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate - the evidence for what makes psychotherapy work*. New York, NY: Routledge.
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: Empiricially, "all must have prizes". *Psychological Bulletin, 122*(3), 203–215. https://doi.org/10.1037/0033-2909.122.3.203.
Weisz, J. R., Jensen-Doss, A., & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist, 61*(7), 671–689. https://doi.org/10.1037/0003-066X.61.7.671.
Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.
Wolitzky-Taylor, K. B., Horowitz, J. D., Powers, M. B., & Telch, M. J. (2008). Psychological approaches in the treatment of specific phobias: A meta-analysis. *Clinical Psychology Review, 28*(6), 1021–1037. https://doi.org/10.1016/j.cpr.2008.02.007.
Yusuf, S., Peto, R., Lewis, J., Collins, R., & Sleight, P. (1985). Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Progress in Cardiovascular Diseases, 27*(5), 335–371.

Ioana R. Podina, PhD, is a psychologist and lecturer at the Faculty of Psychology and Educational Sciences (FPES), University of Bucharest (UB). Dr. Podina is an internationally certified psychotherapist (Albert Ellis Institute, New York, USA). She currently coordinates the Laboratory of Clinical Cognitive Sciences (FPES, UB). Dr. Podina is also the Chairman of the Research Committee of the Society for the Exploration of Psychotherapy Integration (SEPI) and a representative for Romania of the Society for Psychotherapy Research (SPR). Her main research interests are focused on the development of evidence-based psychotherapeutic interventions, investigation of the mechanism of change in psychotherapy and the identification of predictors of relapse in psychopathology. Dr. Podina has published several scientific articles, specifically meta-analyses some of which were published in Clinical Psychology Review and other peer reviewed journals.

Andreea Vîslă, PhD, received her Ph.D. in Psychology from Babes-Bolyai University, Romania. She is currently a postdoctoral researcher in the Department of Psychology at the University of Zurich where she is involved in projects aiming at investigating the enduring efficacy of psychological interventions for depression and anxiety disorders. Dr. Visla's research interests focus on mechanisms involved in psychopathology and psychotherapeutic treatment, process and outcome research, randomized clinical trials research designs, researcher and provider/therapist effect, and cross-cultural differences in psychopathology. In investigating these topics, Dr. Visla employs meta-analytic approaches, multilevel longitudinal models, and ecological momentary assessment.

Liviu A. Fodor is currently a Ph.D. student enrolled in the "Evidence-based assessment and psychological interventions" Doctoral School. He also currently works as assistant researcher at the International Institute for the Advanced Studies of Psychotherapy and Applied Mental Health, Babeş-Bolyai University. His research interests include the multilevel analysis of technologically augmented interventions used in the treatment of anxiety disorders. Liviu does research in Quantitative Psychology, Cognitive Science and Clinical Psychology.

Christoph Flückiger, PhD, is professor at the Department of Psychology for Psychological Interventions and Psychotherapy at University of Zurich and head of the affiliated psychotherapy laboratory. Dr. Flückiger does research predominantly in Clinical Psychology, Psychotherapy and Quantitative, meta-analytic Psychology. He is currently Incoming Editor (2019) for the journal "Psychotherapy Research" of the International Society of Psychotherapy Research (SPR).