

VU Research Portal

The effects of cognitive behavioral therapy are not systematically falling: a revision of Johnsen & Friborg.

Cristea, I.A.; Hollon, S.D.; David, D.; Karyotaki, E.; Cuijpers, Pim; Stefan, S.I.

published in

Psychological Bulletin
2017

DOI (link to publisher)

[10.1037/bul0000062](https://doi.org/10.1037/bul0000062)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Cristea, I. A., Hollon, S. D., David, D., Karyotaki, E., Cuijpers, P., & Stefan, S. I. (2017). The effects of cognitive behavioral therapy are not systematically falling: a revision of Johnsen & Friborg. *Psychological Bulletin*, 143(3), 326-340. <https://doi.org/10.1037/bul0000062>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

COMMENT

The Effects of Cognitive Behavioral Therapy Are Not Systematically Falling: A Revision of Johnsen and Friborg (2015)

Ioana A. Cristea

Babeş-Bolyai University and University of Padova

Simona Stefan

Babeş-Bolyai University

Eirini Karyotaki

Vrije University Amsterdam and EMGO Institute for Health and Care Research, Amsterdam, the Netherlands

Daniel David

Babeş-Bolyai University

Steven D. Hollon

Vanderbilt University

Pim Cuijpers

Vrije University Amsterdam and EMGO Institute for Health and Care Research, Amsterdam, the Netherlands

In a meta-analysis, Johnsen and Friborg (2015) reported a significant negative relationship between publication year and the effect sizes (ESs) of cognitive-behavioral therapy (CBT) for depressive disorders, suggesting its effectiveness was falling. We identified a series of methodological and conceptual caveats and consequently redid the meta-analysis. We used the same inclusion criteria, but only included randomized controlled trials and searched for additional eligible trials. We computed both within-group and between-group ESs for the CBT arm for the Beck Depression Inventory (BDI) and the Hamilton Rating Scale for Depression (HRSD). We assessed risk of bias, sample size, type of control group, and the study's country of origin and conducted subgroup, single, and multiple meta-regression analyses including publication year and other moderators. We identified 30 additional eligible trials. Within-group ESs presented huge heterogeneity estimates (I^2 around 90%). Year of publication was significant in some single meta-regression analyses on the BDI, but not significant in others, in most analyses on the HRSD, and in any of the analyses on between-group ESs. Multiple regression models indicated that either year was not significantly related or that both year and country were significantly related to outcomes, with a temporal trend present solely in US studies. Year of publication does not appear to be a reliable and independent moderator of the effectiveness of CBT for depression. The linear "fall" reported by Johnsen and Friborg (2015) is most likely a spurious finding.

Keywords: cognitive behavioral therapy, depression, year of publication, meta-analysis, moderator

Supplemental materials: <http://dx.doi.org/10.1037/bul0000062.supp>

A recent meta-analysis (Johnsen & Friborg, 2015) sent ripples through the psychotherapy community, when its results suggested that the effects of cognitive-behavioral therapy (CBT) for depression seem to "have declined linearly and steadily since its introduction" (p. 747). The authors conducted a meta-analysis of trials of CBT for depression and used meta-

regression procedures to examine the relationship between the year of publication and the effect sizes (ESs) in the CBT intervention arm, examining outcomes on two widely used depression scales: the Beck Depression Inventory (BDI) and the Hamilton Rating Scale for Depression (HRSD). Their results showed a significant decrease of the ESs over time for both the

Ioana A. Cristea, Department of Clinical Psychology and Psychotherapy, Babeş-Bolyai University and Department of General Psychology, University of Padova; Simona Stefan, Department of Clinical Psychology and Psychotherapy, Babeş-Bolyai University; Eirini Karyotaki, Department of Clinical Psychology, Vrije University Amsterdam and EMGO Institute for Health and Care Research, Amsterdam, the Netherlands; Daniel David, Department of Clinical Psychology and Psychotherapy, Babeş-Bolyai University; Steven D. Hollon, Department of Psychology, Vanderbilt University; Pim Cuijpers, Department of Clinical Psychology, Vrije University Amsterdam and EMGO Institute for Health and Care Research.

Ioana A. Cristea and Simona Stefan were supported for this work by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project PN-II-RU-TE-2014-4–1316 awarded to Ioana A. Cristea.

Correspondence concerning this article should be addressed to Ioana A. Cristea, Department of Clinical Psychology and Psychotherapy, Babeş-Bolyai University, 37 Republicii Street, Cluj-Napoca, CJ 400015, Romania or Department of General Psychology, University of Padova, Via Venezia 8, 35131 Padova, Italy. E-mail: ioanacristea@psychology.ro

BDI (slope = -0.029 , $p < .001$), as well the HRSD (slope = -0.02 , $p = .01$).

These results received a lot of attention in both scientific outlets and the general media, with glum headlines announcing that CBT may have “had its day as a treatment for depression” (Graham, 2015) or wistfully wondering why CBT “is falling out of favour” (Burkeman, 2015). Unfortunately, most of the discussion has focused on the possible explanatory mechanisms behind these findings, with a plethora of speculations ranging from procedural issues related to the delivery of CBT and a possible reduction in the integrity of delivery over time, to more general ones, such as CBT following the temporal dynamics of the “placebo” effect. Much less consideration has been given to the *actual* methodology by which these results came about in Johnsen and Friberg (2015).

Nevertheless, we identified a number of questionable or unclear aspects. *First*, the authors included both randomized and nonrandomized trials. This issue is chief since randomization of participants to treatment groups serves to ensure that sources of bias are equally distributed across the groups, with the only difference among them being the intervention. Nonrandomized trials are subject to a whole array of sources of bias to a degree we have no way of appropriately gauging. Effects in these trials might be due to many factors other than the intervention, such as the passage of time, nonspecific factors like patients’ expectations (the “placebo” effect), variables unbeknownst to the experimenter actually causing the changes, and so forth. *Second*, even if we just consider randomized controlled trials (RCTs), the number of included trials appeared to be relatively small (52) compared to other meta-analyses of CBT for depression (Barth et al., 2013; Cuijpers et al., 2013). *Third*, we noticed that the majority of heterogeneity values (I^2) reported by the authors were extremely high (around 90%), suggesting that the variability among effect sizes was so large and consequently the studies so inconsistent, that they would almost seem to have very little in common. That raises serious doubts about the combination of the included studies in the meta-analysis (Ioannidis, Patsopoulos, & Evangelou, 2007). It is likely that at least one of the sources for this high heterogeneity may be related to the combination of different types of studies, like randomized and nonrandomized. *Fourth*, the authors chose to focus on “completer” analyses instead of *intent-to-treat* (ITT) analyses. Yet landmark meta-epidemiological studies (Nüesch et al., 2009) have demonstrated that not only do effect sizes from trials with patient exclusions (i.e., restricted to completers) tend to benefit the intervention more than trials without exclusions (ITT), but that the differences in ESs between trials with and without exclusions also seem to be more pronounced in meta-analyses with high heterogeneity among included trials, as seems to be the case of Johnsen and Friberg (2015). Also, “completer” data are usually based on a reduced number of participants and particularly exposed to the attrition bias, given that drop-out might not be random and may be related to the intervention. On the other hand, ITT analyses have been criticized for being too conservative (Gupta, 2011), although the available methods for conducting these analyses have evolved over time. Furthermore, it is conceivable that “completer” analyses may be more affected by changes over time in the effects for CBT than ITT analyses. Johnsen and Friberg (2015) did not find differences between studies reporting “completer” versus ITT data in subgroup analyses, but this does not necessarily imply that using one of these data sets is equivalent to the other. *Fifth*, the paper

relies on within-group effect sizes, which are problematic in that they allow the integration of findings from studies with diverse designs in a way that artificially inflates the estimation of the effectiveness of a treatment by leveling differences due to methodological quality and the larger context of the trial. Meta-analyses relying on within-group ESs run the risk of crediting the intervention with a lot of extraneous influences such as spontaneous remission, regression to the mean, background variables that favored the intervention (e.g., participant expectations), and increased heterogeneity due to different time spans between baseline and posttest in included studies. *Finally*, the authors’ main analyses considered publication year as a single moderator in single meta-regression analysis. The authors also ran two-way interaction analysis between year and other variables considered. But an important test that was missing was conducting multiple analyses using more than one of the potential moderators. Looking at each moderator separately or in combinations of two is misleading as it increases the number of multiple tests carried out on the same dataset and thus the risk of Type I error and of seemingly significant findings that may be spurious.

Furthermore, at least two *other* important variables have also been shown to change with the passing of time. One is trial *quality*, with a recent meta-analysis (Chen et al., 2014) that examined historical changes in the quality and quantity of psychotherapy trials for depression revealing that trials improved over time on most quality criteria considered. Another variable is sample size, as earlier CBT trials had small samples and it has been demonstrated for trials of psychotherapy in general that larger studies yield smaller effect sizes (Barth et al., 2013; Cuijpers, van Straten, Bohlmeijer, Hollon, & Andersson, 2010). In fact, meta-analyses of CBT for depression also found evidence of this small sample bias (Barth et al., 2013; Furukawa et al., 2014). Thus, the apparent decline in the efficiency of CBT over time might be masking the fact that trials were simply becoming larger and of better quality (i.e., lower risk of bias). Johnsen and Friberg (2015) did look at both these variables. For quality, they showed that it was positively correlated with time, but did not find a significant two-way interaction between it and year of publication in their relationship with ESs. For sample size, they redid the analyses excluding studies with samples under an arbitrarily defined cut-off of 20, and showed that year remained a significant predictor of ES in single meta-regression analysis. However, this definition of a cut-off point is unfounded and certainly does not exclude the potentially biasing effect of sample size. Similarly, given that at least these three potential moderators (year, quality of the trial, sample size) seem at least partly intercorrelated and each related to depression outcomes, their inclusion in the same meta-regression model would have been necessary. It could be that in this case, one or all stop being predictive of outcome. Moreover, other moderator variables could also be relevant, such as the type of control group used, given that the effects of CBT have been shown to vary with different control groups (Furukawa et al., 2014).

Our aim was to redo the meta-analysis conducted by Johnsen and Friberg (2015), to clarify if year of publication remained a significant moderator of the effects of CBT for depression *after* (a) addressing potentially flawed or questionable methodological choices; (b) adding missing RCTs that satisfied the original authors’ criteria to the analyses; (c) examining between-group in addition to within-group ESs for the trials with an additional

control group; and (d) conducting multiple meta-regression analyses including other moderators shown to be related to the efficiency of CBT.

Method

Identification and Selection of Studies

Our starting point for study selection was represented by the studies included by [Johnsen and Friberg \(2015\)](#). We used the same inclusion criteria. Studies were included if (a) the implemented intervention was “pure” CBT, (b) designed to treat depression, (c) delivered by therapists trained in CBT, (d) in an individual face-to-face format for (e) adults with (f) an unipolar depressive disorder of any kind as their primary psychiatric diagnosis, (g) with a pretreatment BDI score over 13.5 and who (h) did not have acute physical illnesses or bipolar or psychotic disorders, (i) with reported outcomes on either the BDI or the HRSD. Similarly to [Johnsen and Friberg \(2015\)](#), we included studies where participants received medication in addition to CBT, as well as studies on inpatients. The only additional filter we added was that studies had to be RCTs. We also included studies published in a language other than English, because language restriction may easily result in missing relevant studies ([Higgins & Green, 2011](#), section 10.2.2.4).

However, we did not employ other methodological choices we considered questionable. If a study contained more than one subgroup of participants, [Johnsen and Friberg \(2015\)](#) included only the one with the highest severity. There is no justification for this choice, which could in fact be construed as data selection. Therefore, if the study included more subgroups, we included them all and pooled them at a study level. We also averaged results at study level if the study included multiple comparisons, like different variations of CBT (for instance, CBT with varying number of sessions, or theoretically diverse CBT orientations). On the other hand, for studies that included both CBT and a truncated form of it (for instance, CBT without some components), or conversely CBT and CBT with added elements (e.g., motivational interviewing), we only included the “pure” CBT condition, like [Johnsen and Friberg \(2015\)](#). Similar to them, for studies with both a CBT and a CBT plus medication arm, we only included the former.

After selecting the RCTs from the studies included by [Johnsen and Friberg \(2015\)](#), we verified their inclusion criteria against a database of 1,756 papers on the psychological treatment of depression, in order to identify potential missed studies. This database has been described in detail elsewhere ([Cuijpers, van Straten, Warmerdam, & Andersson, 2008](#)) and has been used in a series of published meta-analyses (<http://www.evidencebasedpsychotherapies.org>). It was developed through a comprehensive literature search (from 1966 to January 2015) in which 16,365 abstracts from PubMed (4,007 abstracts), PsycInfo (3,147 abstracts), Embase (5,912 abstracts), and the Cochrane Central Register of Controlled Trials (3,995 abstracts) were examined. This database is continuously updated. These abstracts were identified by combining terms indicative of psychological treatment and depression (both MeSH terms and text words). Two researchers performed the selection of additional studies independently from each other. Disagreements were resolved through discussion.

Study Quality (Risk of Bias)

As a proxy for study quality, [Johnsen and Friberg \(2015\)](#) used a scale for randomized trials called the Randomized Controlled Trial Psychotherapy Quality Rating Scale (RCT-PQRS; [Kocsis et al., 2010](#)). However, a recent analysis ([Armijo-Olivo, Fuentes, Ospina, Saltaji, & Hartling, 2013](#)) of tools of evaluating methodological qualities of RCTs revealed inconsistencies between the items in these tools and the Cochrane Collaboration Risk of Bias (RoB; [Higgins et al., 2011](#)). For the RCT-PQRS, this analysis revealed the scale included items not linked to any type of bias and hence irrelevant for study quality (for example, a priori relevant hypotheses that justify comparison group[s]), and also noted its validity, particularly criterion validity, had not been assessed in any way. For these reasons, as well as to increase the comparability of our results with other meta-analyses of psychotherapy for depression, we used four items from the RoB tool to assess trial quality. These included adequate generation of the allocation sequence (selection bias); concealment of allocation to conditions (selection bias); blinding of assessors to outcome (detection bias); and dealing with incomplete outcome data (attrition bias). The fifth item of the RoB tool, selective outcome reporting, was not included as it is impossible to assess given the vast majority of psychotherapy trials, especially older ones, have not been independently registered.

Blinding of outcome assessors was rated as positive either if the study described proper methods of ensuring blinding or if all outcome measures were self-report scales, thus not requiring the interaction with an assessor. Blinding of participants is impossible to ensure in psychotherapy trials and was not rated. Dealing with incomplete data was assessed as positive if ITT analyses were conducted, meaning the authors employed a method of including all randomized participants in the analysis.

So as to use study quality as a moderator in analyses, we also computed a “risk of bias” score for each study, by giving one point to each criterion for which a study could be rated as low RoB. In this way, each included study had a score denoting *low* RoB, ranging from 0 to 4, with higher values indicating *low* RoB on more criteria. Two independent researchers rated risk of bias, and disagreements were resolved through discussion. We also computed Cohen’s Kappa interrater agreement for each of the four items of the Cochrane Risk of Bias tool, prior to resolving disagreements.

Potential Moderators

Given that one of our goals was to conduct multiple meta-regression analyses looking at multiple predictors in the same model, we chose to focus on variables that had already been studied in the literature of psychotherapy for depression and shown to be related to outcome. To maximize statistical power, we included, as much as possible, variables that would be described in *most* reports, including the older ones.

Some of the moderators considered by [Johnsen and Friberg \(2015\)](#) (gender, baseline severity) have been shown to not be related to outcomes in meta-analyses of CBT or psychotherapy for depression ([Cuijpers et al., 2014](#); [Driessen, Cuijpers, Hollon, & Dekker, 2010](#)). Also, others were more appropriate for separate analyses (e.g., BDI version, ITT vs. “completer” analyses), and we opted for this. Finally, for some variables (e.g., proportion of patients with comorbidity or

taking medication, therapist competence), few reports contained data or the information was often imprecise.

Therefore, as continuous variables, we included year of publication, number of subjects randomized in the CBT arm, and number of CBT sessions. We also included the following categorical moderators: recruitment type (clinical, defined as recruitment from help-seeking general practice populations or outpatient samples vs. others); diagnosis (major depressive disorder diagnosed according to a system such as the *DSM* or the Research Diagnostic Criteria versus others); target group (adults vs. specific populations, such as elderly, postpartum depression, general medical conditions); patient type (outpatient vs. inpatient); type of control group (waitlist, care-as-usual or placebo vs. others); type of CBT (use of Beck manual vs. others); country (U.S. vs. others).

Data Synthesis and Meta-Analysis

Outcome measures. Similar to Johnsen and Friborg (2015), we used the BDI and HRSD as outcome measures. We also coded the version of the BDI (BDI-I; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961 versus BDI-II; Beck, Steer, & Brown, 1996) and HRSD (Hamilton, 1960) used in the study. If a study used multiple versions of the HRSD, the one with 17 items was preferred.

Unlike Johnsen and Friborg (2015), we did not include treatment recovery rates because (a) the rates were based on the BDI and duplicated the information already obtained through the use of the BDI while giving the false impression of a new outcome, and (b) the pooling of absolute response rates is a problematic, insufficiently reliable, procedure in itself, because it typically results in extremely high levels of heterogeneity. Hence it is usually discouraged to pool such rates (Higgins & Green, 2011), sections 9.4.4. and 9.4.8).

We also differed in the choice regarding the preferential use of “completer” versus ITT data, in the cases where both were available. We included both “completer” and ITT data for the studies reporting both and conducted separate analysis: one in which we chose ITT over “completer” outcome data (ITT preferred) and one in which we did the opposite (“completer” preferred).

Effect size calculation. To calculate mean effect sizes, we used the computer program Comprehensive Meta-Analysis/CMA (version 3.3.070).

Johnsen and Friborg (2015) used an unusual combination of two types of ES calculation. For studies not including a no-intervention control group, they calculated a *within* ES for the CBT arm (*prepost within* design in the original paper), computing a standardized mean difference (Cohen’s *d*). Importantly, they did this for both nonrandomized trials, as well as for RCTs in which the CBT arm was being compared to medication or another psychotherapy. However, for RCTs in which CBT was contrasted to no treatment, a waitlist, or care-as-usual, they computed another type of ES (*controlled* design), by calculating the difference between pre- and posttest scores for the intervention and, respectively, the control group, and then standardizing using the change scores.

There are two major problems with these distinct calculations. First, for the main analysis the authors report all studies together, effectively combining ESs calculated with the first method with others calculated with the second (see Table 2 of their paper). Second, for the prepost within design the authors combined RCTs with non-RCTs, while for the other, the controlled design, they in fact included

only one part of RCTs. In this way, data from RCTs were broken across two categories, making it impossible for the reader to assess the relationship between year and ESs across such trials *only*.

Consequently, we opted for two types of analyses. In the first, we computed prepost (*within*) ESs, using the procedures for the standardized mean difference. Like Johnsen and Friborg (2015), we assumed an $r = .70$ for pre to post correlation and we reported the indicator corrected for small sample bias (Hedges & Olkin, 1985), Hedges’ *g*. In the second, we calculated between groups ESs—the difference between the CBT and the control group at posttest (standardized mean difference), corrected for small sample bias as well (Hedges’ *g*). Because we only included RCTs, we opted for the use of posttest means only, a method not requiring the imputation of a prepost correlation, which is usually absent in studies.

Assessment of heterogeneity. Like Johnsen and Friborg (2015), and because we expected considerable heterogeneity among studies, we used a random effects model (Borenstein, Hedges, Higgins, & Rothstein, 2009) to pool mean effect sizes. We assessed heterogeneity by calculating the I^2 statistic, which indicates heterogeneity in percentages. A value of 0% indicates no observed heterogeneity, whereas larger values define increasing heterogeneity, with 25% as a threshold for low, 50% for moderate, and 75% or above for high (Higgins, Thompson, Deeks, & Altman, 2003). We calculated 95% confidence intervals (CIs) around I^2 (Ioannidis et al., 2007), using the noncentral χ^2 -based approach with the heterogi module for STATA (Orsini, Bottai, Higgins, & Buchan, 2006).

Johnsen and Friborg (2015) did not report I^2 for their main analyses (the only I^2 estimations given are in Table 4 of their paper, which reports subgroup analyses). They also did not specify a definition of outliers, but some studies clearly appeared to have huge ESs and were referred to as outliers in their Results section. We defined outliers as studies in which the 95% CI of the ES was outside the 95% CI of the pooled studies, on both sides.

Meta-Regression Analyses

Testing whether moderators were related to ESs through meta-regression was the main focus of the paper. We conducted both single and multiple analyses. Meta-regression analyses were conducted according to a random effects model using the Knapp-Hartung method. More detailed single meta-regression analyses were conducted with year as a moderator for: all included studies; excluding studies on inpatients; excluding outliers; only on studies with low RoB (more than 2 RoB criteria rated as low risk); and separately for the BDI-I, BDI-II, and HRSD-17. These analyses were done both favoring ITT over “completer” analyses and vice versa. Two multiple regression models were used: one in which only predictors shown to be significant in single meta-regression analyses were included (parsimonious model) and another including all predictors (full model).

Summary of Analyses

We present two separate sets of analyses. One includes only the RCTs contained in Johnsen and Friborg (2015), and focuses on within-group effects sizes, like in their main analyses. The added element here, aside from correcting some small methodological lapses, are multiple meta-regression analyses including more potentially relevant predictors. Although in some cases we presented

both, this first analysis gave priority to “completer” data over ITT so as to enhance similarity to the original meta-analysis. The goal is to give the reader a chance to ascertain how stable and specific the moderating effect of year on depression outcomes is. The second set is our version of a “gold standard” analysis, in which we adjust *all* the methodological choices we consider questionable and assess whether under these conditions we can still evidence a systematic, linear decline in the efficacy of CBT. In this latter analysis, we only retain the inclusion criteria of [Johnsen and Friborg \(2015\)](#) and conduct a new meta-analysis that in our view best improves the original.

Results

Selection and Inclusion of Studies

[Johnsen and Friborg \(2015\)](#) included a total of 52 RCTs. We found four of them to not meet their own inclusion criteria: one ([McBride, Atkinson, Quilty, & Bagby, 2006](#)) reported outcomes for a subsample of an already included study ([Quilty, McBride, & Bagby, 2008](#)); another ([Forman, Herbert, Moitra, Yeomans, & Geller, 2007](#)) included participants selected to have a score on the BDI or the Beck Anxiety Inventory over 9, so not necessarily depressed; another ([Brown, Evans, Miller, Burgess, & Mueller, 1997](#)) included participants with alcoholism, not depression, as a primary diagnosis and a BDI score over 10; and finally, another ([Lieberman & Eckman, 1981](#)) primarily targeted suicide attempters, only some of whom were also depressed, and implemented pure behavior therapy. Also, another included study ([Kalapatapu et al., 2014](#)) was a secondary analysis (identified as such even in the title) of another trial ([Mohr et al., 2012](#)), which was not included. We therefore opted for the inclusion of the larger, more complete trial.

For two trials ([Gallagher-Thompson & Steffen, 1994](#); [McLean & Hakstian, 1979](#)), we found it impossible to calculate pre-to-post ESs in the CBT arm for the BDI or HRSD, despite the fact these were reported by [Johnsen and Friborg \(2015\)](#). Two studies ([Hollon et al., 1992](#); [Quilty et al., 2008](#)) had only been included in the analysis for the BDI, even if they also reported data on the HRSD. One study ([Jacobson, Dobson, Fruzzetti, Schmalings, & Salusky, 1991](#)) was wrongly listed by the authors as a controlled (not randomized) trial, even though it was in fact an RCT.

Our database search identified 30 additional trials that satisfied the criteria of [Johnsen and Friborg \(2015\)](#). Out of these, 25 reported data for the BDI and 23 for the HRSD. Thirteen studies included outcome data for both the ITT and “completer” (12 for the BDI and 9 for the HRSD). A full list of the included studies is included in the references.

Characteristics and Risk of Bias of Included Studies

Selected characteristics of included studies are presented in Appendix S1. Risk of bias in the included studies was considerable. The Kappa interrater agreement coefficients were 0.85 for sequence generation, 0.88 for concealment of allocation, 0.82 for blinding for outcome assessors, and 0.76 for dealing with incomplete data, showing good interrater agreement even prior to resolving disagreements. Estimates for each study are given in Appendix S1. Only 17 of the 77 (22%) included studies had low RoB on 3

or 4 of the four criteria considered. Fifteen studies (19%) had high or uncertain RoB on *all* the criteria considered.

Analysis Restricted to RCTs Included in [Johnsen and Friborg \(2015\)](#)

We note that all analyses (main effects, simple and multiple meta-regression) yielded almost identical results when all studies identified as eligible (the studies in the original meta-analysis plus the extra studies identified) were considered. A synthetic version of these results is presented in Appendix S2.

Main Effects of CBT (Within-Group ESs)

“Completer” analyses were favored to maximize similarity with the original meta-analysis. With the correction of some smaller errors described above, we included 45 trials for the BDI resulting in a Hedges’ $g = 1.64$ (95% CI 1.46 to 1.82), and respectively 30 trials for the HRSD, with a g of 1.70 (95% CI 1.51 to 1.88), values very similar to those of [Johnsen and Friborg \(2015\)](#), 1.58 for the BDI, and 1.69 for the HRSD. Heterogeneity was extremely high (84% for the BDI and 78% for the HRSD). Forest plots for both the BDI and the HRSD are presented in the supplementary Figures S3 and S4.

Single Meta-Regression Analyses

Single meta-regression analyses with year of publication and the other moderators, as well as multiple meta-regression, are presented in [Table 1](#) and [Table 2](#). They showed a significant relationship between year of publication and outcome for the BDI in most analyses, with the exception of low RoB studies and studies on the BDI-II only. For the HRSD, the relationship was significant in half of the analyses. Notably, when ITT data were favored, results were similar for the BDI, but *all* indicated nonsignificant relationships between year and outcomes for the HRSD. Country of publication (U.S. vs. others) was also a significant moderator of outcome for the analyses on the BDI, $b_1 = 0.61$, $p = .01$.

Multiple Meta-Regression Analyses

In order to avoid collinearity, we calculated the correlations among the predictors and found them all to be under 0.60. Multiple meta-regression favoring “completers” showed that either year was not significantly related to outcome (both the parsimonious and full models for the HRSD; the full model for the BDI) or that both year and country were significantly related to outcomes (the parsimonious model for the BDI and for the BDI and HRSD taken together). We examined the relationship between year and within-group ESs separately for U.S. and non-U.S. studies. We focused this analysis on the BDI because both year and country were significant in the parsimonious model. Results are displayed in [Figure 1](#). For the 25 studies conducted in the U.S., there was a significant negative relationship between year and outcomes ($b_1 = -0.04$; 95% CI: -0.07 to -0.01 ; $p = .01$) that was not present for the 20 studies conducted outside of the U.S. ($b_1 = -0.02$, $p = .18$). Of note, the ESs from the *earlier* studies were higher for studies conducted in the U.S. than for those conducted outside, and subsequently, over time, started to go down

Table 1
Single Meta-Regression Analyses With Year of Publication, Within-Group Effect sizes, Studies Included in Johnsen & Friborg (2015) Only

| Outcome measure | ITT preferred over “completer” | | | “Completer” preferred over ITT | | |
|--------------------------------|--------------------------------|----------|-----------------------|--------------------------------|----------|-----------------------|
| | Coeff | 95% CI | <i>p</i> ^a | Coeff | 95% CI | <i>p</i> ^a |
| BDI (I + II) | | | | | | |
| All studies | -.03 | -.05–.01 | .0019 | -.04 | -.06–.02 | .001 |
| Inpatients excluded | -.03 | -.05–.01 | .0026 | -.03 | -.06–.01 | .001 |
| Outliers excluded | -.02 | -.03–.00 | .02 | -.02 | -.03–.01 | .005 |
| Inpatients & outliers excluded | -.02 | -.03–.00 | .03 | -.02 | -.03–.01 | .010 |
| Only low RoB studies (>2) | .02 | -.01–.05 | .19 | -.01 | -.05–.03 | .50 |
| BDI-I only | -.03 | -.07–.00 | .03 | -.03 | -.07–.00 | .04 |
| BDI-II only | -.04 | -.10–.02 | .19 | -.04 | -.10–.02 | .14 |
| HRSD (all versions) | | | | | | |
| All studies | -.02 | -.04–.01 | .12 | -.03 | -.05–.00 | .03 |
| Inpatients excluded | -.02 | -.04–.01 | .15 | -.02 | -.05–.00 | .04 |
| Outliers excluded | .01 | -.01–.02 | .37 | -.01 | -.03–.01 | .26 |
| Inpatients & outliers excluded | .01 | -.01–.03 | .24 | -.01 | -.03–.01 | .39 |
| Only low RoB studies (>2) | .05 | -.03–.13 | .19 | .01 | -.07–.09 | .79 |
| Only HRSD-17 | -.02 | -.05–.00 | .10 | -.03 | -.05–.00 | .02 |

Note. BDI = Beck Depression Inventory; HRSD = Hamilton Rating Scale for Depression; ITT = intent-to-treat; RoB = risk of bias.

^a The *p* levels in these columns indicate whether the relationship between year and effect size is significant in meta-regression analyses (significant results are in bold).

to values closer to what non-U.S. studies engendered from the beginning.

“New” Improved Meta-Analysis (Additional Studies Added, Between-Group ESs, ITT Data Preferred Over “Completer”)

Main effects of CBT (between-group ESs). For the BDI (see Figure 2), 29 trials resulted in a Hedges’ *g* of 0.72 (95% CI 0.56 to 0.89), with moderate heterogeneity ($I^2 = 50%$). For the HRSD (see Figure 3), 19 trials aggregated in a Hedges’ *g* of 0.79 (95% CI 0.62 to 0.96), with moderate heterogeneity ($I^2 = 37%$). Only 4 studies had both ITT and “completer” data.

Single meta-regression analyses. Single meta-regression analyses (see Table 3) indicated no significant relationships between year of publication and outcomes, neither for the BDI nor for the HRSD. However, the number of participants in the CBT group and respectively recruitment from clinical samples were significant predictors in all cases (BDI and HRSD averaged together, BDI only, HRSD only).

Multiple meta-regression analyses. Year of publication (see Table 4) was not significant in any of the analyses. The analyses combining the BDI and the HRSD, which were the most statistically powerful, found the number of subjects in the CBT arm to be significantly negatively related to outcome in the parsimonious model. This also was a significant predictor in the full model, together with the study’s country of origin. Multiple regression on the BDI and respectively the HRSD only found recruitment from clinical samples as significantly and negatively associated to outcome, but these analyses were based on a small number of trials, and consequently the association might be spurious.

Discussion

The recent meta-analysis by Johnsen and Friborg (2015) suggested that the efficiency of CBT for depression was declining

over time. We identified some potentially erroneous or unclear methodological aspects in their paper. Building on previous research showing that the quality and sample sizes of depression trials had also changed over time, we wondered whether the apparent decline in the effects of CBT over time might simply be masking a well-documented phenomenon in outcome research: trials simply become larger and better (i.e., less vulnerable to biases) over time. Therefore, we undertook a revision of Johnsen and Friborg (2015), with the goal of examining whether year still remained a significant moderator of depression outcomes after (a) addressing potential methodological flaws and disputable choices, (b) including missing eligible trials, and (c) conducting more extensive and reliable moderator analysis with more multiple predictors.

We identified a series of errors and inconsistencies in the selection of the studies. Four included studies did not satisfy the authors’ own inclusion criteria, and one was a secondary analysis, for which the authors did not include the primary trial. Of course, meta-analyses are difficult to conduct, requiring a lot of effort and resources, and small errors are inevitable and many times inconsequential. However, what appears more worrisome is the fact that we identified *thirty* additional trials that were missed in Johnsen and Friborg (2015), despite the fact they satisfied the authors’ inclusion criteria. This is more than half of the number of RCTs (52) they did include and represents a problem with the potential to alter the results. We used the same inclusion criteria so as to be able to judge our findings against those of Johnsen and Friborg (2015). However, some of these criteria might be considered as unnecessarily strict. For instance, restricting studies to only individual, face-to-face CBT is unsupported, given that other formats of delivery (e.g., guided self-help) were shown to be as effective (Barth et al., 2013; Cuijpers, Donker, van Straten, Li, & Andersson, 2010). Also, the threshold of a pretreatment BDI score of 13.5 is completely arbitrary and unnecessary, particularly since only studies on participants with depressive disorders were in-

Table 2

Single and Multiple Meta-Regression Analyses With All Moderators of Interest, Within-Group Effect Sizes, Studies Included in Johnsen & Friberg (2015) Only ("Completer" Data Preferred Over ITT)

| Outcome measure | Single | | | Multiple: Parsimonious | | | Multiple: Full model | | |
|------------------------------|--------|-------------|-----------------------|------------------------|-----------|-----------------------|----------------------|-------------|-----------------------|
| | Coeff | 95% CI | <i>p</i> ^a | Coeff | 95% CI | <i>p</i> ^a | Coeff | 95% CI | <i>p</i> ^a |
| HRSD + BDI-I + BDI-II | | | | | | | | | |
| Year | -.03 | -.05--.01 | .002 | -.03 | -.05--.01 | .001 | -.02 | -.05--.01 | .12 |
| N | -.00 | -.02--.01 | .58 | | | | .00 | -.01--.02 | .47 |
| RoB | -.13 | -.33--.06 | .17 | | | | -.11 | -.38--.17 | .44 |
| WL/CAU/PLA present | -.00 | -.46--.46 | .99 | | | | -.11 | -.66--.44 | .68 |
| Recr from clinical samples | -.08 | -.58--.42 | .74 | | | | .03 | -.54--.60 | .92 |
| MDD vs other diagnosis | .09 | -.38--.56 | .69 | | | | -.06 | -.60--.48 | .82 |
| Adults vs specific group | .28 | -.21--.77 | .26 | | | | .24 | -.26--.74 | .33 |
| Number of sessions | .04 | -.01--.09 | .12 | | | | -.02 | -.10--.06 | .58 |
| Country (U.S. vs other) | .56 | .15--.98 | .009 | .39 | -.01--.79 | .05 | .59 | -.09--1.26 | .09 |
| Beck's manual vs other | .18 | -.34--.70 | .49 | | | | -.00 | -.64--.64 | .99 |
| Inpatients vs outpatients | -.58 | -2.06--.89 | .43 | | | | -.34 | -2.06--1.37 | .69 |
| BDI (I + II) | | | | | | | | | |
| Year | -.03 | -.06--.01 | .002 | -.03 | -.05--.01 | .008 | -.03 | -.06--.00 | .06 |
| N | -.01 | -.02--.01 | .27 | | | | .00 | -.01--.02 | .80 |
| RoB | -.14 | -.35--.08 | .22 | | | | -.05 | -.37--.27 | .78 |
| WL/CAU/PLA present | -.02 | -.53--.49 | .93 | | | | -.06 | -.67--.55 | .85 |
| Recr from clinical samples | -.13 | -.70--.44 | .64 | | | | -.16 | -.84--.54 | .67 |
| MDD vs other diagnosis | .17 | -.36--.69 | .53 | | | | .14 | -.48--.76 | .65 |
| Adults vs specific group | .38 | -.17--.93 | .17 | | | | .41 | -.16--.99 | .15 |
| Number of sessions | .03 | -.02--.09 | .20 | | | | -.03 | -.12--.05 | .43 |
| Country (U.S. vs other) | .61 | .14--1.07 | .01 | .43 | -.00--.86 | .05 | .41 | -.36--1.18 | .29 |
| Beck's manual vs other | .35 | -.23--.93 | .23 | | | | .31 | -.41--1.02 | .39 |
| Inpatients vs outpatients | -.59 | -2.21--1.02 | .46 | | | | -.27 | -2.13--1.58 | .77 |
| HRSD (all versions) | | | | | | | | | |
| Year | -.03 | -.05--.00 | .02 | -.01 | -.04--.01 | .31 | -.01 | -.05--.03 | .74 |
| N | .00 | -.01--.02 | .71 | | | | .00 | -.02--.02 | .85 |
| RoB | -.10 | -.29--.09 | .27 | | | | -.03 | -.37--.31 | .84 |
| WL/CAU/PLA present | -.20 | -.71--.32 | .44 | | | | .06 | -.65--.77 | .86 |
| Recr from clinical samples | .09 | -.48--.66 | .75 | | | | .13 | -.63--.88 | .73 |
| MDD vs other diagnosis | .00 | -.56--.57 | .99 | | | | -.31 | -1.07--.44 | .40 |
| Adults vs specific group | .48 | -.00--.97 | .05 | .38 | -.07--.82 | .09 | .42 | -.16--1.00 | .15 |
| Number of sessions | .06 | .02--.12 | .01 | .05 | -.01--.10 | .09 | .06 | -.06--.18 | .29 |
| Country (U.S. vs other) | .36 | -.16--.89 | .17 | | | | .16 | -.74--1.05 | .72 |
| Beck's manual vs other | -.05 | -.73--.63 | .88 | | | | -.25 | -1.21--.70 | .58 |
| Inpatients vs outpatients | -.64 | -2.01--.73 | .35 | | | | -.18 | -2.05--1.70 | .85 |

Note. ITT = intent-to-treat; BDI = Beck Depression Inventory; HRSD = Hamilton Rating Scale for Depression; N = number of participants randomized in the CBT arm; RoB = risk of bias; WL = waitlist control; CAU = care-as-usual; PLA = placebo; Recr = recruitment; MDD = major depressive disorder.

^a The *p* levels in these columns indicate whether the relationship between year and effect size is significant in meta-regression analyses (significant results are in bold).

cluded. On the other hand, studies on inpatients have consistently been shown to be a distinct category of trials, subject to a very diverse set of constraints (e.g., participants are often not free to leave the trial), so their combination with outpatient studies is problematic.

We also made a number of methodological choices divergent from Johnsen and Friberg (2015). We restricted our analysis to RCTs, imposed no language restrictions, and calculated both within (pre-to-post) group and between-group ESs. Moreover, all within ESs were calculated with the same procedure, regardless of whether the CBT arm came from a trial with an "inactive" or active (e.g., other psychotherapy) control group. To aid comparison, we conducted two sets of analyses. One included only the RCTs contained in Johnsen and Friberg (2015), and focused on within-group effects sizes, like in their main analyses, while adding the multiple meta-regression results with other predictors of

interest. The second is our version of the most improved analysis, in which we modified all the methodological choices we considered questionable (i.e., additional studies included, between-group ESs, and ITT data preferred over "completer").

The result from our first analysis showed large prepost ESs for CBT on both the BDI and the HRSD, with values very similar to those reported by Johnsen and Friberg (2015). The most important result was the extremely high level of heterogeneity, raising serious doubts about the reliability of the results. This was further corroborated by the fact that a startling *one fourth* (in analysis favoring "completers") to *one third* (in analysis favoring ITT) of the included studies were outliers (i.e., resulted in ESs with 95% CI outside the 95% CI of the pooled ES). However, in our second, improved meta-analysis, estimations of the effects of CBT compared to a control group were, as expected, smaller, but accompanied by only moderate heterogeneity, and with a small number of

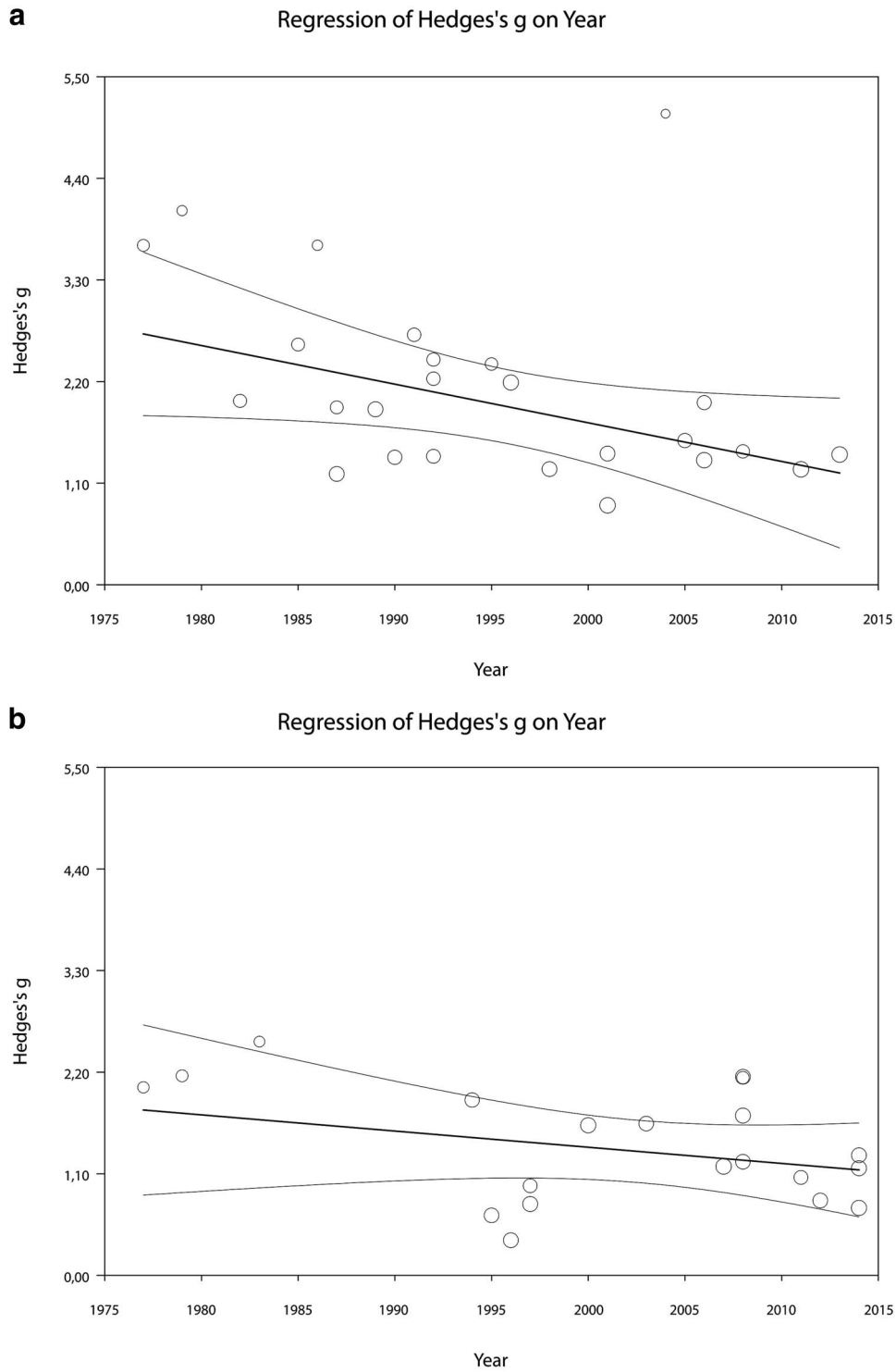


Figure 1. Meta-regression analyses of the effects of year of publication on within-group effect sizes in analyses restricted to RCTs included in Johnsen and Friberg (2015) for the BDI for (a) studies in the U.S., (b) studies outside the US.

BDI (between groups effects)

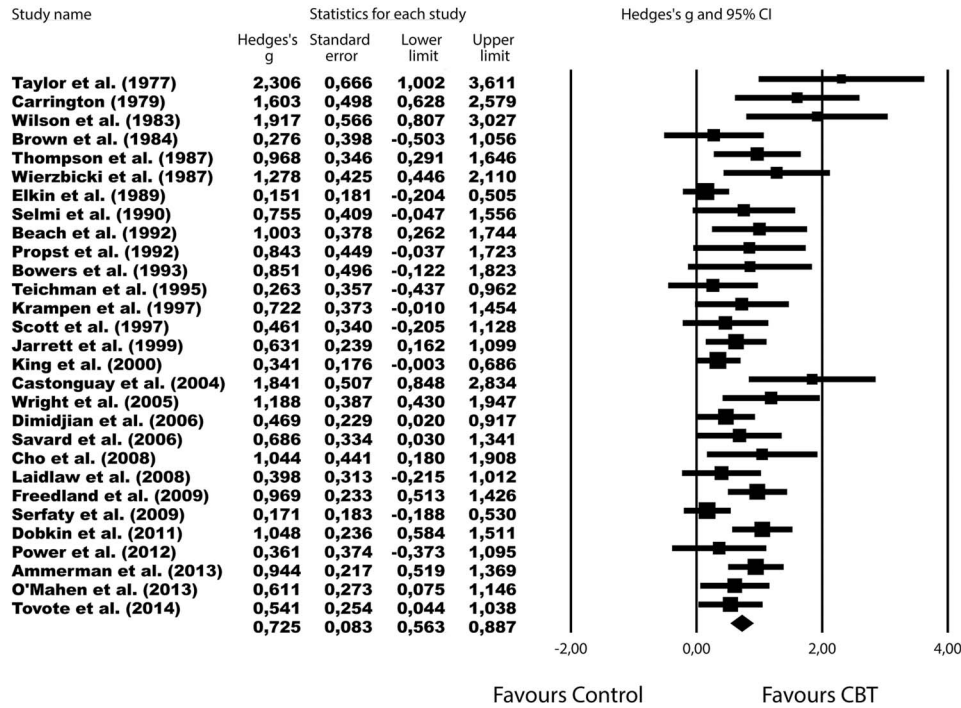


Figure 2. Standardized between-group effect sizes of CBT on the Beck Depression Inventory in new meta-analysis including all eligible trials.

outliers, in line with previous meta-analyses of CBT compared to a control group (Barth et al., 2013; Cuijpers et al., 2013). Given that the collection of studies is similar to, in fact even narrower than, that of previous meta-analyses, we conjecture that heteroge-

neity might have to do with the nature of prepost ESs. The idea that one could be able to gauge the “pure” effect of CBT by combining the CBT arms from different trials is in itself problematic. Between-group ESs have the advantage of being calculated *em-*

HRSD (between groups effects)

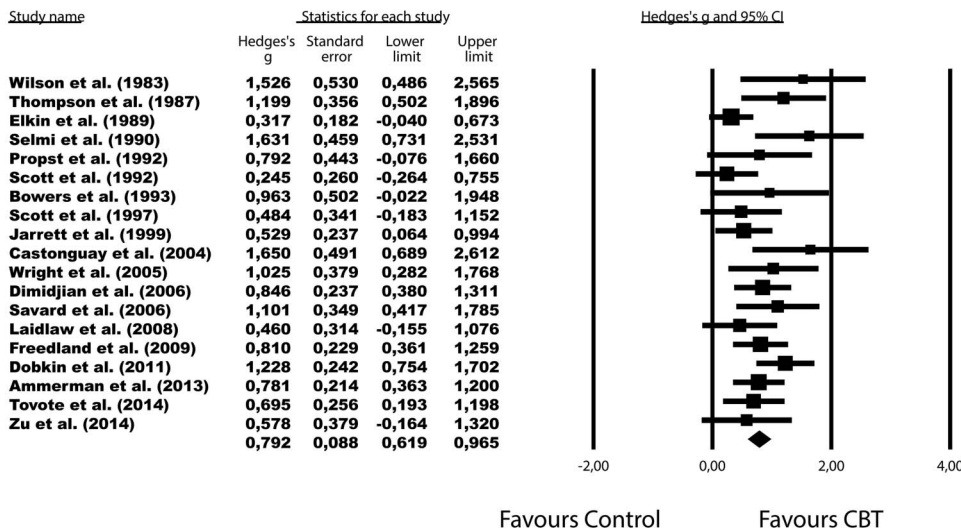


Figure 3. Standardized between-group effect sizes of CBT on Hamilton Rating Scale for Depression in new meta-analysis including all eligible trials.

Table 3
Single Meta-Regression Analyses With Year of Publication For
Between-Group ESs, All Controlled Studies (ITT Data Preferred
Over “Completer”)

| Outcome measure | Coeff | 95% CI | <i>p</i> ^a |
|--------------------------------|-------|----------|-----------------------|
| BDI (I + II) | | | |
| All studies | -.01 | -.03-.01 | .22 |
| Inpatients excluded | -.01 | -.03-.01 | .23 |
| Outliers excluded | -.01 | -.02-.01 | .38 |
| Inpatients & outliers excluded | -.01 | -.02-.01 | .39 |
| Only low RoB studies (>2) | .03 | -.02-.07 | .25 |
| BDI-I only | -.00 | -.03-.02 | .59 |
| BDI-II only | .01 | -.09-.10 | .85 |
| HRSD (all versions) | | | |
| All studies ^b | .00 | -.02-.02 | .77 |
| Inpatients excluded | .00 | -.02-.02 | .75 |
| Only low RoB studies (>2) | .03 | -.03-.08 | .28 |
| Only HRSD-17 | .00 | -.02-.02 | .89 |

Note. BDI = Beck Depression Inventory; HRSD = Hamilton Rating Scale for Depression; ITT = intent-to-treat; RoB = risk of bias.

^a The *p* levels in this column indicate whether the relationship between year and effect size is significant in meta-regression analyses (no results were significant). ^b There were no outliers for this analysis.

bedded in the context of the trial in which participants were randomized. As such, they also include the important assumption that the intervention and the control group are not completely, if at all, independent from each other. Within-group ESs, such as the ones used by Johnsen and Friberg (2015), do not include this non-independence assumption because they effectively detach the intervention group from the context of the trial. Furthermore, between-group effect sizes adjust for differences in target populations, interventions, and study design, because both the treatment and control group have these same characteristics. Within-group effect sizes, on the other hand, do not adjust for these differences in any way. For example, length of treatment is highly important because of natural recovery, which is considerable in depression. In within-group effect sizes, length of treatment can be expected to be an important predictor of the effect size, whereas this is not the case in between-group effect sizes, because both treatment and control group have the same length. Finally, by leveling differences due to methodological quality and the context of the trial, meta-analyses of within-group ESs run the risk of artificially inflating the apparent efficiency on an intervention, ascribing to it a variety of extraneous influences.

Regarding the crux of our revision—the moderating value of year of publication on depression outcomes—our results contradict those of Johnsen and Friberg (2015). In the analyses similar to their original, which used within-group ESs, year was a significant moderator of outcome in single analyses for the BDI, with the exception of studies reporting outcomes only on the BDI-II and of studies with low risk of bias. Publication year was *not*, however, a significant moderator when outcomes were measured on the HRSD, with the exception of the analysis favoring “completer” over ITT data. But even in this last case, excluding outliers or considering only studies with low RoB rendered the relationship non-significant. Single meta-regression also identified another moderator: country where the trial was conducted. Most importantly, nonetheless, multiple meta-regression analyses on the BDI

and HRSD data collapsed together, as well as on the BDI taken separately, evidenced one of two phenomena: *either* year stopped being a significant moderator of outcome (full model) *or both* year and country were significant (parsimonious model). When outcomes were restricted to the HRSD, no predictor (year included) remained significant in multiple meta-regression analyses.

In our subsequent improved meta-analysis, where we included additional, eligible but missed, studies, favored ITT over “completer” data, and computed effects sizes in comparison to a control group (*between-group ESs*), results were even more trenchant. Year of publication was not a predictor of outcome in *any* of the single or multiple meta-regression analyses, regardless of the outcome measure. Indeed, the only significant predictors that consistently emerged were the number of participants in the CBT arm and recruitment from clinical samples, which were both negatively related to outcome. However, given that we used the same inclusion criteria as the original meta-analysis, we could calculate between-group ESs for only 31 trials (the rest of the trials compared CBT exclusively with another psychotherapy or pharmacotherapy). Since Johnsen and Friberg (2015) reported a similar analysis on a larger number of trials, one might argue we had smaller statistical power. However, we should note that level of heterogeneity were just moderate for our analysis on between-group ESs, a factor increasing statistical power.

As an exploratory examination, we attempted to disentangle the link between year of publication and country for within-group ESs by looking at the association between year and depression outcomes separately, for studies published in the U.S. and for studies published in other countries. Our results also indicated that although the significant decrease of ESs over time was present in studies conducted in the U.S., where earlier studies resulted in much higher ESs and subsequently moved toward effect sizes similar to the rest of the world, no significant decrease was found in non-U.S. studies. We note that this finding might, too, be an artifact, given the high levels of heterogeneity accompanying analyses on within-group ESs. Interestingly but perhaps coincidentally, a recent meta-analysis (Tuttle et al., 2015) of drug trials for neuropathic pain has shown a similar geographical pattern in the effectiveness of the placebo arm: ESs for this arm decreased over time in trials conducted in the U.S. However, in that case, study length and sample size were also significant moderators of the within-group ESs, unlike in our analysis, although number of participants in the CBT arm was negatively related to outcome for between-group ESs.

In conclusion, we redid the meta-analysis conducted by Johnsen and Friberg (2015) and uncovered a number of methodological errors. Heterogeneity estimates were extremely high and a large proportion of studies were outliers, effectively precluding the realization of a meta-analysis (Ioannidis et al., 2007). Given that our additional analyses of between-group ESs, as well as other meta-analyses of CBT that included studies with larger and more diverse samples did not come across such high heterogeneity, we surmise the problem may lie with the combination of within-group ESs. Most importantly, we found that the association of year of publication with depression outcomes was unstable and fleeting, present only under certain limited conditions, and completely absent in others. Specifically, this relationship was almost completely absent in the within-group ESs analyses on the HRSD and in *all* analyses that employed between-group ESs. Multiple meta-

Table 4

Single and Multiple Meta-Regression Analyses With All Moderators of Interest for Between-Group ESs, All Controlled Studies (ITT Data Preferred Over “Completer”)

| Outcome measure | Single | | | Multiple: Parsimonious | | | Multiple: Full model | | |
|------------------------------|--------|------------|-----------------------|------------------------|----------|-----------------------|----------------------|------------|-----------------------|
| | Coeff | 95% CI | <i>p</i> ^a | Coeff | 95% CI | <i>p</i> ^a | Coeff | 95% CI | <i>p</i> ^a |
| HRSD + BDI-I + BDI-II | | | | | | | | | |
| Year | -.01 | -.02-.01 | .26 | .00 | -.01-.01 | .93 | .00 | -.02-.02 | .80 |
| N | -.01 | -.02-.01 | .0003 | -.01 | -.02-.00 | .001 | -.01 | -.02-.00 | .01 |
| RoB | -.11 | -.25-.03 | .11 | | | | -.04 | -.23-.15 | .65 |
| Recr from clinical samples | -.43 | -.71-.16 | .003 | -.26 | -.51-.00 | .05 | -.17 | -.54-.19 | .33 |
| MDD vs other diagnosis | -.03 | -.35-.29 | .85 | | | | -.03 | -.32-.26 | .85 |
| Adults vs specific group | -.12 | -.44-.20 | .45 | | | | -.07 | -.46-.32 | .71 |
| Number of sessions | -.02 | -.05-.02 | .39 | | | | -.02 | -.05-.01 | .24 |
| Country (U.S. vs other) | .31 | .02-.61 | .04 | .23 | -.01-.47 | .06 | .36 | .04-.68 | .03 |
| Beck's manual vs other | .05 | -.33-.43 | .79 | | | | -.01 | -.34-.33 | .97 |
| Inpatients vs outpatients | .19 | -.03-1.41 | .75 | | | | -.43 | -1.55-.70 | .44 |
| BDI-I only | | | | | | | | | |
| Year | -.01 | -.03-.02 | .59 | -.00 | -.02-.02 | .89 | -.01 | -.05-.04 | .71 |
| N | -.01 | -.02-.00 | .003 | | | | -.01 | -.03-.01 | .37 |
| RoB | -.09 | -.28-.10 | .34 | | | | .12 | -.30-.53 | .55 |
| Recr from clinical samples | -.66 | -.96-.37 | .0001 | -.67 | -.97-.36 | .0002 | -.70 | -1.91-.50 | .22 |
| MDD vs other diagnosis | -.04 | -.53-.44 | .85 | | | | .29 | -.41-.99 | .37 |
| Adults vs specific group | -.32 | -.79-.15 | .17 | | | | .04 | -.82-.89 | .93 |
| Number of sessions | -.01 | -.06-.04 | .67 | | | | .00 | -.07-.07 | .96 |
| Country (U.S. vs other) | .12 | -.37-.60 | .62 | | | | -.36 | -1.35-.64 | .44 |
| Beck's manual vs other | .22 | -.27-.71 | .36 | | | | .12 | -.51-.76 | .67 |
| Inpatients vs outpatients | .03 | -1.30-1.37 | .96 | | | | -.50 | -2.13-1.12 | .51 |
| HRSD only | | | | | | | | | |
| Year | .00 | -.02-.02 | .77 | -.01 | -.02-.01 | .53 | -.00 | -.03-.03 | .99 |
| N | -.01 | -.02-.00 | .05 | -.01 | -.02-.00 | .09 | -.01 | -.02-.00 | .08 |
| RoB | -.03 | -.21-.15 | .70 | | | | -.04 | -.31-.23 | .72 |
| Recr from clinical samples | -.53 | -.81-.24 | .001 | -.52 | -.85-.19 | .004 | -.39 | -.97-.20 | .16 |
| MDD vs other diagnosis | -.37 | -.72-.02 | .04 | | | | -.01 | -.57-.56 | .97 |
| Adults vs specific group | -.19 | -.56-.18 | .29 | | | | .10 | -.42-.62 | .67 |
| Number of sessions | -.03 | -.07-.01 | .19 | | | | -.03 | -.09-.03 | .30 |
| Country (U.S. vs other) | .23 | -.16-.62 | .23 | | | | .40 | -.12-.92 | .12 |
| Beck's manual vs other | -.15 | -.86-.56 | .66 | | | | .16 | -.74-1.06 | .69 |
| Inpatients vs outpatients | .18 | -1.03-1.38 | .76 | | | | -.53 | -1.97-.91 | .42 |

Note. ITT = intent-to-treat; BDI = Beck Depression Inventory; HRSD = Hamilton Rating Scale for Depression; N = number of participants randomized in the CBT arm; RoB = risk of bias; WL = waitlist control; CAU = care-as-usual; PLA = placebo; Recr = recruitment; MDD = major depressive disorder.

^a The *p* levels in these columns indicate whether the relationship between year and effect size is significant in meta-regression analyses (significant results are in bold).

regression analyses did *not* support the systematic decline in the efficiency of CBT over time, with the exception of studies conducted in the U.S. However, these results should not be construed as implying that the effects of CBT have not declined from the initial studies showing very large ESs. In the case of within-group ESs, roughly a third of the studies were outliers and many of these were early studies (Beck, Hollon, Young, Bedrosian, & Budenz, 1985; Carrington, 1979; McNamara & Horan, 1986; Rush, Beck, Kovacs, & Hollon, 1977), so in a sense there was nowhere to go but down. This is a not-uncommon phenomenon when testing a new therapy, with the first studies showing large magnitude effects (also called the “winner’s curse”). It is worth noting that some more recent studies also found large ESs of CBT (Castonguay et al., 2004; David, Szentagotai, Lupu, & Cosman, 2008; Quilty, Dozois, Lobo, Ravindran, & Bagby, 2014).

Although it would be impossible to exclude with absolute certainty the possibility that year of publication *might* be systematically related to the effects of CBT for depression, for some outcomes in some circumstances, this is not very probable. More

likely, the linear temporal trend (“fall”) reported by Johnsen and Friborg (2015) is simply a spurious finding.

References

References marked with an asterisk indicate studies included in the meta-analysis.

- *Ammerman, R. T., Putnam, F. W., Altaye, M., Stevens, J., Teeters, A. R., & Van Ginkel, J. B. (2013). A clinical trial of in-home CBT for depressed mothers in home visitation. *Behavior Therapy, 44*, 359–372. <http://dx.doi.org/10.1016/j.beth.2013.01.002>
- Armijo-Olivo, S., Fuentes, J., Ospina, M., Saltaji, H., & Hartling, L. (2013). Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: A descriptive analysis. *BMC Medical Research Methodology, 13*, 116. <http://dx.doi.org/10.1186/1471-2288-13-116>
- *Ashman, T., Cantor, J. B., Tsaousides, T., Spielman, L., & Gordon, W. (2014). Comparison of cognitive behavioral therapy and supportive psychotherapy for the treatment of depression following traumatic

- brain injury: A randomized controlled trial. *The Journal of Head Trauma Rehabilitation*, 29, 467–478. <http://dx.doi.org/10.1097/HTR.000000000000098>
- *Banken, D. M. (1993). *Group versus individual cognitive-behavioral treatment for depression*. Pullman, Washington: Washington State University.
- Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., . . . Cuijpers, P. (2013). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: A network meta-analysis. *PLoS Medicine*, 10, e1001454. <http://dx.doi.org/10.1371/journal.pmed.1001454>
- *Beach, S. R. H., & O'Leary, K. D. (1992). Treating depression in the context of marital discord: Outcome and predictors of response of marital therapy versus cognitive therapy. *Behavior Therapy*, 23, 507–528. [http://dx.doi.org/10.1016/S0005-7894\(05\)80219-9](http://dx.doi.org/10.1016/S0005-7894(05)80219-9)
- *Beck, A. T., Hollon, S. D., Young, J. E., Bedrosian, R. C., & Budenz, D. (1985). Treatment of depression with cognitive therapy and amitriptyline. *Archives of General Psychiatry*, 42, 142–148. <http://dx.doi.org/10.1001/archpsyc.1985.01790250036005>
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II, Beck Depression Inventory: Manual*. San Antonio, TX: Psychological Corporation.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561–571. <http://dx.doi.org/10.1001/archpsyc.1961.01710120031004>
- *Beutler, L. E., Moleiro, C., Malik, M., Harwood, M. T., Romanelli, R., Gallagher-Thompson, D., & Thompson, L. (2003). A comparison of the Dodo, EST, and ATI factors among comorbid stimulant-dependent, depressed patients. *Clinical Psychology & Psychotherapy*, 10, 69–85. <http://dx.doi.org/10.1002/cpp.354>
- *Blackburn, I. M., & Moore, R. G. (1997). Controlled acute and follow-up trial of cognitive therapy and pharmacotherapy in out-patients with recurrent depression. *The British Journal of Psychiatry*, 171, 328–334. <http://dx.doi.org/10.1192/bjp.171.4.328>
- *Bodenmann, G., Plancherel, B., Beach, S. R. H., Widmer, K., Gabriel, B., Meuwly, N., . . . Schramm, E. (2008). Effects of coping-oriented couples therapy on depression: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 76, 944–954. <http://dx.doi.org/10.1037/a0013467>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, England: Wiley. <http://dx.doi.org/10.1002/9780470743386>
- *Bowers, W. A. (1990). Treatment of depressed in-patients. Cognitive therapy plus medication, relaxation plus medication, and medication alone. *The British Journal of Psychiatry*, 156, 73–78. <http://dx.doi.org/10.1192/bjp.156.1.73>
- *Bowers, W., Stuart, S., Macfarlane, R., & Gorman, L. (1993). Use of computer-administered cognitive-behavior therapy with depressed inpatients. *Depression*, 1, 294–299. <http://dx.doi.org/10.1002/depr.3050010603>
- Brown, R. A., Evans, D. M., Miller, I. W., Burgess, E. S., & Mueller, T. I. (1997). Cognitive-behavioral treatment for depression in alcoholism. *Journal of Consulting and Clinical Psychology*, 65, 715–726. <http://dx.doi.org/10.1037/0022-006X.65.5.715>
- *Brown, R. A., & Lewinsohn, P. M. (1984). A psychoeducational approach to the treatment of depression: Comparison of group, individual, and minimal contact procedures. *Journal of Consulting and Clinical Psychology*, 52, 774–783. <http://dx.doi.org/10.1037/0022-006X.52.5.774>
- Burkeman, O. (2015, July 3). Why CBT is falling out of favour. *The Guardian*. Retrieved from <http://www.theguardian.com/lifeandstyle/2015/jul/03/why-cbt-is-falling-out-of-favour-oliver-burkeman>
- *Carrington, C. H. (1979). A comparison of cognitive and analytically oriented brief treatment approaches to depression in black women. *Dissertation Abstracts International*, 40(6B). (UMI No. 7926513)
- *Castonguay, L. G., Schut, A. J., Aikens, D. E., Constantino, M. J., Laurenceau, J.-P., Bologh, L., & Burns, D. D. (2004). Integrative cognitive therapy for depression: A preliminary investigation. *Journal of Psychotherapy Integration*, 14, 4–20.
- Chen, P., Furukawa, T. A., Shinohara, K., Honyashiki, M., Imai, H., Ichikawa, K., . . . Churchill, R. (2014). Quantity and quality of psychotherapy trials for depression in the past five decades. *Journal of Affective Disorders*, 165, 190–195. <http://dx.doi.org/10.1016/j.jad.2014.04.071>
- *Cho, H. J., Kwon, J. H., & Lee, J. J. (2008). Antenatal cognitive-behavioral therapy for prevention of postpartum depression: A pilot study. *Yonsei Medical Journal*, 49, 553–562. <http://dx.doi.org/10.3349/ymj.2008.49.4.553>
- *Constantino, M. J., Marnell, M. E., Haile, A. J., Kanther-Sista, S. N., Wolman, K., Zappert, L., & Arnow, B. A. (2008). Integrative cognitive therapy for depression: A randomized pilot comparison. *Psychotherapy*, 45, 122–134. <http://dx.doi.org/10.1037/0033-3204.45.2.122>
- Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *The Canadian Journal of Psychiatry/La Revue canadienne de psychiatrie*, 58, 376–385.
- Cuijpers, P., Donker, T., van Straten, A., Li, J., & Andersson, G. (2010). Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychological Medicine*, 40, 1943–1957. <http://dx.doi.org/10.1017/S0033291710000772>
- Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are over-estimated: A meta-analysis of study quality and effect size. *Psychological Medicine*, 40, 211–223. <http://dx.doi.org/10.1017/S0033291709006114>
- Cuijpers, P., van Straten, A., Warmerdam, L., & Andersson, G. (2008). Psychological treatment of depression: A meta-analytic database of randomized studies. *BMC Psychiatry*, 8, 36. <http://dx.doi.org/10.1186/1471-244X-8-36>
- Cuijpers, P., Weitz, E., Twisk, J., Kuehner, C., Cristea, I., David, D., . . . Hollon, S. D. (2014). Gender as predictor and moderator of outcome in cognitive behavior therapy and pharmacotherapy for adult depression: An “individual patient data” meta-analysis. *Depression and Anxiety*, 31, 941–951. <http://dx.doi.org/10.1002/da.22328>
- *David, D., Szentagotai, A., Lupu, V., & Cosman, D. (2008). Rational emotive behavior therapy, cognitive therapy, and medication in the treatment of major depressive disorder: A randomized clinical trial, posttreatment outcomes, and six-month follow-up. *Journal of Clinical Psychology*, 64, 728–746. <http://dx.doi.org/10.1002/jclp.20487>
- *de Jong, R., Treiber, R., & Henrich, G. (1986). Effectiveness of two psychological treatments for inpatients with severe and chronic depressions. *Cognitive Therapy and Research*, 10, 645–663. <http://dx.doi.org/10.1007/BF01173752>
- *de Jong-Meyer, R., & de Hautzinger, M. (1996). Results of two multi-center treatment studies among patients with endogenous and nonendogenous depression: Conclusions and prospects. *Zeitschrift für Klinische Psychologie*, 25, 155–160.
- *Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmalzing, K. B., Kohlenberg, R. J., Addis, M. E., . . . Jacobson, N. S. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology*, 74, 658–670. <http://dx.doi.org/10.1037/0022-006X.74.4.658>
- *Dobkin, R. D., Menza, M., Allen, L. A., Gara, M. A., Mark, M. H., Tiu, J., . . . Friedman, J. (2011). Cognitive-behavioral therapy for depression in Parkinson's disease: A randomized, controlled trial. *The American Journal of Psychiatry*, 168, 1066–1074. <http://dx.doi.org/10.1176/appi.ajp.2011.10111669>

- *Dozois, D. J. A., Bieling, P. J., Patelis-Siotis, I., Hoar, L., Chudzik, S., McCabe, K., & Westra, H. A. (2009). Changes in self-schema structure in cognitive therapy for major depressive disorder: A randomized clinical trial. *Journal of Consulting and Clinical Psychology, 77*, 1078–1088. <http://dx.doi.org/10.1037/a0016886>
- Driessen, E., Cuijpers, P., Hollon, S. D., & Dekker, J. J. M. (2010). Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. *Journal of Consulting and Clinical Psychology, 78*, 668–680. <http://dx.doi.org/10.1037/a0020570>
- *Dunn, R. J. (1979). Cognitive modification with depression-prone psychiatric patients. *Cognitive Therapy and Research, 3*, 307–317. <http://dx.doi.org/10.1007/BF01185971>
- *Dunner, D. L., Schmalting, K. B., Hendrickson, H., Becker, J., Lehman, A., & Bea, C. (1996). Cognitive therapy versus fluoxetine in the treatment of dysthymic disorder. *Depression, 4*, 34–41. [http://dx.doi.org/10.1002/\(SICI\)1522-7162\(1996\)4:1<34::AID-DEPR4>3.0.CO;2-F](http://dx.doi.org/10.1002/(SICI)1522-7162(1996)4:1<34::AID-DEPR4>3.0.CO;2-F)
- *Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., . . . Parloff, M. B. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry, 46*, 971–982. <http://dx.doi.org/10.1001/archpsyc.1989.01810110013002>
- *Emanuels-Zuurveen, L., & Emmelkamp, P. M. (1996). Individual behavioural-cognitive therapy v. marital therapy for depression in maritally distressed couples. *The British Journal of Psychiatry, 169*, 181–188. <http://dx.doi.org/10.1192/bjp.169.2.181>
- *Emanuels-Zuurveen, L., & Emmelkamp, P. M. (1997). Spouse-aided therapy with depressed patients. *Behavior Modification, 21*, 62–77. <http://dx.doi.org/10.1177/01454455970211003>
- *Floyd, M. D. (1998). *Cognitive therapy for depression: A comparison of individual psychotherapy and bibliotherapy for depressed older adults*. Birmingham, AL: The University of Alabama.
- Forman, E. M., Herbert, J. D., Moitra, E., Yeomans, P. D., & Geller, P. A. (2007). A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. *Behavior Modification, 31*, 772–799. <http://dx.doi.org/10.1177/0145445507302202>
- *Freedland, K. E., Skala, J. A., Carney, R. M., Rubin, E. H., Lustman, P. J., Dávila-Román, V. G., . . . Hogue, C. W., Jr. (2009). Treatment of depression after coronary artery bypass surgery: A randomized controlled trial. *Archives of General Psychiatry, 66*, 387–396. <http://dx.doi.org/10.1001/archgenpsychiatry.2009.7>
- Furukawa, T. A., Noma, H., Caldwell, D. M., Honyashiki, M., Shinohara, K., Imai, H., . . . Churchill, R. (2014). Waiting list may be a nocebo condition in psychotherapy trials: A contribution from network meta-analysis. *Acta Psychiatrica Scandinavica, 130*, 181–192. <http://dx.doi.org/10.1111/acps.12275>
- *Gallagher, D. E., & Thompson, L. W. (1982). Treatment of major depressive disorder in older adult outpatients with brief psychotherapies. *Psychotherapy: Theory, Research, & Practice, 19*, 482–490. <http://dx.doi.org/10.1037/h0088461>
- Gallagher-Thompson, D., & Steffen, A. M. (1994). Comparative effects of cognitive-behavioral and brief psychodynamic psychotherapies for depressed family caregivers. *Journal of Consulting and Clinical Psychology, 62*, 543–549. <http://dx.doi.org/10.1037/0022-006X.62.3.543>
- Graham, S. (2015, July 10). Has cognitive behavioral therapy had its day as a treatment for depression in the UK? *Vice*. Retrieved from <http://www.vice.com/read/has-cbt-had-its-day-as-a-treatment-for-depression>
- Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research, 2*, 109–112. <http://dx.doi.org/10.4103/2229-3485.83221>
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, & Psychiatry, 23*, 56–62. <http://dx.doi.org/10.1136/jnnp.23.1.56>
- *Hautzinger, M., Jong-Meyer, R. D., Treiber, R., Rudolf, G. A. E., & Thien, U. (1996). Wirksamkeit Kognitiver Verhaltenstherapie, Pharmaco-therapie und deren Kombination bei nicht-endogenen, unipolaren Depression [Efficacy of CBT, pharmacotherapy, and their combination for non-endogenous, unipolar depression]. *Zeitschrift für Klinische Psychologie, 25*, 130–145.
- *Hayden, T., Perantie, D. C., Nix, B. D., Barnes, L. D., Mostello, D. J., Holcomb, W. L., . . . Hershey, T. (2012). Treating prepartum depression to improve infant developmental outcomes: A study of diabetes in pregnancy. *Journal of Clinical Psychology in Medical Settings, 19*, 285–292. <http://dx.doi.org/10.1007/s10880-011-9294-8>
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Higgins, J. P. T., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., . . . Cochrane Statistical Methods Group. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal, 343*, d5928–d5928. <http://dx.doi.org/10.1136/bmj.d5928>
- Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for Systematic Reviews of Interventions Version 5.1.0* (updated March 2011). The Cochrane Collaboration. Retrieved from www.cochrane-handbook.org
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *The BMJ, 327*, 557–560. <http://dx.doi.org/10.1136/bmj.327.7414.557>
- *Hollon, S. D., DeRubeis, R. J., Evans, M. D., Wiemer, M. J., Garvey, M. J., Grove, W. M., & Tuason, V. B. (1992). Cognitive therapy and pharmacotherapy for depression. Singly and in combination. *Archives of General Psychiatry, 49*, 774–781. <http://dx.doi.org/10.1001/archpsyc.1992.01820100018004>
- Ioannidis, J. P. A., Patsopoulos, N. A., & Evangelou, E. (2007). Uncertainty in heterogeneity estimates in meta-analyses. *BMJ, 335*, 914–916. <http://dx.doi.org/10.1136/bmj.39343.408449.80>
- *Jacobson, N. S., Dobson, K., Fruzzetti, A. E., Schmalting, K. B., & Salusky, S. (1991). Marital therapy as a treatment for depression. *Journal of Consulting and Clinical Psychology, 59*, 547–557. <http://dx.doi.org/10.1037/0022-006X.59.4.547>
- *Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., . . . Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology, 64*, 295–304. <http://dx.doi.org/10.1037/0022-006X.64.2.295>
- *Jarrett, R. B., Schaffer, M., McIntire, D., Witt-Browder, A., Kraft, D., & Risser, R. C. (1999). Treatment of atypical depression with cognitive therapy or phenelzine: A double-blind, placebo-controlled trial. *Archives of General Psychiatry, 56*, 431–437. <http://dx.doi.org/10.1001/archpsyc.56.5.431>
- 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*, 747–768. <http://dx.doi.org/10.1037/bul0000015>
- Kalapapap, R. K., Ho, J., Cai, X., Vinogradov, S., Batki, S. L., & Mohr, D. C. (2014). Cognitive-behavioral therapy in depressed primary care patients with co-occurring problematic alcohol use: Effect of telephone-administered vs. face-to-face treatment—A secondary analysis. *Journal of Psychoactive Drugs, 46*, 85–92. <http://dx.doi.org/10.1080/02791072.2013.876521>
- *Kennedy, S. H., Konarski, J. Z., Segal, Z. V., Lau, M. A., Bieling, P. J., McIntyre, R. S., & Mayberg, H. S. (2007). Differences in brain glucose metabolism between responders to CBT and venlafaxine in a 16-week randomized controlled trial. *The American Journal of Psychiatry, 164*, 778–788. <http://dx.doi.org/10.1176/ajp.2007.164.5.778>
- *King, M., Sibbald, B., Ward, E., Bower, P., Lloyd, M., Gabbay, M., & Byford, S. (2000). Randomised controlled trial of non-directive coun-

- selling, cognitive-behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care. *Health Technology Assessment*, 4, 1–83.
- Kocsis, J. H., Gerber, A. J., Milrod, B., Roose, S. P., Barber, J., Thase, M. E., . . . Leon, A. C. (2010). A new scale for assessing the quality of randomized clinical trials of psychotherapy. *Comprehensive Psychiatry*, 51, 319–324. <http://dx.doi.org/10.1016/j.comppsy.2009.07.001>
- *Krampen, G. (1997). Autogenes Training vor und begleitend zur metho-
denübergreifenden Einzelpsychotherapie bei depressiven Störungen. *Zeitschrift für Klinische Psychologie [Autogenic training before and in addition to integrative psychotherapy of depressive disorders]*. *Psychiatrie und Psychotherapie*, 45, 214–232.
- *Laidlaw, K., Davidson, K., Toner, H., Jackson, G., Clark, S., Law, J., . . . Cross, S. (2008). A randomised controlled trial of cognitive behaviour therapy vs treatment as usual in the treatment of mild to moderate late life depression. *International Journal of Geriatric Psychiatry*, 23, 843–850. <http://dx.doi.org/10.1002/gps.1993>
- Liberman, R. P., & Eckman, T. (1981). Behavior therapy vs insight-oriented therapy for repeated suicide attempters. *Archives of General Psychiatry*, 38, 1126–1130. <http://dx.doi.org/10.1001/archpsyc.1981.01780350060007>
- *Lopes, R. T., Gonçalves, M. M., Machado, P. P. P., Sinai, D., Bento, T., & Salgado, J. (2014). Narrative therapy vs. cognitive-behavioral therapy for moderate depression: Empirical evidence from a controlled clinical trial. *Psychotherapy Research*, 24, 662–674. <http://dx.doi.org/10.1080/10503307.2013.874052>
- *Luty, S. E., Carter, J. D., McKenzie, J. M., Rae, A. M., Frampton, C. M. A., Mulder, R. T., & Joyce, P. R. (2007). Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *The British Journal of Psychiatry*, 190, 496–502. <http://dx.doi.org/10.1192/bjp.bp.106.024729>
- *Macaskill, D. N. D., & Macaskill, A. (1996). Rational-emotive therapy plus pharmacotherapy versus pharmacotherapy alone in the treatment of high cognitive dysfunction depression. *Cognitive Therapy and Research*, 20, 575–592. <http://dx.doi.org/10.1007/BF02227962>
- *Markowitz, J. C., Kocsis, J. H., Fishman, B., Spielman, L. A., Jacobsberg, L. B., Frances, A. J., . . . Perry, S. W. (1998). Treatment of depressive symptoms in human immunodeficiency virus-positive patients. *Archives of General Psychiatry*, 55, 452–457. <http://dx.doi.org/10.1001/archpsyc.55.5.452>
- McBride, C., Atkinson, L., Quilty, L. C., & Bagby, R. M. (2006). Attachment as moderator of treatment outcome in major depression: A randomized control trial of interpersonal psychotherapy versus cognitive behavior therapy. *Journal of Consulting and Clinical Psychology*, 74, 1041–1054. <http://dx.doi.org/10.1037/0022-006X.74.6.1041>
- McLean, P. D., & Hakstian, A. R. (1979). Clinical depression: Comparative efficacy of outpatient treatments. *Journal of Consulting and Clinical Psychology*, 47, 818–836. <http://dx.doi.org/10.1037/0022-006X.47.5.818>
- *McNamara, K., & Horan, J. J. (1986). Experimental construct validity in the evaluation of cognitive and behavioral treatments for depression. *Journal of Counseling Psychology*, 33, 23–30. <http://dx.doi.org/10.1037/0022-0167.33.1.23>
- *Misri, S., Reebye, P., Milis, L., & Shah, S. (2006). The impact of treatment intervention on parenting stress in postpartum depressed mothers: A prospective study. *American Journal of Orthopsychiatry*, 76, 115–119. <http://dx.doi.org/10.1037/0002-9432.76.1.115>
- *Mohr, D. C., Boudewyn, A. C., Goodkin, D. E., Bostrom, A., & Epstein, L. (2001). Comparative outcomes for individual cognitive-behavior therapy, supportive-expressive group psychotherapy, and sertraline for the treatment of depression in multiple sclerosis. *Journal of Consulting and Clinical Psychology*, 69, 942–949. <http://dx.doi.org/10.1037/0022-006X.69.6.942>
- *Mohr, D. C., Ho, J., Duffecy, J., Reifler, D., Sokol, L., Burns, M. N., . . . Siddique, J. (2012). Effect of telephone-administered vs face-to-face cognitive behavioral therapy on adherence to therapy and depression outcomes among primary care patients: A randomized trial. *JAMA: Journal of the American Medical Association*, 307, 2278–2285. <http://dx.doi.org/10.1001/jama.2012.5588>
- *Murphy, G. E., Carney, R. M., Knesevich, M. A., Wetzel, R. D., & Whitworth, P. (1995). Cognitive behavior therapy, relaxation training, and tricyclic antidepressant medication in the treatment of depression. *Psychological Reports*, 77, 403–420. <http://dx.doi.org/10.2466/pr0.1995.77.2.403>
- *Murphy, G. E., Simons, A. D., Wetzel, R. D., & Lustman, P. J. (1984). Cognitive therapy and pharmacotherapy. Singly and together in the treatment of depression. *Archives of General Psychiatry*, 41, 33–41. <http://dx.doi.org/10.1001/archpsyc.1984.01790120037006>
- Nüesch, E., Trelle, S., Reichenbach, S., Rutjes, A. W. S., Bürgi, E., Scherer, M., . . . Jüni, P. (2009). The effects of excluding patients from the analysis in randomised controlled trials: Meta-epidemiological study. *British Medical Journal*, 339, b3244.
- *O'Mahen, H., Himle, J. A., Fedock, G., Henshaw, E., & Flynn, H. (2013). A pilot randomized controlled trial of cognitive behavioral therapy for perinatal depression adapted for women with low incomes. *Depression and Anxiety*, 30, 679–687. <http://dx.doi.org/10.1002/da.22050>
- Orsini, N., Bottai, M., Higgins, J., & Buchan, I. (2006). *HETEROGI: Stata module to quantify heterogeneity in a meta-analysis*. Retrieved from <http://econpapers.repec.org/software/bococode/s449201.htm>
- *Parker, G., Blanch, B., Paterson, A., Hadzi-Pavlovic, D., Sheppard, E., Manicavasagar, V., . . . Perich, T. (2013). The superiority of antidepressant medication to cognitive behavior therapy in melancholic depressed patients: A 12-week single-blind randomized study. *Acta Psychiatrica Scandinavica*, 128, 271–281.
- *Power, M. J., & Freeman, C. (2012). A randomized controlled trial of IPT versus CBT in primary care: With some cautionary notes about handling missing values in clinical trials. *Clinical Psychology & Psychotherapy*, 19, 159–169. <http://dx.doi.org/10.1002/cpp.1781>
- *Propst, L. R., Ostrom, R., Watkins, P., Dean, T., & Mashburn, D. (1992). Comparative efficacy of religious and nonreligious cognitive-behavioral therapy for the treatment of clinical depression in religious individuals. *Journal of Consulting and Clinical Psychology*, 60, 94–103. <http://dx.doi.org/10.1037/0022-006X.60.1.94>
- *Quilty, L. C., Dozois, D. J. A., Lobo, D. S. S., Ravindran, L. N., & Bagby, R. M. (2014). Cognitive structure and processing during cognitive behavioral therapy vs. pharmacotherapy for depression. *International Journal of Cognitive Therapy*, 7, 235–250. <http://dx.doi.org/10.1521/ijct.2014.7.3.235>
- *Quilty, L. C., McBride, C., & Bagby, R. M. (2008). Evidence for the cognitive mediational model of cognitive behavioural therapy for depression. *Psychological Medicine*, 38, 1531–1541. <http://dx.doi.org/10.1017/S0033291708003772>
- *Rieu, J., Bui, E., Rouch, V., Faure, K., Birmes, P., & Schmitt, L. (2011). Efficacy of ultrabrief cognitive and behavioural therapy performed by psychiatric residents on depressed inpatients. *Psychotherapy and Psychosomatics*, 80, 374–376. <http://dx.doi.org/10.1159/000323406>
- *Rush, A. J., Beck, A. T., Kovacs, M., & Hollon, S. (1977). Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. *Cognitive Therapy and Research*, 1, 17–37. <http://dx.doi.org/10.1007/BF01173502>
- *Savard, J., Simard, S., Giguère, I., Ivers, H., Morin, C. M., Maunsell, E., . . . Marceau, D. (2006). Randomized clinical trial on cognitive therapy for depression in women with metastatic breast cancer: Psychological and immunological effects. *Palliative & Supportive Care*, 4, 219–237. <http://dx.doi.org/10.1017/S1478951506060305>

- *Scott, A. I., & Freeman, C. P. (1992). Edinburgh primary care depression study: Treatment outcome, patient satisfaction, and cost after 16 weeks. *BMJ*, *304*, 883–887. <http://dx.doi.org/10.1136/bmj.304.6831.883>
- *Scott, C., Tacchi, M. J., Jones, R., & Scott, J. (1997). Acute and one-year outcome of a randomised controlled trial of brief cognitive therapy for major depressive disorder in primary care. *The British Journal of Psychiatry*, *171*, 131–134. <http://dx.doi.org/10.1192/bjp.171.2.131>
- *Selmi, P. M., Klein, M. H., Greist, J. H., Sorrell, S. P., & Erdman, H. P. (1991). Computer-administered therapy for depression. *Computers in Medical Practice*, *8*, 98–102.
- *Serfaty, M. A., Haworth, D., Blanchard, M., Buszewicz, M., Murad, S., & King, M. (2009). Clinical effectiveness of individual cognitive behavioral therapy for depressed older people in primary care: A randomized controlled trial. *Archives of General Psychiatry*, *66*, 1332–1340. <http://dx.doi.org/10.1001/archgenpsychiatry.2009.165>
- *Shapiro, D. A., Barkham, M., Rees, A., Hardy, G. E., Reynolds, S., & Startup, M. (1994). Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology*, *62*, 522–534. <http://dx.doi.org/10.1037/0022-006X.62.3.522>
- *Strauman, T. J., Vieth, A. Z., Merrill, K. A., Kolden, G. G., Woods, T. E., Klein, M. H., . . . Kwapiil, L. (2006). Self-system therapy as an intervention for self-regulatory dysfunction in depression: A randomized comparison with cognitive therapy. *Journal of Consulting and Clinical Psychology*, *74*, 367–376. <http://dx.doi.org/10.1037/0022-006X.74.2.367>
- *Taylor, F. G., & Marshall, W. L. (1977). Experimental analysis of a cognitive-behavioral therapy for depression. *Cognitive Therapy and Research*, *1*, 59–72. <http://dx.doi.org/10.1007/BF01173505>
- *Teichman, Y., Bar-el, Z., Shor, H., Sirota, P., & Elizur, A. (1995). A comparison of two modalities of cognitive therapy (individual and marital) in treating depression. *Psychiatry*, *58*, 136–148.
- *Teismann, T., Dymel, W., Schulte, D., & Willutzki, U. (2011). Ressourcenorientierte Akutbehandlung unipolarer Depressionen: Eine randomisierte kontrollierte Psychotherapiestudie [Resource-focused treatment for unipolar depression: A randomized controlled psychotherapy study]. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, *61*, 295–302. <http://dx.doi.org/10.1055/s-0030-1270453>
- *Thompson, L. W., Coon, D. W., Gallagher-Thompson, D., Sommer, B. R., & Koin, D. (2001). Comparison of desipramine and cognitive/behavioral therapy in the treatment of elderly outpatients with mild-to-moderate depression. *The American Journal of Geriatric Psychiatry*, *9*, 225–240. <http://dx.doi.org/10.1097/00019442-200108000-00006>
- *Thompson, L. W., Gallagher, D., & Breckenridge, J. S. (1987). Comparative effectiveness of psychotherapies for depressed elders. *Journal of Consulting and Clinical Psychology*, *55*, 385–390. <http://dx.doi.org/10.1037/0022-006X.55.3.385>
- *Tovote, K. A., Fleeer, J., Snippe, E., Peeters, A. C. T. M., Emmelkamp, P. M. G., Sanderman, R., . . . Schroevers, M. J. (2014). Individual mindfulness-based cognitive therapy and cognitive behavior therapy for treating depressive symptoms in patients with diabetes: Results of a randomized controlled trial. *Diabetes Care*, *37*, 2427–2434. <http://dx.doi.org/10.2337/dc13-2918>
- Tuttle, A. H., Tohyama, S., Ramsay, T., Kimmelman, J., Schweinhardt, P., Bennett, G. J., & Mogil, J. S. (2015). Increasing placebo responses over time in U.S. clinical trials of neuropathic pain. *Pain*, *156*, 2616–2626. <http://dx.doi.org/10.1097/j.pain.0000000000000333>
- *Wagner, B., Horn, A. B., & Maercker, A. (2014). Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial. *Journal of Affective Disorders*, *152–154*, 113–121. <http://dx.doi.org/10.1016/j.jad.2013.06.032>
- *Watson, J. C., Gordon, L. B., Stermac, L., Kalogerakos, F., & Steckley, P. (2003). Comparing the effectiveness of process-experiential with cognitive-behavioral psychotherapy in the treatment of depression. *Journal of Consulting and Clinical Psychology*, *71*, 773–781. <http://dx.doi.org/10.1037/0022-006X.71.4.773>
- *Wierzbicki, M., & Bartlett, T. S. (1987). The efficacy of group and individual cognitive therapy for mild depression. *Cognitive Therapy and Research*, *11*, 337–342. <http://dx.doi.org/10.1007/BF01186284>
- *Wiles, N., Thomas, L., Abel, A., Ridgway, N., Turner, N., Campbell, J., . . . Lewis, G. (2013). Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: Results of the CoBalT randomised controlled trial. *The Lancet*, *381*, 375–384. [http://dx.doi.org/10.1016/S0140-6736\(12\)61552-9](http://dx.doi.org/10.1016/S0140-6736(12)61552-9)
- *Wilson, P. H., Goldin, J. C., & Charbonneau-Powis, M. (1983). Comparative efficacy of behavioral and cognitive treatments of depression. *Cognitive Therapy and Research*, *7*, 111–124. <http://dx.doi.org/10.1007/BF01190064>
- *Wright, J. H., Wright, A. S., Albano, A. M., Basco, M. R., Goldsmith, L. J., Raffield, T., & Otto, M. W. (2005). Computer-assisted cognitive therapy for depression: Maintaining efficacy while reducing therapist time. *The American Journal of Psychiatry*, *162*, 1158–1164. <http://dx.doi.org/10.1176/appi.ajp.162.6.1158>
- *Zu, S., Xiang, Y.-T., Liu, J., Zhang, L., Wang, G., Ma, X., . . . Li, Z.-J. (2014). A comparison of cognitive-behavioral therapy, antidepressants, their combination and standard treatment for Chinese patients with moderate-severe major depressive disorders. *Journal of Affective Disorders*, *152–154*, 262–267. <http://dx.doi.org/10.1016/j.jad.2013.09.022>

Received November 3, 2015
 Revision received May 18, 2016
 Accepted May 23, 2016 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://notify.apa.org/> and you will be notified by e-mail when issues of interest to you become available!