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...
THE EFFECTS OF CBT ARE NOT SYSTEMATICALLY FALLING

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 (I²) 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 (Ioanidis, 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 effects 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 Friborg (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 Friborg (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 Friborg (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 Friborg (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 Friborg (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 Friborg (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 Friborg (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; DiRiessen, 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 outpatients 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 ESPs 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 $\chi^2$-based approach with the heterogei 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 HRSD17. 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, Quality, & 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 (Liberman & 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, Schmaling, & 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 “complete” 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 ($b_1 = -0.04; 95\% \text{ CI}: -0.07 \text{ to } -0.01; p = .01$) 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
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 respectively the HRSD only found recruitment from clinical samples as significantly and negatively associated to outcomes, while the BDI and respectively the HRSD only had both ITT and “completer” data).

**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 & Friborg (2015) Only (“Completer” Data Preferred Over ITT)

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Single</th>
<th></th>
<th></th>
<th>Multiple: Parsimonious</th>
<th></th>
<th></th>
<th>Multiple: Full model</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff 95% CI p&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Coeff 95% CI p&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Coeff 95% CI p&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Coeff 95% CI p&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Coeff 95% CI p&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Coeff 95% CI p&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Coeff 95% CI p&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Coeff 95% CI p&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Coeff 95% CI p&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>HRSD + BDI-I + BDI-II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>-0.03 -0.05--0.01 .002</td>
<td>-0.03 -0.05--0.01 .001</td>
<td>-0.02 -0.04--0.01 .008</td>
<td>-0.02 -0.04--0.01 .008</td>
<td>-0.02 -0.04--0.01 .008</td>
<td>-0.02 -0.04--0.01 .008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>-0.00 -0.02--0.01 .58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RoB</td>
<td>-1.13 -1.33--1.06 .17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WL/CAU/PLA present</td>
<td>-0.00 -0.06--0.00 .99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recr from clinical samples</td>
<td>-0.08 -0.08--0.02 .74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDD vs other diagnosis</td>
<td>0.09 -0.06--0.22 .93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adults vs specific group</td>
<td>0.28 -0.21--0.35 .63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of sessions</td>
<td>0.04 -0.01--0.06 .12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country (U.S. vs other)</td>
<td>0.56 .15--0.98 .009</td>
<td>0.39 -0.01--0.79 .05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beck’s manual vs other</td>
<td>0.18 -0.34--0.34 .49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inpatients vs outpatients</td>
<td>-0.58 -2.06--0.89 .43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDI (I + II)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>-0.03 -0.06--0.01 .002</td>
<td>-0.03 -0.05--0.01 .008</td>
<td>-0.01 -0.04--0.01 .31</td>
<td>-0.01 -0.05--0.03 .74</td>
<td>-0.01 -0.05--0.03 .74</td>
<td>-0.01 -0.05--0.03 .74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>-0.01 -0.02--0.00 .27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RoB</td>
<td>-1.14 -3.05--1.02 .22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WL/CAU/PLA present</td>
<td>-0.02 -0.53--0.49 .93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recr from clinical samples</td>
<td>-1.13 -0.70--0.44 .63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDD vs other diagnosis</td>
<td>0.17 -0.56--0.29 .53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adults vs specific group</td>
<td>0.38 -0.27--0.57 .17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of sessions</td>
<td>0.03 -0.02--0.09 .20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country (U.S. vs other)</td>
<td>0.61 1.41-1.07 .01</td>
<td>0.43 -0.08--0.86 .05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beck’s manual vs other</td>
<td>0.35 -0.23--0.36 .73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inpatients vs outpatients</td>
<td>-0.59 -2.21-1.20 .46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRSD (all versions)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>-0.03 -0.05--0.00 .02</td>
<td>-0.01 -0.04--0.01 .31</td>
<td>-0.01 -0.05--0.03 .74</td>
<td>-0.01 -0.05--0.03 .74</td>
<td>-0.01 -0.05--0.03 .74</td>
<td>-0.01 -0.05--0.03 .74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>0.00 -0.01--0.02 .71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RoB</td>
<td>-0.10 -0.29--0.09 .27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WL/CAU/PLA present</td>
<td>-0.20 -0.71--0.32 .44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recr from clinical samples</td>
<td>0.09 -0.48--0.66 .75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDD vs other diagnosis</td>
<td>0.00 -0.56--0.57 .99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adults vs specific group</td>
<td>0.48 -0.00--0.97 .08</td>
<td>0.38 -0.07--0.82 .09</td>
<td>0.42 -0.16--1.00 .15</td>
<td>0.42 -0.16--1.00 .15</td>
<td>0.42 -0.16--1.00 .15</td>
<td>0.42 -0.16--1.00 .15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of sessions</td>
<td>0.06 -0.02--0.12 .01</td>
<td>0.05 -0.10--0.10 .09</td>
<td>0.06 -0.06--0.20 .29</td>
<td>0.06 -0.06--0.20 .29</td>
<td>0.06 -0.06--0.20 .29</td>
<td>0.06 -0.06--0.20 .29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country (U.S. vs other)</td>
<td>0.36 -0.16--0.99 .17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beck’s manual vs other</td>
<td>-0.05 -0.73--0.68 .88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inpatients vs outpatients</td>
<td>-0.64 -2.01--1.73 .35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.

<sup>a</sup> 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).

included. 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 Friborg (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 Friborg (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 Friborg (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 Friborg (2015) for the BDI for (a) studies in the U.S., (b) studies outside the US.
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 heterogeneity 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-

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

**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”)

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Coeff</th>
<th>95% CI</th>
<th>p²</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDI (I + II)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All studies</td>
<td>−.01</td>
<td>−.03–.01</td>
<td>.22</td>
</tr>
<tr>
<td>Inpatients</td>
<td>−.01</td>
<td>−.03–.01</td>
<td>.23</td>
</tr>
<tr>
<td>Outliers</td>
<td>−.01</td>
<td>−.02–.01</td>
<td>.38</td>
</tr>
<tr>
<td>Inpatients &amp; outliers</td>
<td>−.01</td>
<td>−.02–.01</td>
<td>.39</td>
</tr>
<tr>
<td>Only low RoB studies (&gt;2)</td>
<td>.03</td>
<td>−.02–.07</td>
<td>.25</td>
</tr>
<tr>
<td>BDI-I only</td>
<td>−.00</td>
<td>−.03–.02</td>
<td>.59</td>
</tr>
<tr>
<td>BDI-II only</td>
<td>.01</td>
<td>−.09–.10</td>
<td>.85</td>
</tr>
<tr>
<td>HRSD (all versions)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All studies</td>
<td>.00</td>
<td>−.02–.02</td>
<td>.77</td>
</tr>
<tr>
<td>Inpatients</td>
<td>.00</td>
<td>−.02–.02</td>
<td>.75</td>
</tr>
<tr>
<td>Only low RoB studies (&gt;2)</td>
<td>.03</td>
<td>−.03–.08</td>
<td>.28</td>
</tr>
<tr>
<td>Only HRSD-17</td>
<td>.00</td>
<td>−.02–.02</td>
<td>.89</td>
</tr>
</tbody>
</table>

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

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).

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 Friborg (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 Friborg (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 Friborg (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 Friborg (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-
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 systemically 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 Friiborg (2015) is simply a spurious finding.

### References

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


THE EFFECTS OF CBT ARE NOT SYSTEMATICALLY FALLING

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.


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!