




# Is Behavioral Activation (BA) More Effective than Cognitive Therapy (CT) in Severe Depression? A Reanalysis of a Landmark Trial

Lorenzo Lorenzo-Luaces<sup>1,2</sup>  · Keith S. Dobson<sup>2</sup>

Published online: 29 April 2019  
© The Author(s) 2019

## Abstract

A landmark study (Jacobson et al., *JCCP*, 64:295–304, 1996) suggested that behavioral activation (BA) is as effective as cognitive therapy (CT) in the treatment of major depression. A conceptual replication supported the efficacy of BA and suggested BA is more effective than CT for severe depression (Dimidjian et al., *JCCP*, 74:658–670, 2006), though these findings have never been replicated.

Outcome data from the participants in the BA and CT condition of the Jacobson et al. (*JCCP*, 64:295–304, 1996) trial were analyzed with the same analytic approach used by Dimidjian et al. (*JCCP*, 74:658–670, 2006). The sample was stratified on the Hamilton Rating Scale for Depression (HRSD) as higher-severity ( $HRSD \geq 20$ ) and lower-severity ( $HRSD \leq 19$ ). Treatment differences in change over time on the HRSD, Beck Depression Inventory (BDI), as well as response ( $\geq 50\%$  change) and remission on each scale, were examined. Tests of moderation with severity as a categorical or continuous variable were conducted, and we explored the effect of severity by treatment on relapse. The results of Dimidjian et al. (*JCCP*, 74:658–670, Dimidjian et al. 2006) were not replicated. Tests of moderation with severity as a continuously measured variable (with the BDI or HRSD) also failed to find that BA was more effective than CT for more severe depression. No differences by severity emerged over the follow-up period.

BA and CT may be roughly equivalent in the treatment of mild, moderate, and severe depression.

**Keywords** Behavioral activation · Cognitive therapy · Moderators · Depression

---

✉ Lorenzo Lorenzo-Luaces  
lolorenz@indiana.edu

<sup>1</sup> Department of Psychological and Brain Sciences, Indiana University–Bloomington, 1101 E 10th St, Bloomington, IN 47405, USA

<sup>2</sup> Department of Psychology, University of Calgary, Calgary, Alberta, Canada

## Introduction

Contemporary behavioral activation (BA) emerged as a treatment for depression following a landmark dismantling study (Jacobson et al. 1996). Jacobson et al. conjectured that the full CT for depression treatment could be divided into three broadly defined components: (1) behavioral activation (BA); (2) challenging automatic thoughts (ATs); and (3) modifying core beliefs (CBs). They randomized 151 participants to three conditions lasting a maximum of 20 sessions: (1) 100% BA; (2) all the elements of BA and AT work; and (3) “full CT” including BA, AT, and a minimum of 8 sessions devoted to CBs. Across various metrics, no statistically or clinically significant differences were observed during acute treatment or over the 2-year follow-up period (Gortner et al. 1998).<sup>1</sup>

The BA condition of the Jacobson trial was expanded to feature functional analysis of behavior as a way of understanding as well as explaining depression and countering depressotypic patterns of behaviors (Martell et al. 2001). This form of BA was tested against CT, paroxetine, and a placebo control group in a sample of 240 patients (Dimidjian et al. 2006). In contrast to the earlier trial, however, the sample of that study was recruited and stratified based on the Hamilton Rating Scale for Depression (HRSD; Hamilton 1960) into a “higher severity” ( $HRSD \geq 20$ ) and “lower severity” ( $HRSD \leq 19$ ) groups. Within the lower-severity group, there were no differences between the three active treatment groups. Within the high-severity group, BA outperformed CT, especially in regard to response (i.e.,  $\geq 50\%$  reduction) on the Beck Depression Inventory (BDI; Beck et al. 1961) with 76% of high-severity patients meeting criteria for response in BA compared to 48% in CT and 49% with paroxetine. A subsequent trial also found BA to be superior to treatment with antidepressant medications (Moradveisi et al. 2013) and more effective in treating more severe depression.

Despite optimism regarding the promise of BA for severe depression, the data do not unequivocally suggest BA should be preferred to CT for more severe depression. In the recent COBRA trial, a large ( $N=440$ ) study that examined the efficacy of BA in primary care, Richards et al. (2016) did not find that BA was more efficacious than CT for severe depression. Additionally, process research does not support a differential effect of cognitive vs. behavioral interventions by severity. In one study of 60 patients with moderate to severe depression, Sasso et al. (2015) did not find that behavioral interventions were superior to cognitive ones among patients with more severe depression. By contrast, Hawley et al. (2017) reported that patient’s use of cognitive skills predicted subsequent symptom change irrespective of symptom severity. Unexpectedly, in that study, patient’s use of behavioral skills was predictive of greater symptom change among patients with milder, rather than more severe, symptoms of depression.

The current study examined the hypothesis that BA is more efficacious than CT in severe depression using data from the Jacobson et al. (1996) trial. We

---

<sup>1</sup> There is confusion in the literature regarding the use of “cognitive therapy (CT)” versus “cognitive behavioral therapy (CBT).” CBT can be applied to refer to a family of interventions that are either cognitive, behavioral, or both. CBT can also refer to a specific intervention package that combines behavioral and cognitive interventions. CT focuses on challenging depressogenic cognitions using a set of strategies that may be cognitive or behavioral. We use the term CT for consistency as it was the one employed by Jacobson et al. (1996).

sought to expand on the analyses by Dimidjian et al. (2006) by conducting a formal test of moderation, which is more appropriate than the stratification analyses Dimidjian et al. reported (Pocock et al. 2002). Stratified or subgroup (e.g., splitting the sample by severity and comparing treatment effects across levels of severity) analyses, are highly subject to spuriousness (Assmann et al. 2000). If one divides a dataset according to a specific variable and conducts statistical tests across subgroups, by chance, one subgroup will have a higher  $p$  value than another, and this  $p$  value may go over the  $p < 0.05$  threshold. These stratified subgroup analyses are also limited by virtue of smaller sample sizes and the possibility of unequal variances across the subsamples. Moderation analyses, in which the interaction between the subgrouping variable of interest and the focal moderator are made to interact to predict outcome are preferable to test subgroup questions (Pocock et al. 2002), in part because they require more power. In addition to testing the moderator findings during the acute phase of treatment, we also explored whether there were long-term differences in BA and CT according to severity, a question that was not explored in the follow-up study to the Dimidjian et al. trial (Dobson et al. 2008).

## Methods

### Sample

The sample for the present analyses consisted of 107 of the 151 participants described by Jacobson et al. (1996). We only included the “pure” BA and “full” CT condition used in that study because they mirror the ways BA and CT are conducted. Participants in the study met the Diagnostic and Statistical Manual of Mental Disorders 3rd edition, revised (American Psychiatric Association 1987) definition of major depression as assessed by the Structured Clinical Interview for DSM-III-R (Spitzer et al. 1992). Additionally, participants were required to have both a score  $\geq 20$  on the BDI (Beck et al. 1961) and a score  $\geq 14$  on the HRSD (Hamilton 1960). Exclusion criteria included bipolar disorder, past or present psychosis, panic disorder, current substance abuse, organic brain syndrome, mental retardation, or the presence of imminent suicidal risk or psychosis, or active outpatient treatment. The study was approved by the University of Washington Institutional Review Board.

### Treatment

Randomization was based on matching for prior depressive episodes, severity, comorbid dysthymia, depression severity, sex, and marital status. The treatment conditions in the original trial were BA, AT, and CT. Our analyses focused on the BA and CT conditions, but all results reported with the BA vs. CT contrast apply to the contrast of BA vs. AT/CT. The treatment conditions can briefly be described as:

BA—The aim of the behavioral activation condition was to foster meaningful engagement with the environment. Interventions included activity monitoring, assessment of pleasure and mastery, graded task assignment, problem-solving, and social skills training.

CT—The full CT condition included elements of the BA as well as elements of the AT condition which was focused on the identification of cognitive distortions and their modification via completing thought records to assess the validity of beliefs, responding in more functional ways to negative thoughts, and behavioral experiments. It also added a focus on schemas or core beliefs. These interventions are aimed at revealing an underlying assumption that cuts across specific situations (e.g., “I am unlovable”) as well as those that explore its pros and cons, and possible alternatives. Therapists in the trial were required to focus on this kind of work for at least eight sessions.

### Outcome Measures and Analytic Strategy

All participants were assessed and administered the BDI and HRSD before therapy, at the time of termination, and at 6-, 12-, 18-, and 24-month follow-ups. The BDI was administered before every treatment session. The timing of the BDI and HRSD differs somewhat between the current trial and the Dimidjian et al. (2006) trial, as the Dimidjian et al. (2006) trial had significantly fewer BDI assessments (i.e., only pre-, mid-, and post-treatment as well as early termination or as clinically indicated) but used an additional HRSD assessment mid-treatment. Beyond these minor differences, we closely followed the analytic plan outlined by Dimidjian et al. (2006; see page 662). For example, gender was controlled in all of our analyses because it was differentially represented, and controlled for, in that trial.

Response on the BDI and HRSD was defined as a 50% or more reduction in the pre-treatment scores on these measures. Remission was defined as scores  $\leq 7$  on the HRSD and  $\leq 10$  on the BDI. As in Dimidjian et al. (2006), separate analyses were conducted for each outcome metric within the higher (i.e.,  $\text{HRSD} \geq 20$ ) and lower (i.e.,  $\text{HRSD} \leq 19$ ) severity subgroup. A hierarchical linear model (HLM), was used to investigate treatment differences in change over time on the BDI with the full intent to treat (ITT) sample. This HLM model included the mixed effect of time (i.e., the session number with session 1 being ‘0,’ session 20 being ‘1’ and other sessions as fractions) as well as treatment condition (BA vs. CT, coded  $\pm 0.5$ ), gender, baseline BDI, and their interactions with time. Random effects for the intercept and slopes were used for these analyses. Because we only had two observations of the HRSD during acute treatment, a general linear model was conducted in which raw change was regressed on the treatment condition and gender. Because the end-of-treatment HRSD was skewed, for these analyses we transformed the scores according to the two-step variable transformation procedure proposed by Templeton (2011) which retains the mean and standard deviation of the variable. Treatment differences in categorical rates of response, remission, and their combination were examined using the Cochran–Mantel–Haenszel (CMH) tests. All analyses were conducted using last observation carried forward (LOCF).

Formal tests of moderation on severity, assessed categorically and continuously, were also conducted. In these HLMs, the BDI at each session was regressed on the mixed effects of time, treatment condition, baseline severity (when categorical,  $\pm 0.5$ , when continuous, mean centered), and the full factorial of time, treatment condition, and severity. The time by treatment condition by severity interaction indicates whether there were differences in the rate of change across the range of severity. We also

evaluated the effects of treatment, by severity, over the follow-up period using a Cox regression to model depressive relapse (defined as meeting major depressive disorder criteria) among responders on the HRSD who were followed over 2 years. Indices of effect size included Cohen's  $d$  for continuous differences, the  $d$ -type effect size described by Feingold (2013) for HLM, and odds ratios (OR) for categorical differences. Because we coded treatment condition as 0.5 for BA and  $-0.5$  for CT,  $d$ -type effects sizes that are positive or ORs over 1 indicate a superiority of BA over CT.  $d$ -type effects sizes that are negative or ORs less than 1 indicate a superiority of CT over BA.

## Results

The sample was predominantly composed of Caucasian (BA, 92.6%;  $n = 50$ ; CT, 76%;  $n = 38$ ) female (BA, 71.9%;  $n = 41$ ; CT, 76.0%;  $n = 38$ ), adults (BA,  $M = 36.6$ ; CT,  $M = 39.2$ ), only a minority of whom were married (BA, 33.3%;  $n = 19$ ; CT = 30%;  $n = 15$ ). As reported previously (Jacobson et al. 1996), participants randomized to the BA ( $M = 17.38$ ,  $SD = 3.83$ ) condition had lower HRSD scores ( $t(105) = -2.15$ ,  $p = 0.03$ ) than participants randomized to CT ( $M = 19.10$ ,  $SD = 4.41$ ). As a result, there were somewhat fewer high-severity patients in BA (24.6%,  $n = 14/57$ ) than in CT (40.0%,  $n = 20/50$ ;  $\chi^2 = 2.93$ ,  $p = 0.09$ ).

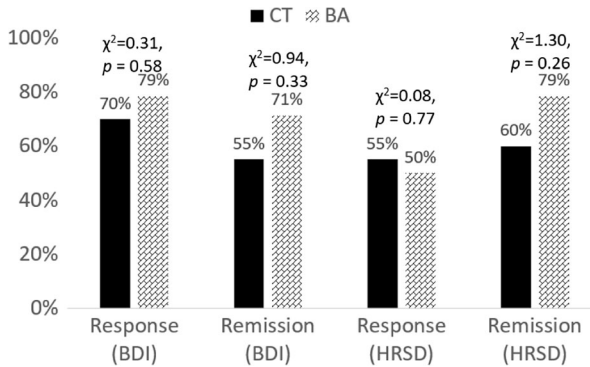
### Replication of Dimidjian et al. 2006

As shown in Table 1, both the higher and lower severity groups experienced improvement over the course of treatment ( $ps < 0.001$ ). Contrary to Dimidjian et al. (2006), for patients high on symptom severity, there were no significant differences in BA vs. CT in changes in depressive symptoms over time ( $B = 0.41$ ,  $SE = 4.24$ ,  $t(27.07) = 0.10$ ,  $p =$

**Table 1** Hierarchical linear models predicting change over time in behavioral activation (BA) vs. cognitive therapy (CT) in lower ( $HRSD \leq 19$ ;  $n = 73$ ) and higher severity ( $HRSD \geq 20$ ;  $n = 34$ ) depressed patients

Low severity	B	SE	df	$t$	$d$
Intercept	14.15	3.56	69.60	3.97***	
Session	-9.95	4.86	55.69	-2.05*	
BA (vs. CT)	-0.69	1.73	68.94	-0.40	
BA*Session	1.93	2.32	55.59	0.83	-0.46
High severity	B	SE	df	$t$	$d$
Intercept	28.54	6.68	30.80	4.28***	
Session	-22.41	8.32	28.04	-2.70*	
BA (vs. CT)	-3.54	3.51	30.61	-1.04	
BA*Session	0.41	4.24	27.07	0.10	0.10

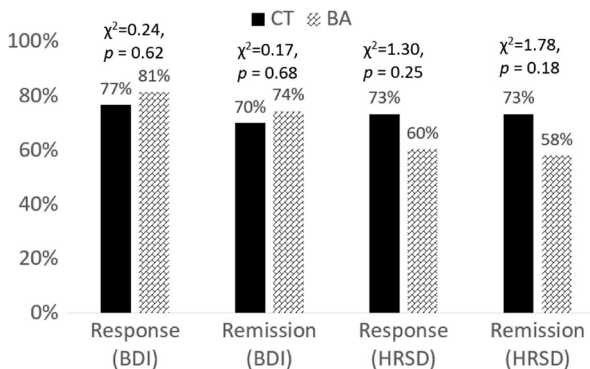
Note. Models controlling for sex, baseline BDI, and their respective effects on change over time. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ . A positive value on 'd' favors BA and a negative value favors CT. Sessions ranged from 1 – 20. For ease of interpretability, they were coded as a '0' for 1, '1' for 20, and fractions of 1 (e.g., session 10 was "0.5")



**Fig. 1** Response and remission in the Hamilton Depression Rating Scale (HRSD) and the Beck Depression Inventory for depressed patients with higher severity (HRSD  $\geq$  20;  $n = 34$ )

0.99,  $d = 0.10$ ). Similarly, there were no significant differences between BA and CT in raw change on the HRSD ( $B = 3.36$ ,  $SE = 4.10$ ,  $t(30) = 0.82$ ,  $p = 0.42$ ,  $\beta = 0.31$ ). Consistent with the findings of Dimidjian et al. (2006), there were no significant differences in change over time on the BDI in BA vs. CT for patients lower on symptom severity ( $B = 1.92$ ,  $SE = 2.31$ ,  $t(55.59) = 0.83$ ,  $p = 0.83$ ,  $d = -0.46$ ). Similarly, for patients lower on symptom severity, there were no significant differences between BA and CT in raw change on the HRSD ( $B = -0.34$ ,  $SE = 2.18$ ,  $t(69) = -0.16$ ,  $p = 0.88$ ,  $d = -0.05$ ).

Figure 1 shows the treatment differences in response and remission on the BDI and HRSD among the more severely depressed participants (HRSD  $\geq$  20). There were no significant differences between the treatments ( $ps > 0.26$ ) across the categorical outcomes. The largest observed difference, in remission rates on the HRSD, favored BA (BA = 78.6%, CT = 60.0%, OR = 3.14, 95% CI = 0.57–17.19,  $\chi^2(1, n = 34) = 1.30$ ,  $p = 0.26$ ). Figure 2 shows the remission and response rates on the BDI and HRSD among the more mildly depressed participants (HRSD  $\leq$  19). None of the contrasts was statistically significant (all  $ps > 0.18$ ). The largest observed difference was on remission on the HRSD, and it favored CT (73.3%) over BA (58.1%) ( $\chi^2(1, n = 73) = 1.78$ ,  $p = 0.18$ , OR = 0.49, 95% CI = 0.17–1.37).



**Fig. 2** Response and remission in the Hamilton Depression Rating Scale (HRSD) and the Beck Depression Inventory for depressed patients with lower severity (HRSD  $\leq$  19;  $n = 73$ )

## Extension of Dimidjian et al. (2006)

**Moderator Analyses** There was no evidence of an interaction between treatment condition and time according to severity when coded as a binary variable on the HRSD ( $B = -1.89$ ,  $SE = 4.48$ ,  $t(81.20) = -0.43$ ,  $p = 0.67$ ). This same result was found when severity was based on the continuous baseline score on the HRSD ( $B = -0.31$ ,  $SE = 0.56$ ,  $t(89.57) = -0.55$ ,  $p = 0.59$ ), and the continuous baseline score on the BDI ( $B = -0.20$ ,  $SE = 0.30$ ,  $t(82.70) = -0.67$ ,  $p = 0.51$ ).

**Severity and Long-Term Outcomes** Seventy of the 107 patients (65%) met response criteria on the HRSD. Complete relapse data was available for most of these patients (91%).<sup>2</sup> BA (54.5%) was not more effective than CT in preventing relapse in severe depression (50.0%;  $HR = 1.35$ ,  $B = 0.30$ ,  $SE = 0.63$ ,  $\chi^2(1) = 0.22$ ,  $p = 0.63$ ). Similarly, there were no differences between BA (42.9%) and CT (36.4%) in relapse rates among the milder cases ( $HR = 0.67$ ,  $B = -0.39$ ,  $SE = 0.50$ ,  $\chi^2(1) = 0.60$ ,  $p = 0.44$ ). Consequently, the interaction between severity group and the treatment condition in predicting relapse hazard was not statistically significant ( $HR = 1.99$ ,  $B = 0.68$ ,  $SE = 0.79$ ,  $\chi^2(1) = 0.76$ ,  $p = 0.38$ ). Further, there were no statistically significant interactions between the treatment conditions and baseline HRSD ( $HR = 1.17$ ,  $B = 0.16$ ,  $SE = 0.10$ ,  $\chi^2(1) = 2.52$ ,  $p = 0.11$ ) or baseline BDI ( $HR = 0.93$ ,  $B = -0.07$ ,  $SE = 0.08$ ,  $\chi^2(1) = 0.81$ ,  $p = 0.93$ ) in predicting outcome.

## Discussion

The current study examined the hypothesis that BA is more effective than CT for severe depression, by employing a very similar analytic strategy to the study that reported this original finding (Dimidjian et al. 2006). In line with the findings from the COBRA trial (Richards et al. 2016), we found no evidence for the superiority of BA relative to CT for more severe depression across acute treatment or a long-term follow-up.

Although our results were clear, there are limitations of the study that warrant consideration. First, this was a secondary analysis of relatively old study data. The BA condition from the Jacobson trial was expanded by Dimidjian et al. (2006) and are thus not fully comparable. Moreover, the study has a relatively small sample size. Several factors and strengths mitigate these limitations. Although the two trials did not follow the exact same manual, Jacobson and Gortner (2000) deemphasized the differences between the two treatments when they stated that they “kept the same set of treatment options, but created a behavior analytic theoretical framework” which guided therapist selection of interventions and was also taught to patients. Moreover, there is no evidence that BA with a focus on behavioral chain analysis is more effective than BA following a different rationale (Nyström et al. 2017). The Jacobson et al. (1996) and Dimidjian et al. (2006) studies were designed and conducted by similar teams, which suggests that the trials are more comparable than

<sup>2</sup> Running analyses imputing outcome data did not change the pattern of results.



randomized controlled trials usually are. Additionally, we closely matched our analytic strategy to mirror the Dimidjian et al. study to ensure that differences in findings are not attributable to statistical artifacts. Despite issues related to sample size, the number of people randomized to BA ( $n = 50$ ) and CT ( $n = 57$ ) in the current study is larger than the numbers randomized in the Dimidjian et al. study (43 and 45, respectively), and studies with 50 or more participants per treatment arm, as in the current case, are rare in depression treatment research (Barth et al. 2013).

Statistical and methodological artifacts cannot be ruled out as an explanation for either the current results or those of Dimidjian et al. (2006) and Moradveisi et al. (2013). As the debates over the issue of reproducibility in psychology have illuminated, spurious effects are not uncommon or unexpected (Pashler and Harris 2012). Effects resembling statistical moderation are especially sensitive to study and analysis design features and may be less likely to replicate than other results (Aguinis and Stone-Romero 1997).

Other factors may also explain why the effects reported by Dimidjian et al. (2006) and Moradveisi et al. (2013) were not replicated in the current study or in the larger COBRA trial. It is possible that the optimal matching of patients to BA vs. CT or medications, may not be contingent on a single variable such as severity but the interaction among several variables which were differentially represented across the trials. Driessen et al. (2016) illustrate how this pattern of results could occur in a study that modeled non-linear interactions among multiple baseline variables. In their study, a small advantage of psychodynamic therapy over CT was evidenced for patients with depression that was both severe and chronic whereas a large advantage of CT was observed if depression was both severe and non-chronic. In mild to moderate depression, there were no advantages of one treatment over the other unless anxiety levels were low, in which case psychodynamic therapy had an advantage over CT. It is possible that the superiority of BA over medications and CT for severe depression in the Dimidjian et al. and Moradveisi et al. trials interacted with unmeasured third variables. Another alternative explanation is that BA may be superior to CT in cases of severe anxiety, but not severe depression (Sasso et al. 2015), as in the Dimidjian et al. trial, the severity grouping was performed with the HRSD which captures symptoms of depression and anxiety (Porter et al. 2017).

In summary, BA and CT appear to be efficacious treatments for depression. While CT has more evidence for its efficacy, BA may be easier to implement by paraprofessionals (Richards et al. 2016). Our results suggest that severity, at least by itself, is not a moderator of their efficacy. Thus, more research is needed on multiple moderators of response to either treatment, as well as the most effective and cost-effective methods for delivery. In lieu of another large RCT like the COBRA trial, an individual patient data meta-analysis could identify moderators of response to BA vs. CT.

### Compliance with Ethical Standards

**Conflict of Interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.



**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology, 82*, 192–206. <https://doi.org/10.1037/0021-9010.82.1.192>.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental health disorders (DSM-III-R)*. Washington, D.C.: American Psychiatric Association.
- Assmann, S. F., Pocock, S. J., Enos, L. E., & Kasten, L. E. (2000). Subgroup analysis and other (mis) uses of baseline data in clinical trials. *The Lancet, 355*, 1064–1069. [https://doi.org/10.1016/S0140-6736\(00\)02039-0](https://doi.org/10.1016/S0140-6736(00)02039-0).
- Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., et al. (2013). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *PLoS Medicine, 10*, e1001454. <https://doi.org/10.1371/journal.pmed.1001454>.
- 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. <https://doi.org/10.1001/archpsyc.1961.01710120031004>.
- Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmalting, K. B., Kohlenberg, R. J., Addis, M. E., et al. (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. <https://doi.org/10.1037/0022-006X.74.4.658>.
- Dobson, K. S., Hollon, S. D., Dimidjian, S., Schmalting, K. B., Kohlenberg, R. J., Gallop, R. J., & Jacobson, N. S. (2008). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the prevention of relapse and recurrence in major depression. *Journal of Consulting and Clinical Psychology, 76*, 468–477. <https://doi.org/10.1037/0022-006X.76.3.468>.
- Driessen, E., Smits, N., Dekker, J. J. M., Peen, J., Don, F. J., Kool, S., et al. (2016). Differential efficacy of cognitive behavioral therapy and psychodynamic therapy for major depression: a study of prescriptive factors. *Psychological Medicine, 46*, 731–744. <https://doi.org/10.1017/S0033291715001853>.
- Feingold, A. (2013). A regression framework for effect size assessments in longitudinal modeling of group differences. *Review of General Psychology, 17*, 111–121. <https://doi.org/10.1037/a0030048>.
- Gortner, E. T., Gollan, J. K., Dobson, K. S., & Jacobson, N. S. (1998). Cognitive-behavioral treatment for depression: relapse prevention. *Journal of Consulting and Clinical Psychology, 66*, 377–384. <https://doi.org/10.1037/0022-006X.66.2.377>.
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23*, 56–62. <https://doi.org/10.1136/jnmp.23.1.56>.
- Hawley, L. L., Padesky, C. A., Hollon, S. D., Mancuso, E., Lapsa, J. M., Brozina, K., & Segal, Z. V. (2017). Cognitive-behavioral therapy for depression using mind over mood: CBT skill use and differential symptom alleviation. *Behavior Therapy, 48*, 29–44. <https://doi.org/10.1016/j.beth.2016.09.003>.
- Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., et al. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology, 64*, 295–304. <https://doi.org/10.1037/0022-006X.64.2.295>.
- Jacobson, N. S., & Gortner, E. T. (2000). Can depression be de-medicalized in the 21st century: scientific revolutions, counter-revolutions and the magnetic field of normal science. *Behaviour Research and Therapy, 38*, 103–117. <https://doi.org/10.1037/0022-006X.64.1.74>.
- Martell, C. R., Addis, M. E., & Jacobson, N. S. (2001). *Depression in context: Strategies for guided action*. New York: Guilford Press.
- Moradveisi, L., Huibers, M. J., Renner, F., Arasteh, M., & Arntz, A. (2013). Behavioural activation v. antidepressant medication for treating depression in Iran: randomised trial. *The British Journal of Psychiatry, 202*, 204–211. <https://doi.org/10.1192/bjp.bp.112.113696>.
- Nyström, M. B. T., Stenling, A., Sjöström, E., Neely, G., Lindner, P., Hassmén, P., Andersson, G., Martell, C., Carlbring, P. (2017). Behavioral activation versus physical activity via the internet: A randomized controlled trial. *Journal of Affective Disorders, 215*, 85–93.

- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7, 531–536. <https://doi.org/10.1177/1745691612463401>.
- Pocock, S. J., Assmann, S. E., Enos, L. E., & Kasten, L. E. (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in Medicine*, 21, 2917–2930. <https://doi.org/10.1002/sim.1296>.
- Porter, E., Chambless, D. L., McCarthy, K. S., DeRubeis, R. J., Sharpless, B. A., Barrett, M. S., et al. (2017). Psychometric properties of the reconstructed Hamilton depression and anxiety scales. *The Journal of Nervous and Mental Disease*, 205, 656–664. <https://doi.org/10.1097/NMD.0000000000000666>.
- Richards, D. A., Ekers, D., McMillan, D., Taylor, R. S., Byford, S., Warren, F. C., et al. (2016). Cost and outcome of Behavioural activation versus cognitive Behavioural therapy for depression (COBRA): A randomised, controlled, non-inferiority trial. *The Lancet*, 388, 871–880. [https://doi.org/10.1016/S0140-6736\(16\)31140-0](https://doi.org/10.1016/S0140-6736(16)31140-0).
- Sasso, K. E., Strunk, D. R., Braun, J. D., DeRubeis, R. J., & Brotman, M. A. (2015). Identifying moderators of the adherence-outcome relation in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 83, 976–984. <https://doi.org/10.1037/ccp0000045>.
- Spitzer, R. L., Williams, J. B., Gibbon, M., & First, M. B. (1992). The structured clinical interview for DSM-III-R (SCID): I: history, rationale, and description. *Archives of General Psychiatry*, 49, 624–629. <https://doi.org/10.1001/archpsyc.1992.01820080032005>.
- Templeton, G. F. (2011). A two-step approach for transforming continuous variables to normal: implications and recommendations for IS research. *Communications of the Association for Information Systems*, 28, 41–58. <https://doi.org/10.17705/1CAIS.02804>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.