Acceptability and Efficacy of Group Behavioral Activation for Depression among Adults: A Meta-Analysis

Mel Simmonds-Buckley

University of Sheffield, UK

Stephen Kellett

Centre for Psychological Services Research, University of Sheffield, UK, and

Sheffield Health and Social Care NHS Foundation Trust, UK

&

Glenn Waller

University of Sheffield, UK

Author Note

This work was supported by the Howard Morton Trust Studentship, University of Sheffield as part of the first author's doctorate. The funders had no role in the study design, collection, analysis or interpretation of the data, writing the manuscript, or the decision to submit the paper for publication.

Correspondence should be addressed to Mel Simmonds-Buckley,

Department of Psychology, University of Sheffield, Cathedral Court, 1 Vicar Lane, Sheffield, S1 2LT, United Kingdom

Email:mksimmonds-buckley1@sheffield.ac.uk

Abstract

The evidence base for behavioral activation (BA) as a frontline treatment for depression is grounded in individual delivery. No valid previous meta-analytic reviews of BA delivered in groups have been conducted. This study therefore examined the efficacy and acceptability of group BA drawn from clinical trial evidence. Randomized controlled trials of group BA were identified using a comprehensive literature search. Depression outcomes at post-treatment/follow-up, recovery and drop-out rates were extracted and analyzed using a random-effects meta-analysis. Treatment moderators were analyzed using meta-regression and subgroup analyses. Nineteen trials were quantitatively synthesized. Depression outcomes post-group BA treatment were superior to controls (SMD 0.72, CI 0.34 to 1.10, k=13, N=461) and were equivalent to other active therapies (SMD 0.14, CI -0.18 to 0.46, k=15, N=526). Outcomes were maintained at follow-up for group BA and moderators of treatment outcome were limited. The drop-out rate for group BA (14%) was no different to other active treatments for depression (17%). Further research is required to refine the conditions for optimum delivery of group BA and define robust moderators and mediators of outcome. However, BA delivered in groups produces a moderate to large effect on depressive symptoms and should be considered an appropriate front-line treatment option.

Keywords: depression; behavioral activation; group delivery; meta-analysis; efficacy

Introduction

When a person is depressed, a widely observed symptom is behavioral avoidance, withdrawal and reduced activity, with these behavioral symptoms often contributing to the maintenance of low mood (Curran, Ekers, Mcmillan, & Houghton, 2012). Given this behavioral component, behavior change has long been a treatment target in the psychotherapy of depression. The initial treatment sessions of cognitive therapy for depression (Beck, Rush, Shaw, & Emery, 1979) focus on behavioral techniques (i.e., activity scheduling) in order to initially lift mood, with evidence of associated early change in depressive symptoms (Ilardi & Craighead, 1994). Purely behavioral treatments for depression that share core techniques around increasing activation and eliciting positive reinforcement have existed since the 1970's.

Treatments can be clustered under four models: Lewinsohn's pleasant events, focusing on increasing access to pleasant events through activity scheduling (Lewinsohn, Sullivan, & Grosscup, 1980); Rehm's self-control therapy, comprising three key elements of self-monitoring, self-evaluation and self-reinforcement (SCT; Rehm, 1984); Martell's contextual behavioral activation (BA), derived from the initial BA segment of Beck's cognitive behavioral therapy (CBT) for depression manual (Martell, Addis, & Jacobson, 2001); and Lejuez's behavioral activation treatment for depression (BATD; (Lejuez, Hopko, & Hopko, 2001). Early versions of BA applied relatively simple methods (e.g., pleasant events), whilst more recent developments of BA (e.g., contextual BA) are more complex. Core differences revolve around the activation approaches used to increase response-contingent positive reinforcement. SCT elaborates on the original pleasant events model by emphasizing the role of self-control in attenuating negative consequences of depression and using self-management skills to reinforce positive behavior change. BATD further expands on the pleasant events approach by relating goals to major life areas (relationships, hobbies etc.) and using activity hierarchies to focus on rewarding achievement of activity goals.

Contextual variants also incorporate values work, but have an additional emphasis on the function of avoidance and approach behaviors as a key strategy for overcoming depression (Kanter et al., 2010).

A central aspect of the BA evidence base is Jacobson's component study (Jacobson et al., 1996), as this emphasized that the cognitive elements of CBT were not necessary to achieve a good outcome with depressed patients. This evidence enabled BA to emerge as a stand-alone depression treatment (Martell et al., 2001). Subsequent BA outcome research has demonstrated that BA is an effective treatment, producing equivalent outcomes to CBT (Cuijpers, van Straten, & Warmerdam, 2007; Dimidjian et al., 2006; Ekers et al., 2014; Mazzucchelli, Kane, & Rees, 2009; Richards et al., 2016). A recent large-scale RCT found that the economic benefits of BA are also considerable, as non-inferior clinical outcomes in comparison to CBT were achieved at a 21% reduced cost (Richards et al., 2016). However, the evidence base for BA is primarily based on individual treatment, with much less focus on the acceptability and effectiveness of group BA delivery.

The importance of understanding the potential of BA as a group therapy relates to its delivery as well as its potential effects. BA works by adopting an 'outside-in' treatment approach, using pragmatic behavioral techniques to increase access to sources of positive reinforcement that in turn then reduce associated depressive thoughts and feelings (Curran et al., 2012). BA is therefore often characterized as a pragmatic and parsimonious treatment for depression (Jacobson et al., 1996). As fewer treatment competencies are required, therapists can be trained in a relatively short time (Ekers, Richards, McMillan, Bland, & Gilbody, 2011). The relative simplicity of BA also makes it well suited to group adaptation, as behavioral treatment principles can be easily taught, grasped and implemented (Dimidjian, Barrera, Martell, Muñoz, & Lewinsohn, 2011). Investigation of indirect comparisons of BA treatment mode have indicated individual and group delivery treatment effects do not differ significantly with group BA producing a moderate effect estimate ($g = 0.62$; Ekers et al.,

2014). During group treatment, patients can additionally benefit from the peer support, normalizing and the learning opportunities created by group dynamics (Yalom & Leszcz, 2005). Groups, if acceptable to patients, are also organizationally efficient, as they optimize scarce therapeutic resources through low therapist to patient ratios (Kellett, Clarke, & Matthews, 2007).

A meta-analysis of group-based BA effectiveness has been reported recently (Chan, Sun, Tam, Tsoi, & Wong, 2017), but had a broad raft of methodological problems. Only seven randomized controlled trials (RCTs) were identified, which does not represent the full evidence base of clinical trials of group BA (as will be seen below). Equally importantly, the seven studies included were actually individual BA (Carlbring et al., 2013; Dimidjian et al., 2006; Ekers et al., 2011; Gawrysiak, Nicholas, & Hopko, 2009; Hopko, Lejuez, LePage, Hopko, & McNeil, 2003; Moradveisi, Huibers, Renner, Arasteh, & Arntz, 2013; Pagoto et al., 2013). Finally, no mention of treatment acceptability issues was made. Any clinical conclusions concerning group BA drawn from the Chan et al. (2017) meta-analysis are therefore seriously flawed.

This meta-analysis therefore focuses on the acceptability and efficacy of group BA compared to standard treatment or waitlist controls and other active therapies and seeks to identify key moderators of outcome. Identifying treatment moderators helps to establish factors that account for variations in treatment effect (i.e. under what conditions and for which patients group BA is most effective). Potential moderators include intervention characteristics (such as type of BA model or number of sessions) and patient characteristics (such as population and depression severity). If differing BA models are not equally effective, it could suggest that different levels of treatment model complexity moderate outcome, and can indicate which models may be more suitable to group adaptation. With regards to amount of treatment, what is the optimum number of group BA sessions? Providing more treatment than required is wasteful of resources, whereas not providing

enough treatment risks creating a 'revolving door' for therapy services (Hansen, Lambert, & Forman, 2002). The dose-response literature suggests a negatively accelerated association between number of sessions and improved outcome, with estimates of 13-18 sessions required to achieve a 50% recovery rate (Hansen et al., 2002; Harnett, O'Donovan, & Lambert, 2010). However, BA has shown significant reductions in depression after much briefer periods of treatment (Armento, McNulty, & Hopko, 2012; Gawrysiak et al., 2009; Hopko, Robertson, & Carvalho, 2009). Meta-analytic investigations of the effectiveness of psychotherapy for depression has shown limited association with the number of sessions, advocating the implementation of briefer treatments (Cuijpers, Huibers, Daniel Ebert, Koole, & Andersson, 2013).

In terms of population-related moderators, which patients are most suitable for group BA? Establishing patient suitability is important as the acceptability of BA is based on assumed ease of application. It has been suggested that BA may provide a useful treatment option for varied and diverse patients, often from underrepresented patient populations (Dimidjian et al., 2011). Similarly, patients can present with differing severities of depression, but the differential effects of baseline severity on group BA treatment outcome are currently unclear. The previous consensus was that severely depressed patients tend to see better outcomes when treated with pharmacotherapy, whereas psychotherapy is indicated when treating mild to moderate depression (Elkin et al., 1995). Recently, this consensus has been questioned, as numerous studies have been unable to demonstrate baseline severity moderating treatment outcome (Driessen, Cuijpers, Hollon, & Dekker, 2010; Weitz et al., 2015). Thus, psychotherapy appears to be as an appropriate treatment for severe depression. BA appears particularly well suited for treating severe depressive phases, as the severely depressed patient may be unable to engage in cognitive work, or may indeed find the work a depressive trigger due to heightened guilt and self-blame (Dimidjian et al., 2006).

Ioannidis & Lau (1999) noted that the meta-analytic method was best employed when

summarizing, synthesizing and quantifying an evidence base that is made up of extant studies with high methodological quality. As randomized controlled trials (RCTs) champion internal as opposed to external validity, then RCTs ensure high methodological quality (Barkham, Stiles, Lambert, & Mellor-Clark, 2010). This meta-analysis therefore solely focuses on RCTs that have been conducted evaluating the effiacacy of group BA, to ensure that the quantititive synthesis was best on the best avaliable evidence. To summarize, this meta-analysis had three aims: (1) assess the efficacy of group BA when compared to passive and active controls, in terms of depression outcomes and recovery rates; (2) explore moderators of outcome in terms of intervention and patient variables; and (3) define the acceptability of group BA by calculating drop-out rates in comparison to passive and active controls.

Method

Identification and Selection of Studies

First, previous meta-analyses of BA were examined and cross-referenced to identify any group-based intervention studies. Second, a comprehensive electronic search was conducted, to identify literature published up until October 2016, which was modified for each of four databases used (MEDLINE, PsycINFO, Cochrane Library and CINAHL). Search terms (expanded using alternative synonyms, and both US and UK spellings) for (i) behavioral activation/therapy (including activity scheduling/pleasant events), (ii) depression and (iii) treatment efficacy were combined using a mixture of MeSH, title, abstract, keywords and text word searches. Filters to human and adult populations were applied (see Appendix A for search strategy). Third, reference lists of identified articles and previous BA reviews were manually searched to identify any additional studies. The primary reviewer (MSB) screened the initial title and abstracts and reviewed the full-texts of all identified studies. Uncertainty regarding study eligibility was debated with two other readers (SK and GW) to reach a consensus decision.

Inclusion Criteria

RCTs of group BA, with adults aged 18 and over with a depressive disorder or elevated symptoms of depression (assessed via a clinical screening interview or self-rated symptoms scored above a defined clinical cut-off on a standardized measure of depression). There was no limitation in terms of co-morbidity, as long as depression was a primary presenting problem. Studies containing child and adolescent participants, individuals with intellectual disability and participants with sub-clinical symptoms of depression were excluded. The methods of studies were analyzed, and the intervention was labelled BA if, and only if, the study delivered a purely behavioral treatment. Studies were labelled BA when the treatment focused on the functional analysis of behavior (in the absence of changing cognitions) and resultant behavioral change, in the pursuit of increasing positive mood. Therefore, mood-activity monitoring, activity scheduling and behavioral activation comprised the behavioral treatment components. The Mazzuchelli et al. (2009) BA treatment definitions were used for this review; i) pleasant events (Lewinsohn et al., 1980); ii) self-control (Rehm, 1984); iii) contextual (Martell et al., 2001) and iv) BATD (Lejuez et al., 2001). Minimum group size was defined as three or more participants in a group in a study. There was no limit on treatment duration or setting.

Comparators included any passive control, treatment as usual (TAU) or active treatment. Control comparators provided patients with a waitlist, TAU consisted of standard routine care in clinical practice settings, such as inpatient or Primary Care Physicians/General Practitioner care and active treatment comparators were other psychotherapies delivered in a therapeutic format that made an additional active attempt to improve depression, including cognitive therapy (CT), cognitive behavior therapy (CBT), problem-solving therapy, supportive therapy and non-specific psychotherapy. No language restrictions were applied, but a publicly available English language translation of the paper was an inclusion criteria. Unpublished studies and dissertations were included if available. Those studies that did not provide sufficient data to calculate effect sizes were excluded.

Outcome Measures

Primary outcome

The primary outcome measure was depressive symptomology measured by any psychometrically validated self-report or clinician-rated measure. A preferred measures hierarchy was used for studies that contained multiple depression outcome measures, so that a single effect size per comparison was calculated. Comparisons of self-report and clinician-rated measures demonstrate that clinician-rated outcomes generate larger effect sizes (Cuijpers, Li, Hofmann, & Andersson, 2010). Where studies used both self and clinician reported outcomes, self-reported outcomes took precedence in order to allow a more conservative estimate of treatment effect. The most commonly used self-report measure (i.e., BDI or BDI-II) was selected. When no self-report measure was available, clinician-rated measures were selected; the Hamilton Rating Scale for Depression (HRSD) took precedence.

Secondary outcomes

When available, information on drop-out and recovery rates was extracted as dichotomous data. Drop-out rates were used as a proxy for treatment acceptability. This was defined as the percentage of non-completers during group BA and control conditions. Non-completers were determined by the original study authors' definition. Recovery rates were the percentage of patients at end of treatment and/or follow-up who scored below the specified clinical threshold on the primary outcome measure. Recovery definition was determined by the original study authors' definition.

Quality Assessment

Methodological quality was assessed using the Cochrane Risk of Bias tool (Cochrane Collaboration, 2011). Due to difficulties blinding participants and personnel in psychotherapy trials, studies were only assessed on four of the risk of bias elements; randomized allocation, allocation concealment, blind outcome assessment and data attrition. Each element was rated for low, high or unclear risk of bias and each study given a score based on the number of

elements meeting criteria of low risk of bias (max score of four; higher scores indicating lower risk of bias). The primary author assessed all the studies and an independent rater assessed 50%. Inter-rater reliability was calculated using Cohen's kappa (Cohen, 1960) (where .21-.40 = fair agreement; .41-.60 = moderate agreement; .61-.80 = substantial agreement; .81-1.0 = almost perfect agreement; Landis & Koch, 1977). The kappa between the primary and independent rater was k=.73, indicating substantial agreement. Discrepancies in ratings were resolved through discussion to produce a final quality rating for each study.

Effect Sizes

Where data were available, outcomes for depression, recovery and drop-out rates were extracted at post-treatment and follow-up (8-weeks or the closest possible time point). Standardized mean differences (SMDs) and standard error (SE) terms were computed for the difference between conditions for each comparison between BA and a comparator condition. SMDs (Cohens d) were calculated by subtracting the mean post-treatment score of the comparator condition from the mean post-treatment score of the BA intervention and dividing the result by the pooled standard deviation (SD) of both conditions post-treatment. Due to the risk of small-sample bias, the J correction was applied to convert SMDs to Hedges g (Hedges & Olkin, 1985). Effect sizes were interpreted according to Cohen's criteria, where 0.2 is indicative of a small effect, 0.5 a moderate effect, and 0.8 a large effect (Cohen, 1992). Dichotomous data for recovery and drop-out rates were calculated as odds ratios (OR); (i.e. the percentage of recovery or drop-out from group BA in comparison to passive or active controls).

A hierarchical procedure was applied to effect size calculations - means and SDs were used wherever possible, followed by effect size data, dichotomous data, and finally t or F-scores. Controlled studies with sub-groups or multiple arms that were comparable were collapsed into one group using Cochrane's recommended method (Cochrane Collaboration, 2011). Studies with multiple comparators within one comparison that could not be collapsed

were included separately, with the number of participants in the shared intervention group split evenly across comparisons. For example, pair-wise comparisons of group BA with both CT and non-directive therapy from Shaw (1977) were both entered into the active therapy meta-analysis (means and SDs unchanged), with the number of patients who received group BA divided out equally between the two, to ensure patients were not included twice (Cochrane Collaboration, 2011).

Meta-analysis

Data were synthesized using Meta-Essentials (Suurmond, van Rhee, & Hak, 2017). Pooled effect sizes and 95% confidence intervals were computed using the inverse of the variance to weight the effect estimates (i.e., outcomes in favor of BAG were indicated by a positive effect size). Due to the expected level of heterogeneity resulting from different comparator types, a random-effects model was used to account for within- and between-study variance. Statistical significance was set at an alpha value of 0.05. Heterogeneity was investigated using the I^2 statistic to indicate percentage of variation and the accompanying Q statistic to report the statistical significance. Heterogeneity benchmarks (Higgins, Thompson, Deeks, & Altman, 2003) were used to identify low (25%), moderate (50%) and high study heterogeneity (75%). Pooled effect sizes were then converted into numbers needed to treat (NNT; Kraemer & Kupfer, 2006). NNT provides an estimate of the number of patients who would need to be treated by the group BA intervention to produce one additional beneficial outcome over a comparator condition.

Subgroup and Moderator Analysis

Sources of heterogeneity within comparisons were investigated using planned subgroup and moderator analyses. Subgroup analysis was used to investigate four categorical variables: control/therapy type (waitlist/TAU and CBT/other psychotherapy); assessment type (clinical interview/elevated symptoms above clinical cut-off); type of BA (pleasant events/self-control/contextual/BATD); and population (young adults/adults/older adults).

Meta-regression was used to investigate five continuous variables: study quality (0-4 risk of bias items); baseline depression (standardized Z-scores); gender (proportion of males); number of group sessions and group size. The beta-coefficient significance threshold was adjusted to $p < 0.01$ to account for multiple testing (Thompson & Higgins, 2002), and a minimum of 10 studies was required to investigate moderators within comparisons (Cochrane Collaboration, 2011).

Publication Bias

Where there were sufficient numbers of studies ($k > 10$), publication bias was assessed via visual inspection of asymmetry on a funnel plot of SEs against effect sizes. Additional statistical analysis of study distribution asymmetry was undertaken using the funnel plot regression method (Macaskill, Walter, & Irwig, 2001). Trim and Fill imputation of missing data gave an adjusted estimate effect, accounting for publication bias (Duval & Tweedie, 2000).

Results

Study Selection

After the removal of duplicates, searches identified 5335 records to be screened (Figure 1). Title and abstract screening identified 78 articles to be retrieved for full-text review. Upon review, 59 were excluded (reasons outlined in Figure 1) leaving a total of 20 studies meeting the inclusion criteria. One remaining study was identified as an outlier and excluded [20] from the quantitative synthesis. This was due to a very large effect size ($d = 5.76$) in favor of group BA compared to waitlist. Removal of this single study was conservative and favored the null hypothesis; this was deemed appropriate to reduce the risk of over-estimation of overall effect of BA.

Insert figure 1 here

Study details and quality ratings are available in Appendix B. Of the N=19 studies included, quality ranged from zero to three quality standards met (max four). Overall study

quality was poor. In particular, nearly all studies provided unclear descriptions of randomization and concealment procedures (see Appendix C for full quality ratings). Only one study was classed as high quality (met three or more quality criteria). Eight studies were deemed medium quality (met 1-2 quality criteria), while the remaining 11 studies were classed as low quality (met 0 quality criteria).

Meta-analysis of Group BA

Study characteristics

Nineteen studies were included across two meta-analytic comparisons. Group BA was compared to controls across 13 studies and active therapies in 12 studies across 15 comparisons. In the control comparisons, nine studies compared BA with a waitlist control and four used TAU. TAU consisted of inpatient (N=3) and outpatient (N=1) standard treatment, with varying levels of daily to weekly contact during the study period. In the active therapy comparisons CBT/CT was the most common comparison psychotherapy (N=5). The treatment comparators included supportive psychotherapy, psychodynamic psychotherapy, non-directive psychotherapy, problem-solving and assertiveness training. All comparator active therapies were delivered in a group format. Participants were recruited from the community (N=10), Universities (N=3) and clinical services (N=6; outpatient N=2, inpatient N=4). Depression was diagnosed via clinical interview (N=17) or self-report symptoms exceeding a depression measure clinical cut-off (N=2). Depression symptomology was assessed via self-report (N=10), clinician report (N=1), or a combination (N=8). The most commonly employed self-report outcome measure was the BDI or BDI-II (N=15), and the most commonly employed clinician-rated outcome measure was the HRSD (N=7). Follow-up duration ranged from 2-32 weeks across N=13 studies. The mean follow-up period was 6 weeks.

BA group studies were conducted on adults in the general population (N=14), students (N=3) and older adults (N=2). Mean depression severity at intake ranged between

mild (N=6), moderate (N=8) and severe (N=4). One study did not report sufficient information to establish baseline severity. Three studies focused on treating a primary problem of depression in conjunction with co-morbid disorders (substance abuse and anxiety). BA treatment type included pleasant events (N=8), self-control (N=6), contextual (N=2) and BATD (N=3). Group sizes ranged from 3-10 participants with a mean of seven, treatment duration ranged from 2-12 sessions, with session duration ranging from 30-120 minutes. Drop-out rates ranged between 0-33% but were unreported in nine studies. Recovery rates ranged from 25-100% but were unreported in 12 studies. Recovery was defined by use of clinical cut-offs on measures (N=5) and MDD diagnosis (N=2). Intent-to-treat analysis was used in N=4 studies, with the remaining 15 studies using completers analyses.

Comparison 1: Group BA versus waitlist/TAU control comparators

Depression at post-treatment; group BA versus waitlist/TAU

Post-treatment outcomes from 13 studies contributed to this analysis, totaling N=461 participants (group BA N=244; control N=217). The overall aggregated SMD was 0.72 (95% CI 0.34 to 1.10; Z = 4.15; p<0.0001) in favor of group BA, suggesting a significant moderate to large effect (Figure 2). Group BA was effective at reducing depressive symptoms at treatment completion, when compared to waitlist and TAU controls. The NNT for group BA was 2.57; one out of every three patients experiences additional benefit from group BA when compared to controls at treatment completion. There was significant between-study heterogeneity contributing to moderate variation in effect (I^2 = 58%; Q = 28.72, p=0.004).

Insert figure 2 and table 1 here

Subgroup analysis and meta-regression results are displayed in Table 1. Significant variation in effect size was associated with type of control condition. A large effect was observed for waitlist controls, but the effect for group BA was small and non-significant when compared to TAU. Treatment effects were not significantly affected by assessment

method, type of BA or sample population. Moderate heterogeneity was evident in the majority of sub-groups. Although not significant, moderating effects of study quality were in the direction of more favorable effects for group BA in lower quality studies. Meta-regression analyses found initial depression severity, gender, number of sessions and group size were not associated with improved treatment outcomes.

Funnel plot inspection gave a slight suggestion of asymmetry (see Appendix D). This indicates that smaller studies may have tended to produce larger effects in favor of group BA. The adjusted effect size produced by Trim and Fill imputation of missing data produced a slightly smaller moderate effect size (0.65, 95% CI 0.25 to 1.05). Testing the extent of asymmetry via funnel plot regression showed sufficient symmetry of study distribution (B = -0.004, t(11) = -0.76, p=0.47).

Depression at follow-up; group BA versus waitlist/TAU

Four studies (waitlist k=1; TAU k=3) had follow-up comparisons with a total of N=129 participants (group BA N=64; control N=65). There was a moderate pooled SMD of 0.69 (95% CI 0.19 to 1.19; Z = 4.42; p<0.0001) in favor of the maintained effects of group BA at follow-up (Figure 2). Group BA therefore appeared effective at sustaining improvement at follow-up compared to controls. The NNT was 2.67, indicating that at follow-up one out of every three participants experienced additional benefit from group BA compared to controls. Studies were statistically homogeneous (I^2 = 0%; Q = 2.25, p=0.52), even when taking a higher significance level threshold (p<0.1) to account for low power from the small number of studies. Limited variance between studies negated the need for further heterogeneity analysis. There were an inadequate number of studies (k<10) to test for publication bias.

Recovery and drop-out rates; group BA versus waitlist/TAU

Two studies (waitlist k=2) reported recovery rates for 118 participants (group BA N=64; control N=54). Recovery rates were significantly higher following group BA than

waitlists (group BA 52%, control 28%), producing a significant odds ratio of 2.99 (95% CI 0.20 to 43.86; Z = 5.17; p = <0.001). More participants recovered after receiving group BA than those allocated to a waitlist condition. All studies were statistically homogeneous (I^2 = 0%; Q = 0.25, p = 0.62).

Five studies (waitlist k=4; TAU k=1) reported drop-out rates for 325 participants (group BA N=185; control N=140). There was no difference in drop-out rates between group BA (15%) versus control conditions (17%), with a non-significant odds ratio of 0.69 (95% CI 0.21 to 2.29; Z = 0.86; p = 0.20). Patient drop-out rates were matched across group BA (15%), waitlist (18%) and TAU (14%). Between-study variance was minimal and not significant (I^2 = 24%; Q = 5.26, p = 0.26). Limited heterogeneity and the small number of studies reporting recovery and drop-out outcomes constrained further investigation into sources of variation in effect sizes. The number of studies of group BA reporting recovery and dropout rates were insufficient to perform any publication bias tests.

Comparison 2: Group BA versus other active psychotherapies

Depression at post-treatment in group BA versus other active psychotherapies

Post-treatment outcomes from 15 comparisons contributed to this analysis, totaling N=526 participants (group BA N=254; active psychotherapies N=272). There was no difference in the effect of group BA when compared to other psychotherapies, with a non-significant SMD of 0.14, tending towards being in favor of group BA (95% CI -0.18 to 0.46; Z = 0.87; p = 0.38) (Figure 3). Group BA was as effective at reducing depressive symptoms as other active psychotherapies. The NNT for group BA was 12.68. This indicates one out of every 13 participants would experience additional benefit post-treatment from being in a group BA treatment, when compared to other psychotherapies. Between-study heterogeneity was moderate and significant (I^2 = 63%; Q = 38.22, p=0.0005).

Insert figure 3 and table 2 here

Further investigation into variations in effect estimate is displayed in Table 2.

Subgroup analyses of different psychotherapies found that group BA compared to CBT/CT therapies resulted in a minimal non-significant effect. When compared to other psychotherapies, group BA resulted in a small (non-significant) effect that leant towards favoring it as a treatment. Significantly differing effect sizes were not evident when comparing different types of BA or the sample populations (all studies used clinical interviews, so assessment type was not assessed as a moderator). There was moderate heterogeneity present in most of the subgroups. Meta-regression analyses found limited evidence of variation in effect sizes according to study quality, initial depression severity, gender, number of sessions or group size.

Funnel plot inspection did not suggest evidence of asymmetry (see Appendix D), with funnel plot regression providing evidence of a symmetrical study distribution (B = 0.005, t(14) = 1.09, p=0.30). Trim and Fill imputation estimated one study was missing and produced an adjusted overall effect estimate of 0.21 (95% CI -0.18 to 0.61), representing a slight increase in favor of group BA, albeit still not reaching significance. The removal of the smallest studies reduced the overall effect estimate to 0.08 (95% CI -0.33 to 0.50), indicating minimal influence of a small study effect. These observations indicate a minimal effect of publication bias and suggest the effect estimate appears reasonably robust.

Depression at follow-up in group BA versus other active psychotherapies

Eight studies performed 10 follow-up comparisons (CBT/CT k=5; other therapy k=5) with a total of 240 participants (group BA N=122; active psychotherapies N=118). There was a small SMD of 0.32 favoring group BA (see Figure 3), but this was not significant (95% CI -0.10 to 0.74; Z = 1.50; p = 0.13). Group BA and the other active psychotherapies therefore produced similar maintained treatment effects at follow-up. The NNT was 6.16, indicating that by follow-up one out of every six patients experienced additional benefit from group BA. Significant between-study heterogeneity was observed representing a moderate level of variance (I^2 = 57%; Q = 21.00, p=0.01). Five comparisons of group BA versus CBT/CT

produced similar effects at follow-up (SMD = 0.07; 95% CI -0.41 to 0.55; Z = 0.27; p = 0.78). BA was compared to other psychotherapies in the remaining five studies at follow-up and showed a moderate (but non-significant) effect in favor of group BA (SMD = 0.59; 95% CI -0.09 to 1.69; Z = 0.27; p = 0.09). The small number of studies prevented any further exploration of moderating variables and publication bias.

Recovery and drop-out rates during group BA versus other active psychotherapies

Seven studies with nine comparisons (CBT/CT k=4; other k=5) reported recovery rates for 351 participants (group BA N=169; other psychotherapies N=182). There was no difference in recovery rates following group BA compared to other psychotherapies (69% during group BA versus 61% during other active psychotherapies) with a non-significant odds ratio of 1.30 (95% CI 0.41 to 4.07; Z = 0.44; p = 0.66). The recovery rate for group BA was comparable to that of other active psychotherapies. Group BA versus CBT/CT had a non-significant OR of 0.39 in favor of CBT/CT (95% CI 0.04 to 4.15; Z = 0.77; p = 0.44). Group BA versus all other therapies had a non-significant OR of 2.72 in favor of group BA (95% CI 0.83 to 8.85; Z = 1.66; p = 0.10). The studies were significantly heterogeneous (I^2 = 61%; Q = 20.42, p=0.009), but there were insufficient studies to examine moderators of variation in effect size or to test publication bias.

Seven studies (CBT/CT k=1; other therapy k=6) reported drop-out rates for 370 participants (group BA N=206; other psychotherapies N=164). There was no difference between drop-out rates during group BA (14%) versus other psychotherapies (17%), with a non-significant odds ratio of 0.71 (95% CI 0.37 to 1.34; Z = 1.06; p = 0.29). Between-study heterogeneity was minimal and non-significant (I^2 = 0%; Q = 5.25, p = 0.51). Subgroup analysis of type of psychotherapy (CBT/CT or other psychotherapies) did not result in significantly different drop-out rates (CBT/CT OR = 0.62; other psychotherapy OR = 71; p = 0.89). Further moderator analysis and tests of publication bias were not conducted, due to insufficient number of studies.

Discussion

The objective of this meta-analysis was to quantify the acceptability and efficacy of BA when delivered in groups to treat depression and explore key potential moderators of outcome. To achieve this objective, only RCTs were selected and this enabled a comparison to be made with both passive and active controls. This analysis was conducted in order to provide guidance to commissioners and clinicians in terms of offering evidence-based treatments for depression. Particularly, this meta-analysis also has provided the first scientifically credible quantitative review of the evidence base for group BA, in contrast to the review conducted by Chan et al. (2017).

Summary of group BA outcomes

In relation to the first aim, the results provide support for the effectiveness of group BA in the treatment of depression across trial contexts. Compared to waitlist comparators, group BA facilitated significantly reduced depressive symptoms at treatment completion and at follow-up, improved recovery rates and equivalent drop-out rates. One out of every three participants would expect to experience additional benefit from receiving group BA, when compared to waitlist. When solely compared to TAU, group BA did not add any additional benefit, with no significant differences in post-treatment outcomes. Compared to other routinely used psychotherapies for depression (including CBT), group BA produced equivalent outcomes at treatment completion and at follow-up, with matched recovery and dropout rates. The results therefore indicate that group BA offers an acceptable, equivalent and useful treatment option in the treatment of depression, both in the short and medium-term.

The moderate to large effects in the reduction of depressive symptoms and increased clinical recovery rates suggests that BA principles translate well into group format settings. The translation of BA theory to group delivery supports the notion that the principles of BA remain simple and parsimonious to deliver, regardless of context (Jacobson et al., 1996). The

magnitude of the group BA treatment effect compared to controls is similar to the effect observed (SMD 0.70-0.87) for individually-delivered BA (Cuijpers et al., 2007; Ekers et al., 2014; Mazzucchelli et al., 2009) and slightly larger than the Ekers et al. delivery format moderator estimate (Ekers et al., 2014). Likewise, the group BA treatment effect is comparable to the individual BA versus other treatments effect (SMD 0.13; Cuijpers et al., 2007). Furthermore, benefits of group BA were still evident at follow-up, suggesting durability of outcomes for this behavioral intervention. Therefore, it is reasonable to conclude that allocating to group BA is not detrimental to patient outcome, and that participants are as likely to engage in group treatment as individual work.

The lack of significantly different group BA outcomes compared to TAU is in contrast to effects seen for individually delivered BA (Ekers et al., 2014) and suggests TAU had a comparatively potent effect in the available studies. All but one of the present TAU studies were conducted in inpatient settings, so it may be that features of inpatient routine care bear similarities with active treatments and provide sufficient potency that is not improved on by group BA. Interestingly, while post-treatment outcomes did not support an added benefit of group BA over TAU, the significant follow-up effects of group BA versus controls were driven by comparisons with TAU. Although follow-up only comprised four studies, a similar pattern (although not quite significant) was seen for post-treatment to follow-up outcomes versus non-CBT therapies. It implies that group BA's advantage over these types of treatments may be in providing more durable beneficial effects in a format that is simpler to disseminate.

Moderators of group BA effectiveness

Analysis of the variation between studies enabled investigation of moderators of group BA effectiveness in order to explore factors that contribute to the treatment effect. Whilst such moderator analyses highlight the magnitude of treatment effect associated with certain patients, treatments and methodological factors, they do not infer causality (Cochrane

Collaboration, 2011). In addition, interpretation needs to be undertaken with caution, as some subgroup arms only had a small number of studies and the high correlation of some variables (e.g., TAU and inpatient settings/BA types) potentially produces unreliable and confounded observed effects.

Group BA was used in studies with a range of participants and varied clinical presentations, and the treatment effect when compared to controls or active therapies was not related to gender, initial depression severity, assessment method or population. The finding that there was no association between the size of treatment effect and initial depression severity is in line with extant evidence (Driessen et al., 2010; Weitz et al., 2015), and contradicts original conclusions that psychotherapy effects are larger for less severe depression (Elkin et al., 1995). The current results imply that, regardless of baseline severity of depression, participants can experience benefit from group BA. Behavioral techniques are easily grasped and implemented by patients, even when (for example) cognitive functioning is impaired during depressed episodes (Lam, Kennedy, McIntyre, & Khullar, 2014). Differences between age population subgroups were not significant, but two of the subgroup arms were very small for control and the active psychotherapy comparisons. Inspection of the size of the effects suggested some variation; group BA was very effective for young adults and adults (versus controls), but much less effective in older adults. It may be the case that BA in groups with older adult participants needs to have relevant treatment adaptations applied, in order to retain clinical effectiveness (Pasterfield et al., 2014).

Various treatment delivery factors (group size, type of BA or number of sessions) were not associated with differences in effectiveness, when compared against controls or active therapy comparisons. Again, statistical interpretation may have been hampered by confounding variables and insufficient comparisons in the subgroup arms for types of BA. Non-significant variation in effect sizes for different types of BA was evident - simpler versions seem to produce the largest treatment effects, but without being statistically

superior. However, the majority of the simpler, older protocols were compared to waitlist controls, while the newer, more complex protocols were compared to TAU. The lack of a definitive advantage of one version of BA highlights that the behavioral treatment model will need further refining and testing to determine the optimal conditions for group delivery.

Number of sessions was not significantly associated with the size of the treatment effect - increasing the number of group sessions did not produce better outcomes. This finding is in line with Cuijpers et al.'s meta-regression analysis (2013) and supports the argument that group BA interventions only need to be brief. Control type did produce differences in treatment effects; waitlist comparisons resulted in a large effect, but TAU comparisons only had a small beneficial effect in favor of group BA. Similar effects have been seen for other types of psychotherapy (Cuijpers, Van Straten, Bohlmeijer, Hollon, & Andersson, 2010; Cuijpers et al., 2013) and highlight the importance of the type of comparator in determining a relevant estimate of effect.

Acceptability

The low drop-out rate for group BA found in this study (14%) implies BA delivered in a group can be well tolerated by patients. A meta-analysis of dropout from one-to-one treatment for major depression found an overall weighted dropout rate of 20% (Cooper & Conklin, 2015). Treatment completion is fundamental to ensure the full benefit of treatment is received which is especially pertinent as early termination of psychotherapy is related to poorer outcomes (Cahill et al., 2003; Hansen et al., 2002). Any claims of the organizational efficiency benefits of group delivery are offset if group attendance is poor. However, the dropout rates observed for group BA in comparison to the active controls (17%) suggests that group delivery does not suppress attendance. The equivalence of the drop-out rates recorded supports the notion that BA in a group format is an acceptable treatment and mirrors meta-analytic findings for individual BA (Ekers, Richards, & Gilbody, 2008).

Clinical and organizational implications

Access to clinically effective group interventions generates a range of organizational benefits, in relation to efficient use of facilities, high therapist to patient ratios and potential reductions to treatment wait-times (Piper, 2008). Recent evidence (Richards et al., 2016) also noted the health economic advantage of BA when delivered on a one-to-one basis. Demand for psychotherapeutic treatment for depression is consistently high, and services can struggle to meet this demand whilst simultaneously ensuring high quality care (Kazdin & Blase, 2011). Frontline depression treatments in clinical services should balance the evidence of clinical effectiveness with issues relating to ease of access, acceptability and efficient use of scarce resources (i.e., balancing both effectiveness and reach). When evaluating a treatment, it is also recommended that it should be compared to the current gold-standard treatment (David, Cristea, & Hofmann, 2018). Compared to CBT, BA has an advantage of a potentially simpler, shorter training for therapists (or even non-specialists; Ekers et al, 2011). This advantage may be particularly relevant in low-income countries, where depression contributes highly to the burden of disease but mental health resources are extremely limited (Patel, 2012; Richards et al., 2016).

There were no differences in subgroup clinical outcomes or drop-out rates when group BA was compared to group CBT (or CT variations) at post-treatment and follow-up. As originally highlighted by Jacobson et al. (1996), this meta-analysis echoes that therapy focused on changing depressogenic cognitions directly might be therapeutically redundant during the treatment of depression. In fact, the comparability of group BA and all other active psychotherapy outcomes is consistent with a large body of evidence that suggests all therapies are as effective as each other (Cuijpers, 2017). Such findings point to common factors shared between therapies producing the treatment benefits (such as therapeutic relationship, demand characteristics), rather than the protocol-specific techniques (Wampold, 2015). If this is the case, it raises questions about CBT as the gold-standard treatment for depression. CBT is recommended as the best treatment for depression (National Institute for

Health and Clinical Excellence [NICE], 2016), although the evidence does not always support that CBT provides treatment effects above and beyond other treatments. In light of the potential dissemination and economic advantages of BA over CBT, conducting a non-inferiority meta-analysis would be a valuable next step.

The treatment effect estimates produced by this meta-analysis are based on RCT evidence, but to what degree do these findings translate into real-world settings? Whilst testing the efficacy of group BA using RCTs is of primary importance, it does not necessarily indicate how effective such group therapy is when delivered in naturalistic settings (Rothwell, 2005). The internally valid conditions of an RCT (e.g. patient exclusion, therapist supervision and treatment fidelity) differ widely from the externally valid conditions of routine practice (e.g., the comorbidity of typical patient populations; Seligman, 1995). Whilst some evidence suggests the outcomes achieved during routine practice are comparable to RCTs (Gibbons et al., 2010; Westbrook & Kirk, 2005), others have found inferior outcomes for naturalistic settings (Barkham et al., 2008; Schindler & Hiller, 2010). Whether the outcomes recorded here can be replicated in routine practice is currently unclear.

Limitations

There is a range of limitations to consider for this meta-analysis. One reviewer screened and extracted all the data, which could introduce potential bias in the data. The number of BA group studies was limited, with the majority of studies also having relatively small sample sizes (Turner, Bird, & Higgins, 2013). For primary outcomes, the number of comparisons was suboptimal for most subgroup analyses of post-treatment outcomes and as discussed above, the resulting moderator interpretations were somewhat restricted. Even fewer studies conducted follow-up depression assessments. The follow-up periods that were reported were generally short and so were too brief to provide a truly valid assessment of the durability of group BA. The measurement period for follow-up assessments were typically between 4-12 weeks and this should be increased to at least one year in future group BA

outcome research. As depression has a chronically relapsing nature, whether the effects of group BA compared to controls or active therapies can be retained in the long-term is still unclear (Steinert, Hofmann, Kruse, & Leichsenring, 2014). Longitudinal tracking of outcomes following group BA, relapse rates and any need for further intervention (e.g., behavioral 'top-up' sessions) would supplement the durability evidence base for group BA. Recovery and drop-out data were not widely reported, meaning investigations of moderators and publication bias were not possible for those outcomes. Future group BA outcome studies should report core information on recovery and drop-out rates as standard and also report average session attendance. In terms of future controlled research, then a randomized patient preference trial (Howard & Thornicroft, 2006) directly comparing individual versus group BA would strengthen the evidence base as indirect comparisons can be confounded by factors unrelated to the treatment effect (e.g., different sample populations, comparisons of effects versus differing levels of control group rigor; Song et al., 2009).

The treatment effect reported for group BA in this meta-analysis may be subject to risk of some over-estimation and imprecision. First, study quality was poor across all studies with only one study deemed to have a low risk of bias. The effect of study quality was not significant, but the lack of variation in study quality meant sub-group analysis had low power. In general, the moderating effect of study quality was in the direction of lower quality studies producing larger effects in favor of BA. Therefore, the degree of sub-optimal study quality may have contributed to an overstated overall treatment effect. It should be noted that the Cochrane risk of bias tool was used to aid comparability and consistency of Cochrane recommended methods, but it may not be the optimal tool to reflect quality issues in psychotherapy research. Use of a quality tool designed specifically for psychotherapy trials, such as the Randomized Controlled Trials of Psychotherapy Quality Rating Scale (RCT-PQRS; Kocsis et al., 2010) may be better suited to capture the most relevant validity factors.

Second, very few studies analyzed outcomes using the intention-to-treat method and

observed effects were mostly based on per protocol analyses. Such 'completer samples' are again at risk of overestimating treatment effects (Heritier, Gebski, & Keech, 2003). Third, the distribution of comparator types across studies was not ideal. With regards to control comparisons, the majority were waitlist conditions. Waitlist controls are prone to overestimating treatment effects in active comparators (Cuijpers et al., 2010). The large difference in the group BA treatment effect compared to waitlists and TAU potentially reflects an overstated waitlist effect. It was also noted that the reporting of what TAU entailed was often vague, which may have contributed to the heterogeneity detected in the studies and makes generalizability of the effect of TAU and group BA similarly difficult to interpret. During the active therapy comparisons, the types of other psychotherapies were very varied, which might have diluted their effect in comparison to group BA. Only CBT or CT treatments were compared in enough studies to allow comparisons by treatment type. However, as CBT is the frontline treatment for depression, this allowed subgroup comparison of group BA with the current gold-standard (David et al., 2018). Fourth, significant variation was evident across BA clinical trials indicating moderate heterogeneity amongst studies, not accounted for by the use of a random-effects model or moderator effects. Results give an indication of the effectiveness of group BA, but the variability increases the statistical imprecision of the effect estimate. Finally, fewer than half the included studies included a treatment integrity check. This means that group BA might not have been delivered in a protocol-adherent way.

Future research directions

This evidence shows that group BA is an effective treatment. However, there is no single version of BA. Direct comparisons in clinical trials of the different versions of group BA are needed to establish the most effective behavioral approach. BA is promoted for its simplicity – therefore, adding complexity or extending treatment without improving outcomes is counterintuitive and needs testing if it is to be justified. Hence, the focus going

forward in the group BA evidence base should be on identifying the most clinically effective and organizationally efficient model for BA to be delivered in a group setting and subsequent implementation into routine practice. This research could also embed longitudinal measures in the method, to allow analysis of what mediates the relationship between BA and outcome. Similarly, the suggestion regarding older adults having a poor response to group BA indicates that moderators of group BA outcomes (e.g., age) need further investigation. The moderating effect of homework compliance on treatment outcomes in relation to other therapies was restricted due to lack of data in the present review. Given the crucial link to BA outcome, this is an area that would merit additional research.

Conclusion

This review provides support for BA as a standalone treatment for depression, but has shown for the first time that a group delivery format can be adopted with confidence. Group BA appears to work across a broad population of participants, regardless of depression severity. Furthermore, group BA appears as clinically effective and acceptable as CBT, the frontline treatment for depression (NICE, 2016). In light of the high and increasing demand for depression treatment, BA should be considered as a frontline intervention, on a par with CBT. Future research should focus on establishing the optimal delivery, mediators, moderators and long-terms effects of group BA, based on high quality efficacy studies and assess the degree to which outcomes then translate in routine practice settings.

References

Armento, M. E. A., McNulty, J. K., & Hopko, D. R. (2012). Behavioral activation of religious behaviors (BARB): Randomized trial with depressed college students. Psychology of Religion and Spirituality, 4, 206–222. https://doi.org/10.1037/a0026405

Barkham, M., Stiles, W. B., Connell, J., Twigg, E., Leach, C., Lucock, M., … Angus, L. (2008). Effects of psychological therapies in randomized trials and practice-based studies. British Journal of Clinical Psychology, 47, 397–415. https://doi.org/10.1348/014466508X311713

Barkham, M., Stiles, W. B., Lambert, M. J., & Mellor-Clark, J. (2010). Building a Rigorous and Relevant Knowledge Base for the Psychological Therapies. In Developing and Delivering Practice-Based Evidence: A Guide for the Psychological Therapies (pp. 21–61). Chichester, UK: John Wiley & Sons Ltd. https://doi.org/10.1002/9780470687994.ch2

Beck, A. T., Rush, J., Shaw, B. F., & Emery, G. (1979). Cognitive therapy of depression. New York: Guilford Press.

Cahill, J., Barkham, M., Hardy, G., Rees, A., Shapiro, D. A., Stiles, W. B., & Macaskill, N. (2003). Outcomes of patients completing and not completing cognitive therapy for depression. British Journal of Clinical Psychology, 42, 133–143. https://doi.org/10.1348/014466503321903553

Carlbring, P., Hägglund, M., Luthström, A., Dahlin, M., Kadowaki, Å., Vernmark, K., & Andersson, G. (2013). Internet-based behavioral activation and acceptance-based treatment for depression: A randomized controlled trial. Journal of Affective Disorders, 148, 331–337. https://doi.org/10.1016/J.JAD.2012.12.020

Chan, A. T. Y., Sun, G. Y. Y., Tam, W. W. S., Tsoi, K. K. F., & Wong, S. Y. S. (2017). The effectiveness of group-based behavioral activation in the treatment of depression: An updated meta-analysis of randomized controlled trial. Journal of Affective Disorders,

208, 345–354. https://doi.org/10.1016/j.jad.2016.08.026

Cochrane Collaboration. (2011). Cochrane Handbook for Systematic Reviews of Interventions. In J. P. T. Higgins & S. Green (Eds.) (version 5). Available from www.handbook.cochrane.org. Retrieved from www.handbook.cochrane.org.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37–46. https://doi.org/10.1177/001316446002000104

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cooper, A. A., & Conklin, L. R. (2015). Dropout from individual psychotherapy for major depression: A meta-analysis of randomized clinical trials. Clinical Psychology Review, 40, 57–65. https://doi.org/10.1016/j.cpr.2015.05.001

Cuijpers, P. (2017). Four decades of outcome research on psychotherapies for adult depression: An overview of a series of meta-analyses. Canadian Psychology, 58, 7–19. https://doi.org/10.1037/cap0000096

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, 58, 376–385. https://doi.org/10.1177/070674371305800702

Cuijpers, P., Huibers, M., Daniel Ebert, D., Koole, S. L., & Andersson, G. (2013). How much psychotherapy is needed to treat depression? A metaregression analysis. Journal of Affective Disorders, 149, 1–13. https://doi.org/10.1016/j.jad.2013.02.030

Cuijpers, P., Li, J., Hofmann, S. G., & Andersson, G. (2010). Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: A meta-analysis. Clinical Psychology Review, 30, 768–778. https://doi.org/10.1016/j.cpr.2010.06.001

Cuijpers, P., Van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The

effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. Psychological Medicine, 40, 211–223. https://doi.org/10.1017/S0033291709006114

Cuijpers, P., van Straten, A., & Warmerdam, L. (2007). Behavioral activation treatments of depression: A meta-analysis. Clinical Psychology Review, 27, 318–326. https://doi.org/10.1016/j.cpr.2006.11.001

Curran, J., Ekers, D., Mcmillan, D., & Houghton, S. (2012). Behavioural Activation. In W. Dryden (Ed.), Cognitive Behaviour Therapies (pp. 236–260). London, UK: Sage Publications Ltd. https://doi.org/10.4135/9781446288368.n11

David, D., Cristea, I., & Hofmann, S. G. (2018). Why cognitive behavioral therapy is the current gold standard of psychotherapy. Frontiers in Psychiatry, 9, 4. https://doi.org/10.3389/fpsyt.2018.00004

Dimidjian, S., Barrera, M., Martell, C., Muñoz, R. F., & Lewinsohn, P. M. (2011). The origins and current status of behavioral activation treatments for depression. Annual Review of Clinical Psychology, 7, 1–38. https://doi.org/10.1146/annurev-clinpsy-032210-104535

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, 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. https://doi.org/10.1037/0022-006X.74.4.658

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. https://doi.org/10.1037/a0020570

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel- plot–based method of testing

and adjusting for publication bias in meta-analysis. Biometrics, 56, 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

Ekers, D., Richards, D., & Gilbody, S. (2008). A meta-analysis of randomized trials of behavioural treatment of depression. Psychological Medicine, 38, 611–623. https://doi.org/10.1017/S0033291707001614

Ekers, D., Richards, D., McMillan, D., Bland, J. M., & Gilbody, S. (2011). Behavioural activation delivered by the non-specialist: Phase II randomised controlled trial. British Journal of Psychiatry, 198, 66–72. https://doi.org/10.1192/bjp.bp.110.079111

Ekers, D., Webster, L., Van Straten, A., Cuijpers, P., Richards, D., & Gilbody, S. (2014). Behavioural activation for depression; An update of meta-analysis of effectiveness and sub group analysis. PLoS ONE, 9, e100100. https://doi.org/10.1371/journal.pone.0100100

Elkin, I., Gibbons, R. D., Shea, M. T., Sotsky, S. M., Watkins, J. T., Pilkonis, P. A., & Hedeker, D. (1995). Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. Journal of Consulting and Clinical Psychology, 63, 841–847. https://doi.org/10.1037/0022-006X.63.5.841

Gawrysiak, M., Nicholas, C., & Hopko, D. R. (2009). Behavioral activation for moderately depressed university students: Randomized controlled trial. Journal of Counseling Psychology, 56, 468–475. https://doi.org/10.1037/a0016383

Gibbons, C. J., Fournier, J. C., Stirman, S. W., Derubeis, R. J., Crits-Christoph, P., & Beck, A. T. (2010). The clinical effectiveness of cognitive therapy for depression in an outpatient clinic. Journal of Affective Disorders, 125, 169–176. https://doi.org/10.1016/j.jad.2009.12.030

Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. Clinical Psychology: Science

and Practice, 329–343. https://doi.org//10.1093/clipsy.9.3.329

Harnett, P., O'Donovan, A., & Lambert, M. J. (2010). The dose response relationship in psychotherapy: Implications for social policy. Clinical Psychologist, 14, 39–44. https://doi.org/10.1080/13284207.2010.500309

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. New York, NY: Academic Press. https://doi.org/https://doi.org/10.1016/C2009-0-03396-0

Heritier, S. R., Gebski, V. J., & Keech, A. C. (2003). Inclusion of patients in clinical trial analysis: The intention-to-treat principle. Medical Journal of Australia, 179, 438–440. https://doi.org/her10586_fm [pii]

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. British Medical Journal, 327, 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hopko, D. R., Lejuez, C. W., LePage, J. P., Hopko, S. D., & McNeil, D. W. (2003). A brief behavioral activation treatment for depression: A randomized pilot trial within an inpatient psychiatric hospital. Behavior Modification, 27, 458–469. https://doi.org/10.1177/0145445503255489

Hopko, D. R., Robertson, S. M. C., & Carvalho, J. P. (2009). Sudden gains in depressed cancer patients treated with behavioral activation therapy. Behavior Therapy, 40, 346–356. https://doi.org/10.1016/J.BETH.2008.09.001

Howard, L., & Thornicroft, G. (2006). Patient preference randomised controlled trials in mental health research. The British Journal of Psychiatry, 188, 303–304. https://doi.org/https://doi.org/10.1192/bjp.188.4.303

Ilardi, S. S., & Craighead, W. E. (1994). The role of nonspecific factors in cognitive-behavior therapy for depression. Clinical Psychology: Science and Practice, 1, 138–155. https://doi.org/10.1111/j.1468-2850.1994.tb00016.x

Ioannidis, J. P., & Lau, J. (1999). Pooling research results: Benefits and limitations of meta-

analysis. The Joint Commission Journal on Quality Improvement, 25, 462–469. https://doi.org/10.1016/S1070-3241(16)30460-6

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. https://doi.org/10.1037//0022-006X.64.2.295

Kanter, J. W., Manos, R. C., Bowe, W. M., Baruch, D. E., Busch, A. M., & Rusch, L. C. (2010). What is behavioral activation?. A review of the empirical literature. Clinical Psychology Review, 30, 608–620. https://doi.org/10.1016/j.cpr.2010.04.001

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

Kellett, S., Clarke, S., & Matthews, L. (2007). Delivering group psychoeducational CBT in Primary Care: Comparing outcomes with individual CBT and individual psychodynamic-interpersonal psychotherapy. British Journal of Clinical Psychology, 46, 211–222. https://doi.org/10.1348/014466506X146188

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. https://doi.org/10.1016/j.comppsych.2009.07.001

Kraemer, H. C., & Kupfer, D. J. (2006). Size of treatment effects and their importance to clinical research and practice. Biological Psychiatry, 59, 990–996. https://doi.org/10.1016/J.BIOPSYCH.2005.09.014

Lam, R. W., Kennedy, S. H., McIntyre, R. S., & Khullar, A. (2014). Cognitive dysfunction in major depressive disorder: Effects on psychosocial functioning and implications for treatment. Canadian Journal of Psychiatry, 59, 649–654.

https://doi.org/10.1177/070674371405901206

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159. https://doi.org/10.2307/2529310

Lejuez, C. W., Hopko, D. R., & Hopko, S. D. (2001). A brief behavioral activation treatment for depression: Treatment manual. Behavior Modification, 25, 255–286. https://doi.org/10.1177/0145445501252005

Lewinsohn, P. M., Sullivan, J. M., & Grosscup, S. J. (1980). Changing reinforcing events: An approach to the treatment of depression. Psychotherapy: Theory, Research & Practice, 17, 322–334. https://doi.org/10.1037/h0085929

Macaskill, P., Walter, S. D., & Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. Statistics in Medicine, 20, 641–654. https://doi.org/10.1002/sim.698

Martell, C. R., Addis, M. E., & Jacobson, N. S. (2001). Depression in context: Strategies for guided action. Depression in context: Strategies for guided action. New York, NY: W.W. Norton. https://doi.org/10.1176/appi.ajp.160.7.1366

Mazzucchelli, T., Kane, R., & Rees, C. S. (2009). Behavioral activation treatments for depression in adults: A meta-analysis and review. Clinical Psychology: Science and Practice, 16, 383–411. https://doi.org/10.1111/j.1468-2850.2009.01178.x

Moradveisi, L., Huibers, M. J. H., Renner, F., Arasteh, M., & Arntz, A. (2013). Behavioural activation v. antidepressant medication for treating depression in Iran: Randomised trial. British Journal of Psychiatry, 202, 204–211. https://doi.org/10.1192/bjp.bp.112.113696

National Institute for Health and Clinical Excellence (NICE). (2016). Depression: The treatment and management of depression in adults (update). (Clinical Guideline 90). London. Available at: www.nice.org.uk/CG90. Retrieved from www.nice.org.uk/CG90

Pagoto, S., Schneider, K. L., Whited, M. C., Oleski, J. L., Merriam, P., Appelhans, B., … Crawford, S. (2013). Randomized controlled trial of behavioral treatment for comorbid

obesity and depression in women: The Be Active Trial. International Journal of Obesity, 37, 1427–1434. https://doi.org/10.1038/ijo.2013.25

Pasterfield, M., Bailey, D., Hems, D., McMillan, D., Richards, D., & Gilbody, S. (2014). Adapting manualized Behavioural Activation treatment for older adults with depression. The Cognitive Behaviour Therapist, 7, 1–11. https://doi.org/10.1017/S1754470X14000038

Patel, V. (2012). Global mental health: From science to action. Harvard Review of Psychiatry, 20, 6–12. https://doi.org/10.3109/10673229.2012.649108

Piper, W. E. (2008). Underutilization of short-term group therapy: Enigmatic or understandable? Psychotherapy Research, 18, 127–138. https://doi.org/10.1080/10503300701867512

Rehm, L. P. (1984). A self-management therapy program for depression. International Journal of Mental Health, 13, 34–53. https://doi.org/10.1080/00207411.1984.11448975

Richards, D. A., Ekers, D., McMillan, D., Taylor, R. S., Byford, S., Warren, F. C., … Finning, K. (2016). Cost and outcome of behavioural activation versus cognitive behavioural therapy for depression (COBRA): A randomised, controlled, non-inferiority trial. The Lancet, 6736, 1–10. https://doi.org/10.1016/S0140-6736(16)31140-0

Rothwell, P. M. (2005). External validity of randomised controlled trials: "to whom do the results of this trial apply?" Lancet, 365, 82–93. https://doi.org/10.1016/S0140-6736(04)17670-8

Schindler, A., & Hiller, W. (2010). Therapieeffekte und Responseraten bei unipolar depressiven Patienten einer verhaltenstherapeutischen Hochschulambulanz [Therapy effects and response rates of cognitive-behavioral treatment for unipolar depressive patients in an outpatient clinic]. Zeitschrift Fur Klinische Psychologie Und Psychotherapie, 39, 107–115. https://doi.org/10.1026/1616-3443/a000019

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The Consumer Reports

study. American Psychologist, 50, 965–974. https://doi.org/10.1037/0003-066X.50.12.965

Shaw, B. F. (1977). Comparison of cognitive therapy and behaviour therapy in the treatment of depression. Journal of Consulting and Clinical Psychology, 45, 543–551. https://doi.org//10.1037/0022-006X.45.4.543

Song, F., Loke, Y. K., Walsh, T., Glenny, A. M., Eastwood, A. J., & Altman, D. G. (2009, April 3). Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: Survey of published systematic reviews. BMJ. British Medical Journal Publishing Group. https://doi.org/10.1136/bmj.b1147

Steinert, C., Hofmann, M., Kruse, J., & Leichsenring, F. (2014). Relapse rates after psychotherapy for depression - Stable long-term effects? A meta-analysis. Journal of Affective Disorders, 168, 107–118. https://doi.org/10.1016/j.jad.2014.06.043

Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison, and validation of Meta-Essentials: A free and simple tool for meta-analysis. Research Synthesis Methods, 8, 537–553. https://doi.org/10.1002/jrsm.1260

Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? Statistics in Medicine, 21, 1559–1573. https://doi.org/10.1002/sim.1187

Turner, R. M., Bird, S. M., & Higgins, J. P. T. (2013). The impact of study size on meta-analyses: Examination of underpowered studies in Cochrane reviews. PLoS ONE, 8, e59202. https://doi.org/10.1371/journal.pone.0059202

Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. World Psychiatry, 14, 270–277. https://doi.org/10.1002/wps.20238

Weitz, E. S., Hollon, S. D., Twisk, J., Van Straten, A., Huibers, M. J. H., David, D., … Cuijpers, P. (2015). Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs pharmacotherapy: An individual patient data

meta-analysis. JAMA Psychiatry, 72, 1102–1109. https://doi.org/10.1001/jamapsychiatry.2015.1516

Westbrook, D., & Kirk, J. (2005). The clinical effectiveness of cognitive behaviour therapy: Outcome for a large sample of adults treated in routine practice. Behaviour Research and Therapy, 43, 1243–1261. https://doi.org/10.1016/j.brat.2004.09.006

Yalom, I. D., & Leszcz, M. (2005). The theory and practice of group psychotherapy (5th ed.). New York, NY: Basic Books.

Table 1. Subgroup and meta-regression analysis of Group BA versus controls (post-treatment)

Subgroup analysis		**No. of Comparisons**	**SMD (g)**	**95% CI**	**I^2(%)[a]**	**P (between subgroups)**	**NNT**
Control type	Waitlist	9	1.02**	0.69 to 1.35	32	0.01**	1.89
	TAU	4	0.20	-0.28 to 0.68	0		8.89
Assessment method	Clinical interview	11	0.75*	0.38 to 1.12	55*	0.72	2.57
	> clinical cut-off	2	0.53	-0.68 to 1.73	84*		3.42
BA type	Pleasant events	7	1.01**	0.62 to 1.40	4	0.08	1.91
	Self-control	3	0.87**	0.27 to 1.46	63*		2.17
	Contextual	1	0.40	-0.64 to 1.44	-		4.49
	BATD	2	0.00	-0.71 to 0.70	0		-
Population	Adults general	10	0.81**	0.44 to 1.17	25	0.26	2.31
	Young adults	2	1.06*	0.24 to 1.89	0		1.83
	Older adults	1	-0.09	-1.23 to 1.05	-		-19.71
Meta-regression analysis		**No. of Comparisons**	**B-coefficient**	**95% CI**	**SE**	**P**	**NNT**
Quality (risk of bias)	(0-4 criteria)	13	-0.31	-0.69 to 0.07	0.18	0.08	-
Initial depression severity	(z scores)	13	0.02	-0.35 to 0.38	0.17	0.93	-
Gender	(% of males)	13	-0.01	-0.02 to 0.00	0.01	0.08	-
Number of sessions	(2-12 sessions)	13	-0.11	-0.23 to 0.01	0.06	0.05	-
Group size	(3-10 patients)	13	0.02	-0.20 to 0.23	0.10	0.86	-

Note: *significant at $p < .05$ threshold; **significant at Bonferroni adjusted $p < .01$ threshold. [a] P value of Q-statistic as I^2 does not have a test of significance; [b]Effect non-significant when controlling for control type. Positive effect size indicates in favor of group BA. Abbreviations: TAU: treatment as usual; SMD: standardized mean difference; CI: confidence interval; SE: standard error; NNT: Numbers needed to treat; BATD: behavioral activation treatment for depression.

Table 2. Subgroup and meta-regression analysis of Group BA versus active therapy (post-treatment)

Subgroup analysis		**No. of Comparisons**	**SMD (g)**	**95% CI**	**I^2(%)[a]**	**P (between subgroups)**	**NNT**
Therapy type	CBT/CT	6	-0.10	-0.59 to 0.39	7	0.22	-17.74
	Other therapies	9	0.30	-0.10 to 0.70	30		5.95
BA type	Pleasant events	8	0.02	-0.45 to 0.49	11	0.74	88.62
	Self-control	5	0.23	-0.34 to 0.80	51		7.74
	Contextual	1	0.75	-0.53 to 2.03	-		2.48
	BATD	1	0.01	-1.13 to 1.15	-		177.24
Population	Adults general	11	0.24	-0.13 to 0.62	11	0.53	7.42
	Young adults	3	-0.12	-0.90 to 0.67	48		-14.79
	Older adults	1	-0.33	-1.55 to 0.89	-		-5.42
Meta-regression analysis		**No. of Comparisons**	**B-coefficient**	**95% CI**	**SE**	**P**	**NNT**
Quality (risk of bias)	(0-4 criteria)	15	-0.39	-0.91 to 0.12	0.24	0.10	-
Initial depression severity	(z scores)	14	-0.43	-1.01 to 0.16	0.30	0.15	-
Gender	(% of males)	15	0.00	-0.02 to 0.01	0.01	0.62	-
Number of sessions	(4-12 sessions)	15	-0.09	-0.22 to 0.04	0.07	0.17	-
Group size	(3-10 patients)	15	-0.09	-0.31 to 0.13	0.11	0.43	-

Note: *significant at $p < .05$ threshold; **significant at Bonferroni adjusted $p < .01$ threshold. [a] P value of Q-statistic as I^2 does not have a test of significance. Positive effect size indicates in favor of group BA. Abbreviations: CBT/CT: cognitive behavioral therapy/cognitive therapy; SMD: standardized mean difference; CI: confidence interval; SE: standard error; NNT: Numbers needed to treat; BATD: behavioral activation treatment for depression.

Figure 1. PRISMA flowchart of study selection.

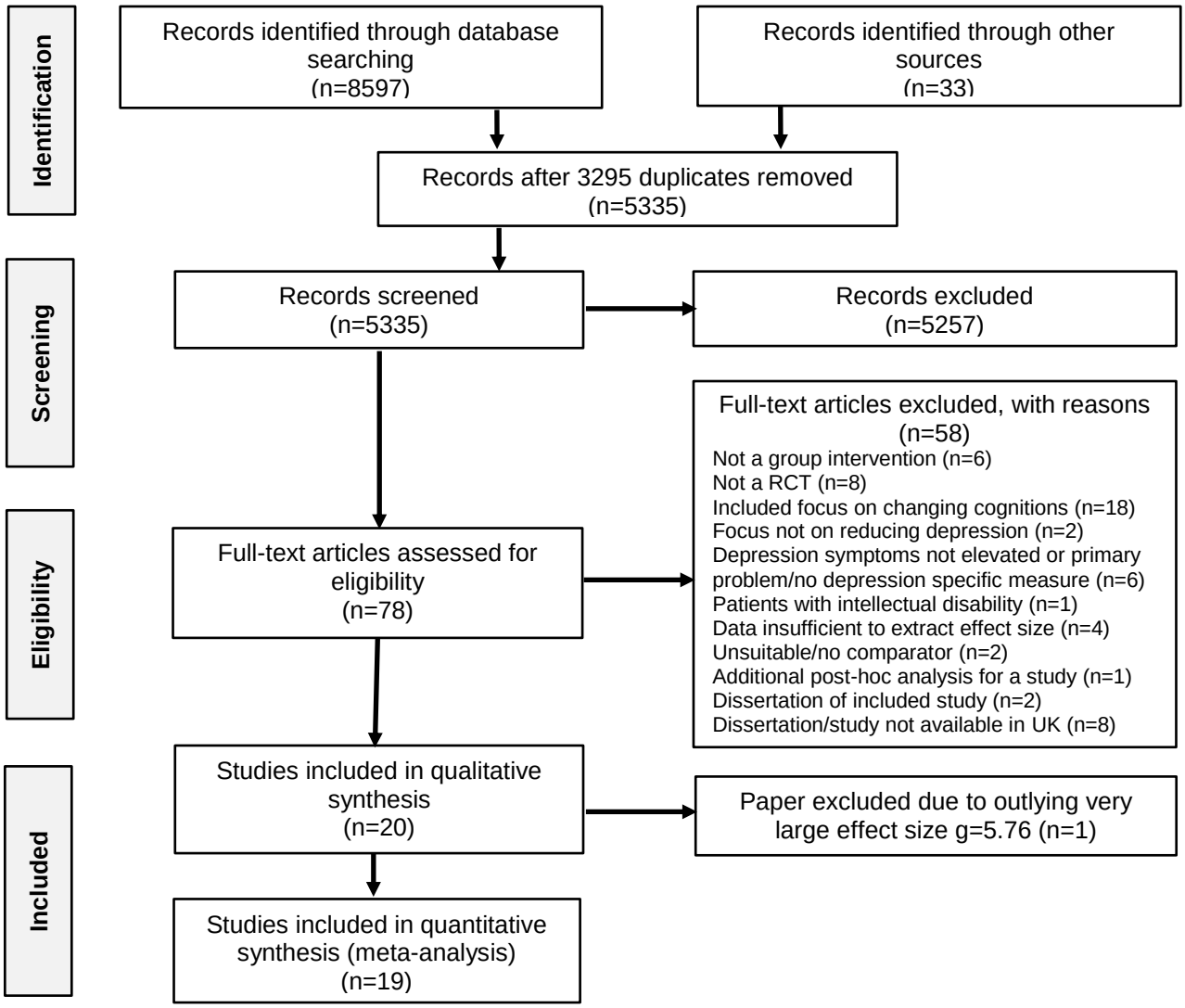

Figure 2. Forest plot of post-treatment and follow-up depression symptom effect sizes for group BA versus waitlist/TAU.

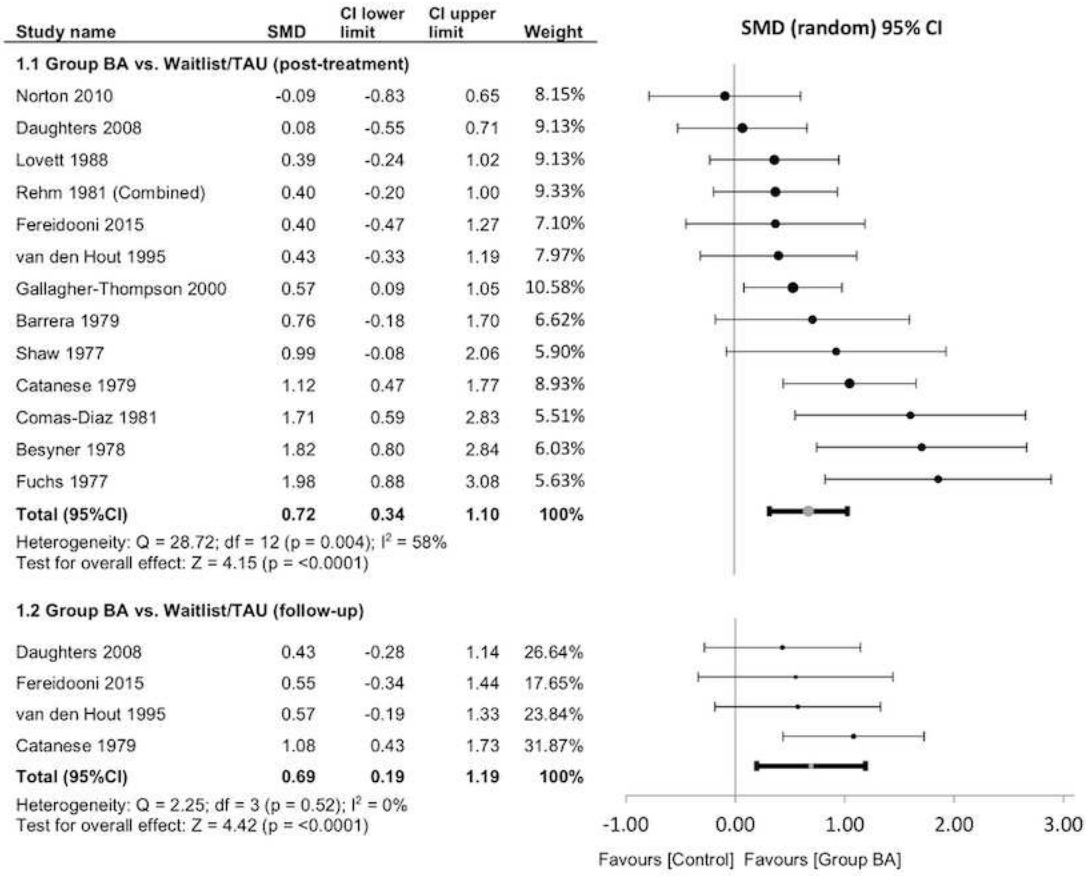

Figure 3. Forest plot of post-treatment and follow-up depression symptom effect sizes for group BA versus active treatment.