
Article:
Simmonds-Buckley, M. orcid.org/0000-0003-3808-4134, Kellett, S. orcid.org/0000-0001-6034-4495, Hague, J. et al. (1 more author) (2025) Improving outcomes via treatment augmentations to behavioural activation for depression in routine practice: a cohort comparison study. The Cognitive Behaviour Therapist, 18. e3. ISSN 1754-470X

https://doi.org/10.1017/s1754470x24000412

Abstract

Background. Behavioural activation (BA) is an empirically supported treatment for depression, but some patients do not clinically benefit from this approach. *Aims.* This study evaluated the effectiveness of two treatment augmentations to an extant manualized 8-session group version of BA delivered in routine practice. Treatment augmentations were dose-response psychoeducation to improve group attendance and implementation intentions to improve clinical outcomes. *Method.* A cohort comparison design, using propensity score matching, compared attendance and clinical outcomes for group BA (*n*=31, drawn from a sample of n=161, from 22 BA groups) with treatment-augmented group BA (*n*=31 from 3 BA+ groups). *Results.* There was no effect of the two treatment augmentations on attendance rates, but clinical outcomes were significantly improved. *Conclusions.* More efforts should be made to improve outcomes for empirically supported interventions in routine services, with treatment augmentations tested in well controlled studies.

Key learning aims:

(1) To learn about the utility of adapting existing cognitive-behavioural treatments as opposed to developing new cognitive-behavioural treatments.

(2) To learn about the potential of propensity score matching in routinely collected datasets.

(3) To learn about group delivery of behavioural activation treatment protocols.

(4) To better understand how to enhance and evaluate group BA treatment protocols using low cost and theoretically or empirically-informed treatment augmentations.

When empirically supported interventions are delivered in routine practice there is marked difference in effectiveness compared to the clinical trials (Gyani et al., 2013; Hansen et al., 2002), with outcomes up to 12% lower in routine practice and particularly when patients do not receive evidence-based interventions (Barkham et al., 2008; Hansen et al., 2002). A primary challenge of translational science for psychotherapy is therefore enabling patients in routine services to have the same quality of intervention and chance of recovery as a participant in a clinical trial (Strauman, Eddington & McCrudden, 2007). Translation efforts have fallen into three main categories: (1) clear treatment guidelines and protocols, (2) routine outcome monitoring, (3) treatment augmentations and (4) enhancing treatment competency. Treatment guidelines ensure that only empirically supported treatments are recommended (American Psychiatric Association [APA], 2010; National Institute for Health and Clinical Excellence [NICE], 2018) and then increasing access enables these empirically supported treatments to be delivered in routine services (Jaycox et al., 2003). Regular supervision, use of treatment manuals and auditing of treatment integrity ensure consistent delivery of evidence-based practice (Bambling et al., 2006; Wilson, 1996; Power et al., 2022). Outcome monitoring consists of tracking sessional outcomes and enables therapists to take associated remedial action (Wampold, 2015; Delgadillo et al., 2018; Delgadillo et al., 2022; Lambert, 2017). Such studies tend to show that when therapists feedback whether their patient is following a symptom trajectory likely to end in remission/recovery, this is associated with better outcomes. Treatment augmentation has been achieved through a variety of practical-technological (e.g., smart phone applications - Ly et al., 2015; and automatic text messaging – Aguilera et al., 2017) and theoretical innovations (Avishai et al., 2018). Treatment competency concerns the effective training and supervision of clinical supervisors and the availability and usage of valid and reliable measures.

A range of meta-analyses have shown that one-to-one and group behavioural activation

(BA) is an effective and efficacious treatment for depression (Ekers et al., 2014; Richards et al., 2016; Simmonds-Buckley et al., 2019; Pott et al., 2021). Behavioural activation is a time-limited psychotherapeutic approach that aims to change the manner which a patient interacts with their immediate environment through the action of three mechanisms; (1) increasing contact with positive reinforcers of healthy behaviours, (2) reducing avoidance behaviours that limit contact to positive reinforcers and (3) understanding and then addressing any blocks to activation (Uphoff et al., 2019). Recovery rates however indicate at least 40% of BA patients do not experience a statistical clinically significant and reliable change on depression outcome measures (Hansen et al., 2002; Hopko et al., 2011), indicating the need to consider treatment augmentation. It is acknowledged that patients may meet their idiosyncratic treatment goals during BA and not reach the statistical threshold for reliable change on nomothetic outcome measures. The parsimonious nature of BA makes it particularly well-suited to treatment augmentation, without unduly affecting the theoretical integrity of the approach (Hopko et al., 2003). Treatment augmentation need to target key barriers and facilitators of change (van Bokhoven et al., 2003) and in the context of BA then *treatment acceptability* and *treatment engagement* are viable targets for treatment augmentation.

Patients need to receive an adequate 'dose' of therapy in routine services to facilitate outcome, often referred to as the dose-response effect (Robinson et al., 2020). Patients' dropout of depression treatment due to the debilitating and demotivating impact of low mood, wider systemic factors and poor treatment fit (Barrett et al., 2008). The recent Gaskell et al., (2023) meta-analysis of the outcomes archived in routinely delivered interventions noted that the lower outcomes achieved in routine practice compared to clinical trials may be due to multiple factors including poor therapist attitudes to protocol-delivered interventions. Treatment acceptability also suffers when there is a discrepancy between patient expectations about rate of improvement and required number of sessions (Swift et al., 2011). When patient expectations of the duration

of therapy differ from the service offer, the likelihood of drop-out increases (Callahan et al., 2009; Mueller & Pekarik, 2000). Psychoeducation on dose-response evidence has been seen to effectively align patient expectations about treatment duration to a dose that would more likely invoke meaningful symptom improvement and this encourages patients to stay in treatment for longer (Swift & Callahan, 2011). Dose-response psychoeducation was selected as a treatment augmentation due to it being easy to implement and also being able to be drawn from extant BA evidence-based information (Kellett, Simmonds-Buckley, Bliss, & Waller (2017).

Treatment engagement is important during BA as the majority of the change work occurs via activation homework (Hopko et al., 2011). Patients' implementation of BA via homework is crucial in generating a sense of progress and so drives positive clinical outcomes (Beck & Tomkins, 2007). The Kazantzis et al., (2010) meta-analysis found an effect size of d = 0.63 for therapies without homework versus d = 1.08 for therapies with homework. Meta-analyses of the relationship between homework adherence and outcome find most modest effects (r = .22 in Kazantzis, Deane & Ronan, 2000; r = .26 in Mausbach, Moore, Roesch, Cardenas & Patterson, 2010). Whilst homework may be planned, it does not guarantee successful completion, so creating the "intention-behaviour gap" (Sheeran & Webb, 2016). Implementation intentions are a simple behaviour change technique that has been shown to close the intention-behaviour gap and so increases goal attainment (Wang, Wang & Gai, 2021). The technique involves the generation of specific plans about how, when and where goals will be acted upon and these are crystalized using a brief 'if-then' format (Gollwitzer, 1999). Establishing 'if-then' plans link intended actions to environmental cues, thus making actions more immediate and automatic and so removing the need for unhelpful procrastination (Webb & Sheeran, 2008). As failure to engage in activation has been identified as a contributing factor to non-response in BA (Hopko et al., 2011), then formation of implementation intentions was selected as the second treatment augmentation. This was also based on evidence of their highly acceptability to patients

(Lucock et al., 2018) and that implementation intentions have been shown to double the rate of activation-related goal attainment during the treatment of depression (Fritzsche et al., 2016). Implementation intentions would be considered in BA theory as one of raft of contingency-management strategies (Kanter et al., 2010).

To summarize, few empirical studies have used translational science approaches to test the effectiveness of treatment augmentations to improving depression outcomes in routine practice (Portela et al., 2015). The main aim of this cohort comparison study was therefore to test whether treatment augmentations to BA groups (BAG) delivered in routine practice treatment were effective. The study hypothesis was that in comparison to the BAG treatment, augmented BAG would have better attendance rates and improved clinical outcomes. The study was conducted in an Improving Access to Psychological Therapies (IAPT) service. These services are now called NHS Talking Therapies for Anxiety and Depression. National evaluations (NHS Digital, 2020) benchmark annual outcomes against a target of a 50% recovery rate. This highlights the need to improve outcomes for the 50% that do not recovery during interventions delivered in Talking Therapies services.

Methods

Ethics, Design and Setting

The study received ethical and research governance approval and was publicly registered (references removed to ensure blinded review) and authors have abided by the Ethical Principles of Psychologists and Code of Conduct as set out by the BABCP and BPS. Participants therefore gave informed consent to participate in the study and for publication. A cohort comparison design was used to compare outcomes for routine delivery of standard BAG with an augmented BAG (i.e. BAG+). Both BAG and BAG+ were delivered in the 'high intensity therapy' tier of a single NHS Talking Therapies service (see Clark 2018 for full description of IAPT stepped care approach). A sample size analysis using G*Power (Faul et al., 2007) indicated $n = 27$ would be

needed in each group (total *N*=54) to detect a small to medium effect size (d = 0.35) with .80 power using a repeated measures between-subjects ANOVA at p = .05. Retrospective anonymized routine outcome data from patients who had previously received BAG therefore formed the historical control. Samples were matched using propensity score matching (PSM). This method enables cohort comparisons to mimic the features of an RCT through balancing pre-treatment covariates in the groups of interest (Austin, 2011) – full details of the PSM procedure are described in the Data Analysis section.

Participants

The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) diagram summarizing patient flow and sample selection is presented in Figure 1. Inclusion criteria were: (a) seeking treatment for a primary presenting problem of depression; b) referred following assessment by a Psychological Wellbeing Practitioner (PWP; Clark, 2018) identifying depression as the presenting problem; (c) attended at least one BAG treatment session and (d) were at least 18-years old. The single exclusion criterion was not meeting criteria for depression caseness prior to commencing BAG (i.e., a PHQ-9 score <10). Out of 34 patients who attended BAG+ across three groups, 31 met the criteria for depression, so had their outcomes included in the analysis. Out of 178 patients who had attended BAG across 22 groups, 161 met the inclusion criteria. From the available pool of 161 BAG patients, 31 were propensity score matched to the 31 eligible BAG+ patients to ensure clinical equivalence at baseline assessment prior to outcome comparison.

Outcome Measures

The outcome measures consisted of the NHS Talking Therapies minimum dataset (Patient Health Questionnaire-9 [PHQ-9], Kroenke et al., 2001; Generalized Anxiety Disorder-7 [GAD-7], Spitzer et al., 2006; and Work and Social Adjustment Scale [WSAS], Mundt et al., 2002). Caseness on the PHQ-9 is a score ≥10, on the GAD-7 ≥8 and on the WSAS >20. For BAG and

BAG+ cohorts, measures were completed at the start of every group session. Patients receiving BAG+ completed a demographic information sheet to capturing age, gender, ethnicity, current antidepressant medication and any previous episodes of depression and treatment. Anonymized clinical outcome measures and brief demographic information for the BAG cohort were accessed via clinical records of routinely collected data.

Group Behavioural Activation; Facilitation, Delivery and Adherence

BAG or BAG+ groups were always facilitated by two UK Association for Behavioural and Cognitive Psychotherapies (BABCP) accredited CBT therapists. All groups were delivered in the same primary care settings. For consistency, when a group was set up, then the same two facilitators delivered all sessions. Due to BAG+ being delivered as a service innovation on BAG, then the eight BAG+ facilitators had also previously delivered BAG. There were seven female facilitators and one male facilitator. All had completed the same BABCP accredited 1-year CBT training programme that contained a two-day workshop on BA (i.e., a Post Graduate Diploma in High Intensity Psychological Interventions). Length of time qualified varied from 2-6 years. All group facilitators attended a quarterly BAG peer supervision group and had one-hour fortnightly one-to-one clinical supervision. Supervision time was therefore matched between BAG and BAG+. A one-hour training intervention was provided for the pool of facilitators (N=8) to support them to deliver BAG+. This introduced the dose-effect psychoeducation and enabled practice in helping patients set implementation intentions. A questionnaire evaluation of the training intervention illustrated that >80% rated understanding the theory and evidence base for implementation intentions and having confidence in using the approach. BAG and BAG+ consisted of eight, weekly, two-hour manualized sessions based on an extant treatment protocol and patient workbook (Houghton et al., 2008; Martell et al., 2010). The BAG protocol was enhanced with two treatment augmentations to produce a BAG+ version of the intervention. Apart from the two treatment augmentations in BAG+, the intervention

received by both cohorts was identical and were matched in terms of time (i.e. augmentations were integrated into extant duration of each BAG+ group).

Supplementary materials summarize the protocol and also details how and where the treatment augmentations were integrated. The first treatment augmentation was a data-informed psychoeducation enhancement targeted at increasing attendance rates. The psychoeducation consisted of dose-effect evidence taken from a pilot BAG outcome study (Kellett et al., 2017). The psychoeducation was included in the pre-treatment information pack sent to patients and stated that: (1) attending at least 4 sessions was required to enable change; (2) BAG was effective regardless of the severity of depression and (3) BAG was effective at also reducing co-morbid anxiety symptoms. The second treatment augmentation was teaching patients how to set 'implementation intentions' when planning homework at the end of each group. Implementation intentions were (1) introduced and modelled by the facilitators at the end of the first session, (2) the workbooks contained if-then planning sheets and (3) a session-specific example of an implementation intention homework plan was provided for every session. Patients worked in pairs and formed implementation intentions using the worksheet at each session for how each idiosyncratic homework assignment would be implemented. Patients silently repeated their homework implementation intention to themselves three times and then repeated it verbally to their group partner. This is standard practice in implementation intention work in clinical samples (Avishai et al., 2018).

Treatment adherence was assessed using an adapted version of an adherence check that has previously been used in a BA trial (Ekers et al., 2011). The checklist included a *general adherence* section (split into items related to the behavioural rationale and items related to between-session work), a *session specific adherence* section, and an overall assessment of whether the session couFld be rated as BA. An item relating to 'use of implementation intentions' was included to check adherence to the BAG+ augmentation. A customized page of

the checklist was adapted for every BAG session to distinguish aspects that would not be expected to be present due to session content. The session specific *mood dependence* item from the BAG checklist was changed to a general adherence item in the BAG+ checklist. After each BAG session, the two facilitators independently completed the adherence checklist. BAG+ adherence was checked after every session delivered throughout the duration of the study. BAG adherence was checked for the delivery of the final two groups of the existing BAG protocol. Full details of the adherence checks summarizing the mean rating for the presence of evidence in each category for BAG and BAG+ are provided in the Supplementary Materials (Figures S1 and S2).

All BAG and BAG+ sessions were rated as representative of BA therapy, indicating that patients received a protocol-adherent group treatment. All the adherence items were deemed to have been present in the sessions, with the majority rated as having very clear or sufficient evidence (BAG = 73%; BAG+ = 85%). Adherence checks of the implementation intentions augmentation showed they were not used in BAG delivery (as expected), but were present during BAG+, with sufficient or very clear evidence in over 90% of sessions. Inter-rater reliability between group facilitators was assessed using Cohen's kappa (Cohen, 1960). Adherence agreement was k=.57 and k=.44 for BAG and BAG+ respectively, indicating moderate agreement (Landis & Koch, 1977).

Data Analysis Plan

The data analysis plan had three main stages. In the first stage, the entire eligible BAG+ sample (n=31) were matched to a comparative subsample of BAG patients (n=31) using propensity score matching (PSM). All patients in the existing BAG archived database who met the inclusion requirements were eligible for matching (N=161). Samples were matched on age, baseline depression (PHQ-9 score), baseline functioning (WSAS score) and employment status. These were selected due evidence identifying these as key influencers of outcomes for

interventions delivered in these settings (Delgadillo et al., 2016). A one-to-one, nearest neighbor matching procedure without replacement was applied with a propensity score within a caliper tolerance of 0.2. To ensure adequate matching, mean difference (standardized differences/proportions) and distribution (variance ratios and five number summaries - minimum, 25th percentile, median, 75th percentile and maximum) diagnostics were performed on the covariates across BAG and BAG+ prior to and post-matching. Unmatched and matched sample demographics are reported in the results.

During the second stage, in order to assess the impact of clustering in the data, intraclass correlation coefficients (ICCs) were used to estimate the level of variance attributable to BAG group level factors. ICCs and the associated design effect (DE) for all the outcome measures were calculated. A DE of greater than two was used as an indication of significant co-dependence and therefore being unsuitable for analysis on a single-level (i.e., would require use of a multi-level model; Muthen & Satorra, 1995). BAG was delivered in 13 groups in total (BAG = 10 and BAG+ = 3). The average cluster size was 4.77. ICCs calculated for PHQ-9 (-0.04), GAD-7 (-0.05) and WSAS outcomes (0.06) produced design effects of 0.85, 0.81 and 1.23 respectively. As all the DEs were less than two, single level analyses were deemed appropriate. Outcomes were analyzed using the intention-to-treat (ITT) principle, including all patients who entered group treatment in the analysis. As outcomes were collected at every session, missing data were accounted for using last observation carried forward (LOCF) imputation.

The final stage evaluated the effect of the BAG+ augmentations on attendance and recovery rates and these were calculated for both cohorts. Attendees at every session were defined as ‘treatment completers’, 4-7 sessions as ‘partial attenders’, and attendees at three sessions or fewer were deemed as ‘dropouts.’ Average session attendance in BAG and BAG+ was also calculated. Reliable and clinically significant change criteria were applied to

depression outcomes to determine recovery rates (Jacobson & Truax, 1991). Talking Theapies reliable change thresholds were used and defined as 'deterioration' being when there was a reliable increase in PHQ-9 scores of ≥6 (i.e. an increase in depression), 'nonresponse' when no reliable change on the PHQ-9 occurred in either direction (i.e., neither improvement or deterioration in depression), 'improvement' when there was a reliable decrease in PHQ-9 scores of ≥6 (i.e. a decrease in depression), and 'recovery' when there was an decrease in PHQ-9 scores of ≥6 (i.e., a reduction in depression) in addition to the pre-post score moving from above to below the PHQ-9 clinical cut-off. Clinical outcomes, attendance and recovery rates were compared for the BAG and BAG+ cohorts using chi-square and odds-ratios for binary outcomes and analysis of variance (ANOVA)/T-tests (one-tailed) and Cohen's *d* within and between group effect sizes for continuous outcomes. Effect sizes of 0.2, 0.5 and 0.8 were considered small, moderate, and large effect respectively (Cohen, 1992).

Results

The results are presented in three sections: (1) sample matching and description; (2) treatment acceptability evaluation and (3) clinical outcome evaluation.

Sample matching and sample description

The matched dataset (*N*=62) was checked to ensure sufficient distribution of covariates across the samples in comparison to the unmatched sample (*N*=192). Supplementary materials contain the comparison of baseline covariates in BAG and BAG+ in the overall unmatched sample and after PSM matching and also the variance and distribution of the continuous covariates before and after matching. Standardized differences demonstrated that imbalance in all the specified covariates across BAG and BAG+ were reduced to below the specified threshold after matching (d <0.10), so indicating minimal differences. The results combined therefore suggested that PSM procedures were successful at matching BAG and BAG+ pre-intervention. Supplementary materials also describe the characteristics of the BAG and BAG+ patients included in the final

sample. Pre-treatment depression was classified as 37% (*n*=23) severe depression, 48% (*n*=30) moderately severe depression, and 15% (*n*=9) moderate depression. Nearly 89% (*n*=55) also met clinical caseness for anxiety (a GAD-7 score of ≥8), with 42% (*n*=26) classified as severe, 37% (*n*=23) as moderate, 18% (*n*=11) as mild, and 3% (*n*=2) as experiencing minimal anxiety. On the WSAS, 73% met caseness for impairment.

Treatment acceptability outcomes

There was no difference ($t(60) = 0.92$, $p = .180$) regarding number of sessions attended during BAG (Mean 4.6, $SD = 2.6$) versus BAG+ (Mean 5.2, $SD = 2.4$). Table 1 shows no significant differences between BAG and BAG+ regarding treatment engagement rates. Therefore, the treatment augmentations to BAG+ did not improve acceptability of the intervention.

Clinical outcomes

Table 2 presents the means and pre-post effect sizes, between-treatment effect sizes and pairwise t-test statistics of BAG and BAG+. Depression symptoms significantly decreased following BAG treatment ($F(1, 60) = 45.22$, $p < .001$), with significantly greater reductions in depression symptoms during BAG+ ($F(1, 60)=2.91$, $p = .047$). Within-group treatment reductions in depression symptoms represented moderate-large and large effects for BAG and BAG+ respectively. The lower post-treatment depression scores for BAG+ were representative of a small between-groups effect (d=0.43). Both BAG and BAG+ both produced small-moderate pre-post reductions in anxiety (GAD-7) and impaired functioning (WSAS). No differences were found between BAG and BAG+ in terms of anxiety ($F(1, 60)=1.98$, $p = .082$) or functioning outcomes ($F(1, 60)=0.40$, $p = .265$).

Figure 2 displays the session-by-session PHQ-9 scores during BAG versus BAG+. Both treatments produced early session reductions in depression and whilst BAG scores plateaued over later sessions, BAG+ outcomes continued to decrease. Significant differences in PHQ-9 scores between BAG and BAG+ were apparent by the 8th session ($t(60) = 1.71$, $p = .049$).

Table 3 summarizes case-by-case outcomes to show that BAG+ produced a significantly lower number of nonresponse outcomes. Patients who received BAG+ treatment were three times less likely to experience nonresponse at the end of treatment. The reduced nonresponse outcomes were explained by significantly more patients in BAG+ experiencing improvement in depression symptoms or full recovery (change in caseness, in addition to improvement). No patient experienced a reliable deterioration in their depression after attending either group intervention. Overall, the two treatment augmentations appeared to improve outcomes in BAG+ with evidence of significantly greater reductions in depression symptoms and significantly fewer BAG+ patients experiencing a nonresponse outcome.

Discussion

This study took an implementation science approach. This tested whether two treatment augmentations to an evidenced-based intervention for depression delivered in routine practice (i.e. group BA) improved the acceptability and effectiveness of that intervention. The use of practice-based data ensure that the results of this study have high external validity. Fidelity to the BAG and BAG+ treatment protocols were checked and was sufficient and PSM procedures then ensured a fair comparison between BAG and BAG+. Except for the treatment augmentations, all BAG and BAG+ patients received the same matched 8-session manualized group BA depression intervention. The combined treatment augmentations did not result in greater attendance rates during BAG+ and so did not improve the acceptability of the intervention. The overall attendance rates (29-42%) were poorer than those reported in the general literature (Swift & Greenberg, 2012) and specifically in Talking Therapies services (Kellett et al., 2021). However, the combined treatment augmentations appeared to increase effectiveness, as BAG+ patients were approximately three times more likely to have improved

depression outcomes. The results additionally provide further evidence that BA is clinically effective when delivered in a group in routine practice, producing moderate to large reductions in depression (Simmonds-Buckley et al., 2019).

Importantly, there were no cases of depression symptom deterioration suggesting that the treatment augmentations were safe. Improved outcomes were more evident in the individual outcomes, with nonresponse outcomes reduced from approximately 77% (BAG) to 52% (BAG+). Those fewer patients experiencing symptom stasis (i.e., those patients that did not achieve at least a reliable improvement) during BAG+ was a consequence of 26% more patients experiencing reliable improvement, with 19% of those also moving below the clinical cut-off for depression. Converting these improvements, and those of other treatment enhancements reported in the literature, into comparable effect sizes suggest the current findings are comparable to other treatment augmentations. The beneficial effect of BAG+ on recovery rates created a moderate effect (d=0.56). The augmentations were low-cost. The evidence base for low-cost treatment augmentations is somewhat mixed. Kellett et al., (2004) used practice-based evidence to better match patients to group CBT and improved the effectiveness of a group intervention and implementation intentions have been found to increase attendance at low intensity group psychoeducational sessions (Avishai et al, 2018). Text-messaging also shown moderate benefits for increasing attendance (d=0.5) but have more limited effects on clinical outcomes (Aguilera et al., 2017). Delgadillo et al. (2015) found that telephone text message appointment reminders and treatment orientation psychoeducation did not increase attendance.

Theoretical and Clinical Implications

The brief training intervention to support the treatment augmentations appeared feasible as it achieved its aim of enabling facilitators to change their BA practice. BAG was augmented with two simple strategies easily integrated into the extant group structure and protocol. One was a psychoeducational augmentation informed by BAG evidence targeting attendance and the other

was a theory-informed augmentation targeting outcome. BAG+ patients experienced greater clinical improvements despite attending the same number of group treatment sessions as the BA patients. It is unknown whether patients engaged with the pre-course materials, although therapists did re-visit the information during the first BAG+ session. As depression is known to have a considerable effect on attention and memory (Otte et al., 2016), it is possible that patients found it difficult to process and retain the psychoeducation information.

These findings build on the promising use of implementation intentions in mental health contexts and demonstrate the potential of integrating 'if-then' plans into existing treatment protocols (Lucock et al., 2018; Toli et al., 2016). The behavioural foundation of BA aided the integration of implementation techniques into the process of setting of bespoke and idiosyncratic homework activities (Toli et al., 2016). Implementation intentions have previously been shown to promote engagement in personally valued activities (Fritzsche et al., 2016). The separation observed in depression outcome trajectories during BAG+ and BAG from session three may be due to the implementation intentions mechanism accounting for the BAG+ improvements (i.e. via more effective homework completion). However, as homework completion was not monitored, it is difficult to draw firm conclusions. BA highlights the importance of context in both the maintenance of depression and the breaking of depressive cycles via changes in behaviours (Martell et al., 2001). Similarly, implementation intentions promote the use of contextual cues to initiate pre-planned actions (Sheeran & Webb, 2016). The mechanisms of 'if-then' planning therefore was relatively easy to integrate into the principles and practice of BA. Fritzche et al. (2016) has previously illustrated the utility of implementation intentions in managing low mood.

The development of new depression treatments, whilst welcome, are clearly not the only methods for improving depression outcomes. Treatment augmentations hold the potential of producing meaningful change with quicker time frames for implementation and evaluation. The

present results advocate the development and testing of small theory-informed augmentations to currently available evidenced-based therapies as a method for improving patient outcomes in routine services. It should be noted, however, that although BAG+ reduced the rate of depression nonresponse, approximately half the patients receiving treatment still failed to experience meaningful change. Individual and clinical factors interact to make patients more suited to certain psychotherapies, with treatment matching then shown to reduce this mismatch (Beutler et al., 2018).

Limitations and Future Research Directions

The lack of randomization and a true control condition means that the results of this study are limited by lack of internal validity. The use of a historical control group means patients were not able to be randomized to treatments and the effects found cannot therefore be attributed with true confidence to treatment augmentations. Future studies should therefore consider randomizing participants into BAG versus BAG+ to better test whether the treatment augmentations are efficacious. Because two augmentations were delivered in BAG+ it is impossible to disaggregate their separate effects. Potential confounds from the differences in how the cohorts were recruited into the study, how data were collected, temporal trends or small changes in service delivery over the time-period, unknown usage of antidepressant medication (in the BAG historical control cohort) and lack of information on any concurrent treatments (both cohorts) could account for the differences found. In BAG+ patients signed a consent form to participate and may have primed them to respond positively. The baseline comparisons between study groups were limited and did not include data on medication and employment status. The variables used to match participants could have been expanded to include the GAD-7 score, employment status, medication, long-term health condition status or disability and the amount of time waiting between referral and treatment starting.

The direct association of the treatment augmentations on the intended outcomes cannot

also be fully justified. For example, the dose-response psychoeducation augmentation may have strengthened trust in the efficacy of BAG+ and so created a placebo effect on outcomes. The two quality improvements may therefore have had a synergistic effect. Drawing a conclusion of a direct connection between implementation intentions and clinical outcome is particularly undermined due to the lack of information concerning amount/quality of homework completed. Monitoring homework compliance across the duration of BAG would improve the method of future studies, particularly as homework compliance has been shown to fall across the duration of CBT interventions for depression (Gaynor, Lawrence & Nelson-Gray, 2006). The durability of the outcome differences observed were not assessed, as no follow-up was conducted. Long-term follow-up studies of the durability of treatment augmentations would be welcome.

Treatment completion rates of the full eight-sessions were sub-optimal, so session-by-session scores were utilized to ensure a pre-post score was available for the whole sample. The LOCF method does have acknowledged limitations (Lachin, 2016) and the results are based solely on self-report data with associated risks of social desirability bias (Tourangeau & Yan, 2007). Using a combination of self and independent-assessor rated outcomes would therefore strengthen the methodology of future BAG studies. The adherence check used in the study relied on the BAG+ therapists to self-rate adherence and because of the historical control nature of the study, only two BAG groups were able to be rated for adherence. Future research would therefore benefit from independent verification of the delivery of treatment augmentation strategies. Because limited adherence data was possible from BAG groups, and the adherence measure for BAG excluded the implementation intention item, the study could not absolutely verify that implementation intentions were absent from all BAG groups. As the BAG+ facilitators were introduced to the study and knew that the aim was to improve outcomes, this may have primed facilitators to deliver BAG+ groups more effectively. In terms of attendance the use of categorical response variables (i.e., attendance cut-offs of 3 or less, 4-7, or > 7 group

BA sessions) was stipulated in the study protocol, but it is acknowledged that such cut-offs are somewhat arbitrary. Finally, although checks were performed to assess the suitably of single-level analyses for the clustered data (by nature of group delivery), it has been argued ICCs as low as 0.01 can still violate dependency assumptions (Baldwin et al., 2011). It would be helpful if the whole service level reporting of the Talking Therapies programme (NHS Digital, 2022) disaggregated outcomes by high intensity intervention. This would have enabled the benchmarking of the results from the current study to treatment outcomes seen for group BA groups in Talking Therapies services more widely.

Conclusions

Treatment augmentations to extant evidence-based therapies offer a simple and direct means by which services can potentially improve outcomes. Interventions need to be 'tweaked and tested' rather than replaced. The variety of processes that help to produce positive change during psychotherapy for depression usefully provides multiple targets for treatment augmentation. Future research should continue to establish the processes that enable treatments to exert their positive influence and then target these for treatment augmentation in controlled studies.

Key practice points

(1) Because much of the change work of BA is completed between sessions when activation is put in place, then CBT therapists should pay attention to any factors that reduce the likelihood of homework completion and reward homework completion.

(2) When homework is not completed, the CBT therapists should apply the TRAP and TRAC approach to the analysis of this behaviour.

(3) The reviewing and mutual design of homework exercises at each session is an effective way of socializing patients to CBT and provides a containing structure to sessions.

(4) The delivery of BA in groups holds promise in terms of enabling patients to learn from each

other and be a support to each other in term of understanding the function of their behaviors and how to adopt an ‘outside-in’ approach to change.

Further reading

Martell, C.R., Dimidjian, S., & Herman-Dunn, R. (2022). Behavioral Activation for Depression, Second Edition; A Clinician's Guide. Guilford Press.

Kanter, J. W. Busch, A. M., Laura C. Rusch, L. C. (2009). Behavioral Activation: Distinctive Features: 5 (CBT Distinctive Features). Routledge.

Veale, D. (2008). Behavioral activation for depression. *Advances in Psychiatric Treatment*, 14, 29-36

References

Aguilera, A., Bruehlman-Senecal, E., Demasi, O., & Avila, P. (2017). Automated text messaging as an adjunct to cognitive behavioural therapy for depression:A clinical trial. *Journal of Medical Internet Research*, *19*, e148. https://doi.org/10.2196/jmir.6914

American Psychiatric Association. (2010). *Practice Guideline For The Treatment Of Patients With Major Depressive Disorder*. 3rd ed. American Psychiatric Association.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*, 399–424. https://doi.org/10.1080/00273171.2011.568786

Avishai, A., Oldham, M., Kellett, S., & Sheeran, P. (2018). Sustaining attendance at a mental health service : A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *86*, 1056–1060. https://doi.org/10.1037/ccp0000341

Baldwin, S. A., Murray, D. M., Shadish, W. R., Pals, S. L., Holland, J. M., Abramowitz, J. S., Andersson, G., Atkins, D. C., Carlbring, C., Carroll, K. M., Christensen, A., Eddington, K. M., Ehlers, A., Feaster, D. J., Keijsers, G. P. J, Koch, E., Kuyken, W., Lange, A., Lincoln, T., … Watson, J. (2011). Intraclass correlation associated with therapists: estimates and applications in planning psychotherapy research. *Cognitive Behaviour Therapy*, *40*, 15–33. https://doi.org/10.1080/16506073.2010.520731

Bambling, M., King, R., Raue, P., Schweitzer, R., & Lambert, W. (2006). Clinical supervision: Its influence on client-rated working alliance and client symptom reduction in the brief treatment of major depression. *Psychotherapy Research*, *16*, 317–331. https://doi.org/10.1080/10503300500268524

Barkham, M., Stiles, W. B., Connell, J., Twigg, E., Leach, C., Lucock, M., Mellor-Clark, J., Bower, P., King, M., Shapiro, D. A., Hardy, G. E., Greenberg, L., & 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

Barrett, M. S., Chua, W. J., Crits-Christoph, P., Gibbons, M. B., & Thompson, D. (2008). Early withdrawal from mental health treatment: Implications for psychotherapy practice. *Psychotherapy: Theory, Research, Practice, Training*, *45*, 247-267. https://doi.org/10.1037/0033-3204.45.2.247

Beck J.S., & Tompkins M.A. (2007). Cognitive therapy. In N. Kazantzis & L. LL'Abate (Eds.), *Handbook of homework assignments in psychotherapy* (pp. 51-63). Springer. https://doi.org/10.1007/978-0-387-29681-4_4

Beutler, L. E., Kimpara, S., Edwards, C. J., & Miller, K. D. (2018). Fitting psychotherapy to patient coping style: A meta-analysis. *Journal of Clinical Psychology, 74,* 1980–1995. doi: 10.1002/jclp.22684

Callahan, J. L., Aubuchon-Endsley, N., Borja, S. E., & Swift, J. K. (2009). Pretreatment expectancies and premature termination in a training clinic environment. *Training and Education in Professional Psychology*, *3*(2), 111. https://doi.org/10.1037/a0012901

Clark, D. M. (2018). Realizing the mass public benefit of evidence-based psychological therapies: the IAPT program. *Annual review of clinical psychology*, *14*. https://doi.org/10.1146/annurev-clinpsy-050817-084833

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

Delgadillo, J., de Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., Ali, S., Aguirre, E., Appleton, M., Nevin, J, O'Hayon, H., Patel, U., Sainty, A., Spencer, P., & McMillan, D. (2018). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: A multisite, open-label, cluster randomised controlled trial. *The Lancet Psychiatry*, *5*, 564–572. https://doi.org/10.1016/S2215-0366(18)30162-7

Delgadillo, J., Ali, S., Fleck K, et al. (2022) Stratified care vs stepped care for depression: A cluster randomized clinical trial. *JAMA Psychiatry.* 2022;79(2):101–108. doi:10.1001/jamapsychiatry.2021.3539

Delgadillo, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Behaviour Research and Therapy*, *79*, 15–22. https://doi.org/10.1016/j.brat.2016.02.003

Delgadillo, J., Moreea, O., Murphy, E., Ali, S., & Swift, J. K. (2015). Can low-cost strategies improve attendance rates in brief psychological therapy? Double-blind randomized controlled trial. *Journal of Clinical Psychology*, *71*, 1139–1152. https://doi.org/10.1002/jclp.22228

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

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Fritzsche, A., Schlier, B., Oettingen, G., & Lincoln, T. M. (2016). Mental contrasting with implementation intentions increases goal-attainment in individuals with mild to moderate depression. *Cognitive Therapy and Research*, *40*, 557–564. https://doi.org/10.1007/s10608-015-9749-6

Gaynor, S.T,, Lawrence, P.S,, & Nelson-Gray, R.O. (2006). Measuring homework compliance in cognitive-behavioral therapy for adolescent depression: review, preliminary findings, and implications for theory and practice. *Behavior Modification, 30,* 647-72. doi: 10.1177/0145445504272979.

Gaskell, C., Simmonds-Buckley, M., Kellett, S., Stockton, C., Somerville, E., Rogerson, E., & Delgadillo, J. (2023). The effectiveness of psychological interventions delivered in routine practice: systematic review and meta-analysis. *Administration Policy and Mental Health, 50(1):* 43-57. doi: 10.1007/s10488-022-01225-y.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*, 493–503. https://doi.org/10.1037/0003-066X.54.7.493

Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: Lessons from year one of IAPT. *Behaviour Research and Therapy*, *51*, 597–606. https://doi.org/10.1016/j.brat.2013.06.004

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*, *9*(3), 329. https://doi.org/10.1093/clipsy.9.3.329

Hopko, D. R., Lejuez, C. W., Ruggiero, K. J., & Eifert, G. H. (2003). Contemporary behavioral activation treatments for depression: Procedures, principles, and progress. *Clinical Psychology Review*, *23*(5), 699-717. https://doi.org/10.1016/S0272-7358(03)00070-9

Hopko, D. R., Magidson, J. F., & Lejuez, C. W. (2011). Treatment failure in behavior therapy: Focus on behavioral activation for depression. *Journal of Clinical Psychology*, *67*, 1106–1116. https://doi.org/10.1002/jclp.20840

Houghton, S., Curran, J., & Saxon, D. (2008). An uncontrolled evaluation of group behavioural activation for depression. *Behavioural and Cognitive Psychotherapy*, *36*, 235–239. https://doi.org/10.1017/S1352465808004207

Improving Access to Psychological Therapies (2019). *The Improving Access to Psychological Therapies Manual: Appendices and helpful resources.* The National Collaborating Centre for Mental Health. https://www.england.nhs.uk/publication/the-improving-access-to-psychological-therapies-manual/

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Jaycox, L. H., Miranda, J., Meredith, L. S., Duan, N., Benjamin, B., & Wells, K. (2003). Impact of a primary care quality improvement intervention on use of psychotherapy for depression. *Mental Health Services Research*, *5*, 109–120. https://doi.org/10.1023/A:1023233612022

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(6):* 608-20. doi: 10.1016/j.cpr.2010.04.001

Kazantzis, N., Whittington, C., & Dattilio, F. (2010). Meta-analysis of homework effects in cognitive and behavioral therapy: A replication and extension. *Clinical Psychology: Science and Practice*, *17*, 144–156. https://doi.org/10.1111/j.1468-2850.2010.01204.x

Kazantzis, N,, Deane, F. P., & Ronan, K.R. (2000). Homework assignments in cognitive and

behavioral therapy: A meta-analysis. *Clinical Psychology: Science and Practice*, 7(2), 189–202. doi: 10.1093/clipsy.7.2.189

Kellett, S., Newman, D., Matthews, L., & Swift, A. (2004). Increasing the effectiveness of large group format CBT via the application of practice-based evidence. *Behavioural and Cognitive Psychotherapy*, *32*, 231–234. https://doi.org/10.1017/S1352465804001213

Kellett, S., Simmonds-Buckley, M., Bliss, P., & Waller, G. (2017). Effectiveness of group behavioural activation for depression: A pilot study. *Behavioural and Cognitive Psychotherapy*, *45*, 401–418. https://doi.org/10.1017/S1352465816000540

Kellett, S., Wakefield, S., Simmonds-Buckley, M., & Delgadillo, J. (2021). The costs and benefits of practice-based evidence: Correcting some misunderstandings about the 10-year meta-analysis of IAPT studies. *British Journal of Clinical Psychology*, *60*, 42-47. https://doi.org/10.1111/bjc.12268

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*, 606–613. https://doi.org/10.1046/J.1525-1497.2001.016009606.X

Kwan, B. M., Dimidjian, S., & Rizvi, S. L. (2010). Treatment preference, engagement, and clinical improvement in pharmacotherapy versus psychotherapy for depression. *Behaviour Research and Therapy*, *48*, 799–804. https://doi.org/10.1016/j.brat.2010.04.003

Lachin, J. M. (2016). Fallacies of last observation carried forward analyses. *Clinical Trials*, *13*, 161–168. https://doi.org/10.1177/1740774515602688

Lambert, M. J. (2017). Maximizing psychotherapy outcome beyond evidence-based medicine. *Psychotherapy and Psychosomatics*, *86*, 80–89. https://doi.org/10.1159/000455170

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

Lucock, M., Bartys, S., Cupac, J., Delgadillo, J., Denton, C., Gaines, S., McMillan, D., Prestwich, A., & Stebbings, R. (2018). Using implementation intentions to prevent relapse after psychological treatment for depression – the SMArT intervention. *Behavioural and Cognitive Psychotherapy*, *46*, 626–632. https://doi.org/10.1017/S1352465818000255

Ly, K. H., Topooco, N., Cederlund, H., Wallin, A., Bergström, J., Molander, O., Carlbring, P., & Andersson, G. (2015). Smartphone-supported versus full behavioural activation for depression: A randomised controlled trial. *PLOS ONE*, *10*, e0126559. https://doi.org/10.1371/journal.pone.0126559

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

Martell, C. R., Dimidjian, S., & Herman-Dunn, R. (2010). *Behavioral Activation for depression: A clinicians guide.* Guilford Publications.

Minami, T., Serlin, R. C., Wampold, B. E., Kircher, J. C., & Brown, G. J. (2008). Using clinical trials to benchmark effects produced in clinical practice. *Quality and Quantity*, *42*, 513-525. https://doi.org/10.1007/s11135-006-9057-z

Mausbach, B. T., Moore, R., Roesch, S., Cardenas, V., & Patterson, T.L. (2010). The relationship between homework compliance and therapy outcomes: An updated meta-analysis. *Cognitive Therapy and Research*, 34(5), 429–438. doi: 10.1007/s10608-010-9297-z

Mueller, M., & Pekarik, G. (2000). Treatment duration prediction: Client accuracy and its relationship to dropout, outcome, and satisfaction. *Psychotherapy*, *37*, 117–123. https://doi.org/10.1037/h0087701

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The Work and Social

Adjustment Scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry*, *180*, 461–464. https://doi.org/10.1192/bjp.180.5.461

Muthen, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, *25*, 267–316.

National Institute for Health and Clinical Excellence. (2018). *Depression in adults: Recognition and management (update).* (Clinical Guideline 90). London. Available at: https://www.nice.org.uk/guidance/cg90/chapter/1-Guidance.

Otte, C., Gold, S. M., Penninx, B. W., Pariante, C. M., Etkin, A., Fava, M., Mohr, D. C., & Schatzberg, A. F. (2016). Major depressive disorder. *Nature Reviews Disease Primers*, *2*, 1–21. https://doi.org/10.1038/nrdp.2016.65

Portela, M. C., Pronovost, P. J., Woodcock, T., Carter, P., & Dixon-Woods, M. (2015). How to study improvement interventions: A brief overview of possible study types. *Postgraduate Medical Journal*, *91*, 343–354. https://doi.org/10.1136/postgradmed-2014-003620rep

Pott, S. L., Delgadillo, J., & Kellett, S. C. (2021). Is behavioral activation an effective and acceptable treatment for co-occurring depression and substance use disorders? A meta-analysis of randomized controlled trials. *Journal of Substance Abuse Treatment*, 108478. https://doi.org/10.1016/j.jsat.2021.108478

Power N., Noble, L. A., Simmonds-Buckley, M., Kellett, S., Stockton, C., Firth, N., & Delgadillo, J. (2022). Associations between treatment adherence-competence-integrity (ACI) and adult psychotherapy outcomes: A systematic review and meta-analysis. *Journal of Consulting and Clin Psychology, 90(5),* 427-445. doi: 10.1037/ccp0000736.

Richards, D. A., Ekers, D., McMillan, D., Taylor, R. S., Byford, S., Warren, F. C., Barrett,

B., Farrand, P. A., Gilbody, S., Kuyken, W., O'Mahen, H., Watkins, E., Wright, K., A., Hollon, S. D., Reed, N., Rhodes, S., Fletcher, E., & 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

Robinson, L., Delgadillo, J., & Kellett, S. (2020). The dose-response effect in routinely delivered psychological therapies: A systematic review. Psychotherapy Research, 30(1), 79-96. doi: 10.1080/10503307.2019.1566676.

Roth, A. D, & Pilling, S. (2008). Using an evidence-based methodology to identify the competences required to deliver effective cognitive and behavioural therapy for depression and anxiety disorders. Behavioural and Cognitive Psychotherapy, 36, 129-47.

Sheeran, P., & Webb, T. L. (2016). The intention–behavior gap. *Social and Personality Psychology Compass*, *10*, 503–518. https://doi.org/10.1111/spc3.12265

Simmonds-Buckley, M., Kellett, S., & Waller, G. (2019). Acceptability and efficacy of group behavioral activation for depression among adults: A meta-analysis. *Behavior Therapy*. https://doi.org/10.1016/j.beth.2019.01.003

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*, 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Strauman, T. J., Eddington, K. M., & McCrudden, M. C. (2007). Affective science and psychotherapy: in search of synergy. In J. Rottenberg & S. L. Johnson (Eds.), *Emotion and psychopathology: Bridging affective and clinical science* (pp. 285–303). American Psychological Association. https://doi.org/10.1037/11562-013

Swift, J. K., & Callahan, J. L. (2011). Decreasing treatment dropout by addressing expectations for treatment length. *Psychotherapy Research*, *21*, 193–200. https://doi.org/10.1080/10503307.2010.541294

Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *80*, 547-559. https://doi.org/10.1037/a0028226

Toli, A., Webb, T. L., & Hardy, G. E. (2016). Does forming implementation intentions help people with mental health problems to achieve goals? A meta-analysis of experimental studies with clinical and analogue samples. *British Journal of Clinical Psychology*, *55*, 69–90. https://doi.org/10.1111/bjc.12086

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*, 859–883. https://doi.org/10.1037/0033-2909.133.5.859

Uphoff, E., Ekers, D., Dawson, S., Richards, D., & Churchill R. (2019). Behavioural activation therapies for depression in adults. *Cochrane Database Systematic Review, 12;2019(4):*CD013305. doi: 10.1002/14651858.CD013305.

van Bokhoven, M. A., Kok, G., & van der Weijden, T. (2003). Designing a quality improvement intervention: A systematic approach. *Quality and Safety in Health Care*, *12*, 215–220. https://doi.org/10.1136/qhc.12.3.215

Wang, G., Wang, Y., & Gai, X. (2021). A meta-analysis of the effects of mental contrasting with implementation intentions on goal attainment. *Frontiers in Psychology*, *12*, 565202. https://doi.org/10.3389/fpsyg.2021.565202

Wampold, B. E. (2015). Routine outcome monitoring: Coming of age-with the usual developmental challenges. *Psychotherapy*, *52*, 458–462. https://doi.org/10.1037/pst0000037

Webb, T. L., & Sheeran, P. (2008). Mechanisms of implementation intention effects: The role of goal intentions, self-efficacy, and accessibility of plan components. *British Journal of Social Psychology*, *47*(3), 373-395. https://doi.org/10.1348/014466607X267010

Wilson, G. T. (1996). Manual-based treatments: The clinical application of research findings. *Behaviour Research and Therapy*, *34*, 295–314. https://doi.org/10.1016/0005-7967(95)00084-4

Table 1; attendance during BAG and BAG+

Note: BAG; Behavioural Activation in Groups (existing intervention cohort), BAG+; Behavioural Activation in Groups (enhanced intervention cohort).

Attender status	BAG (n = 31)	BAG+ (n = 31)	Chi-squared (p value)	Odds ratio (BAG+:BAG)
Treatment completers (8 sessions)	3 (10%)	4 (13%)	0.16 (p=.344)	1.38
Partial attenders (4-7 sessions)	15 (48%)	18 (58%)	0.58 (p=.223)	1.48
Drop-outs (1-3 sessions)	13 (42%)	9 (29%)	1.13 (p=.144)	0.57

Table 2; *means, standard deviations (SD) and effect sizes (d) for BAG and BAG+*

	BAG (n = 31)	BAG+ (n = 31)	Between-group d (BAG vs BAG+)	T-score
Primary outcome				
PHQ-9				
Pre-treatment mean (SD)	18.7 (4.1)	18.4 (4.0)	0.07	0.28 (p=.390)
Post-treatment mean (SD)	15.5 (5.1)	13.0 (6.4)	0.43	1.69 (p=.049)
Pre-post mean change (SD)	-3.2 (4.8)	-5.4 (5.3)	0.43	1.71 (p=.047)
Pre-post d	0.78 (r=.50)	1.35 (r=.55)	-	
Secondary outcomes				
GAD-7				
Pre-treatment mean (SD)	14.6 (4.2)	13.0 (5.0)	0.35	1.35 (p=.091)
Post-treatment mean (SD)	12.7 (4.9)	9.7 (5.1)	0.60	2.38 (p=.011)
Pre-post mean change (SD)	-1.9 (3.6)	-3.3 (4.5)	0.36	1.41 (p=.082)
Pre-post d	0.45 (r=.69)	0.66 (r=.62)	-	
WSAS				
Pre-treatment mean (SD)	24.6 (8.8)	25.0 (8.1)	-0.04	-0.17 (p=.435)
Post-treatment mean (SD)	20.3 (8.9)	19.2 (10.9)	0.10	0.41 (p=.342)
Pre-post mean change (SD)	-4.4 (6.2)	-5.9 (10.6)	0.19	0.71 (p=.240)
Pre-post d	0.49 (r=.75)	0.72 (r=.42)	-	

Note: Pre-post effect sizes (d) have been calculated by dividing the pre-post difference by the pre-SD as recommended by Minami et al. (2008) (for reference the correlation (r) between pre-post scores is reported in brackets). BAG; Behavioural Activation in Groups (existing intervention cohort), BAG+; Behavioural Activation in Groups (enhanced intervention cohort), PHQ-9; Patient health questionnaire, GAD-7 generalized anxiety disorder scale, WSAS; work and social adjustment scale.

Table 3; *individual outcomes for BAG and BAG+*

Post-treatment PHQ-9 recovery status	BAG (*n* = 31)	BAG+ (*n* = 31)	Chi-squared (p value)	Odds ratio (BAG+:BAG)
Recovered	10% (3)	29% (9)	3.72 (*p*=.027)	3.82
Improved	23% (7)	48% (15)	4.51 (*p*=.002)	3.21
Nonresponse	77% (24)	52% (16)	4.51 (*p*=.002)	0.31
Deteriorated	0% (0)	0% (0)	-	-

Note. 'Recovered' represents the proportion of patients who showed clinically significant change in addition to reliable improvement. Therefore, 'Recovered' and 'Improved' categories are not mutually exclusive. BAG; Behavioural Activation in Groups (existing intervention cohort), BAG+; Behavioural Activation in Groups (enhanced intervention cohort), PHQ-9; Patient health questionnaire.

Figure 1; STROBE flow diagram of patient selection

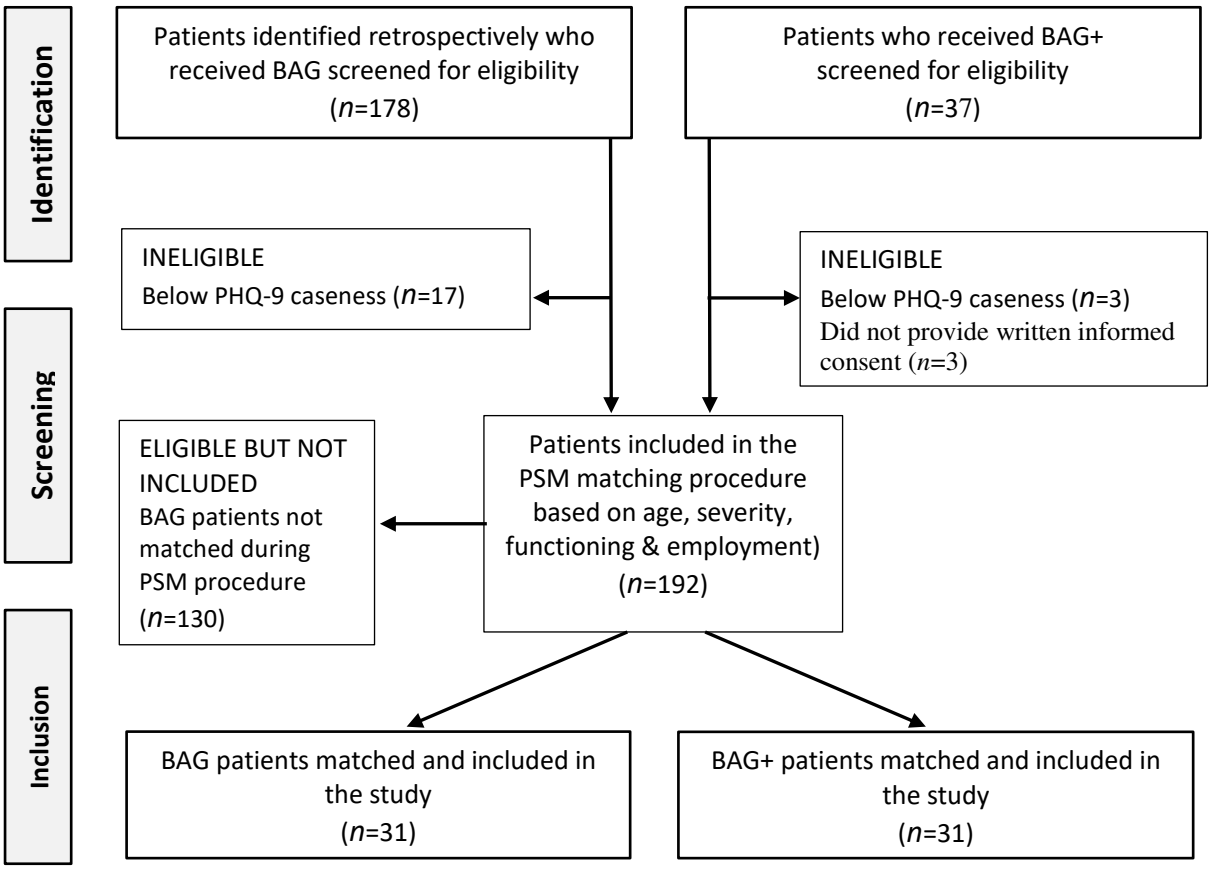
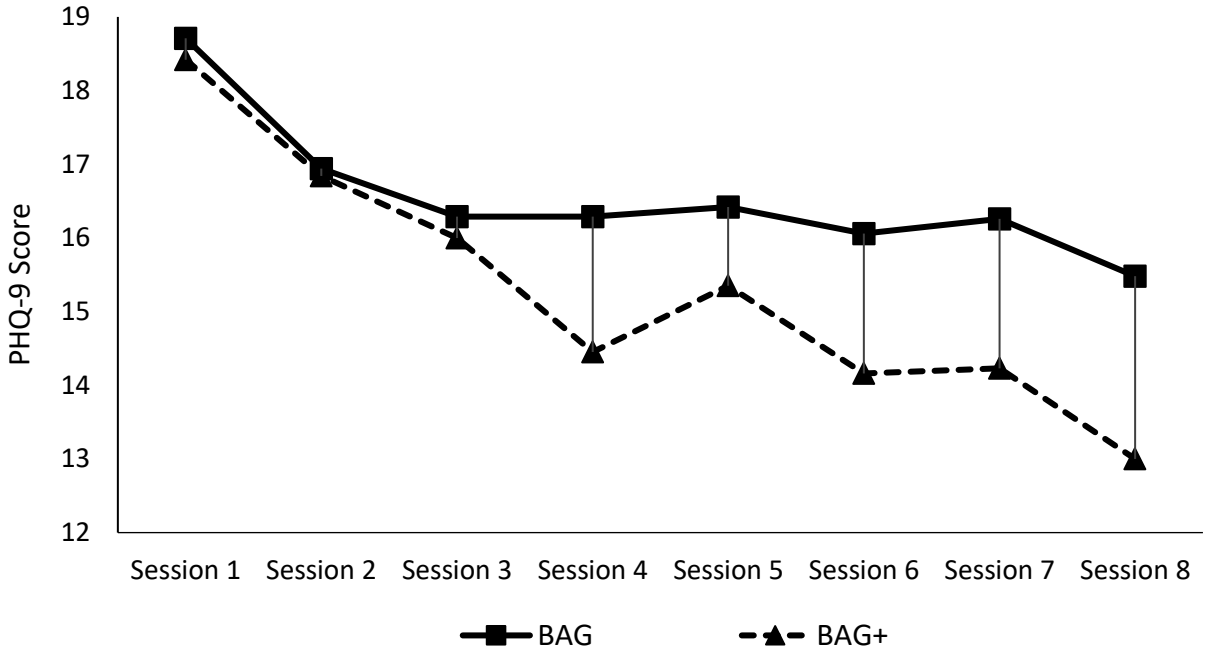


Figure 2; session-by-session PHQ-9 scores during BAG and BAG+

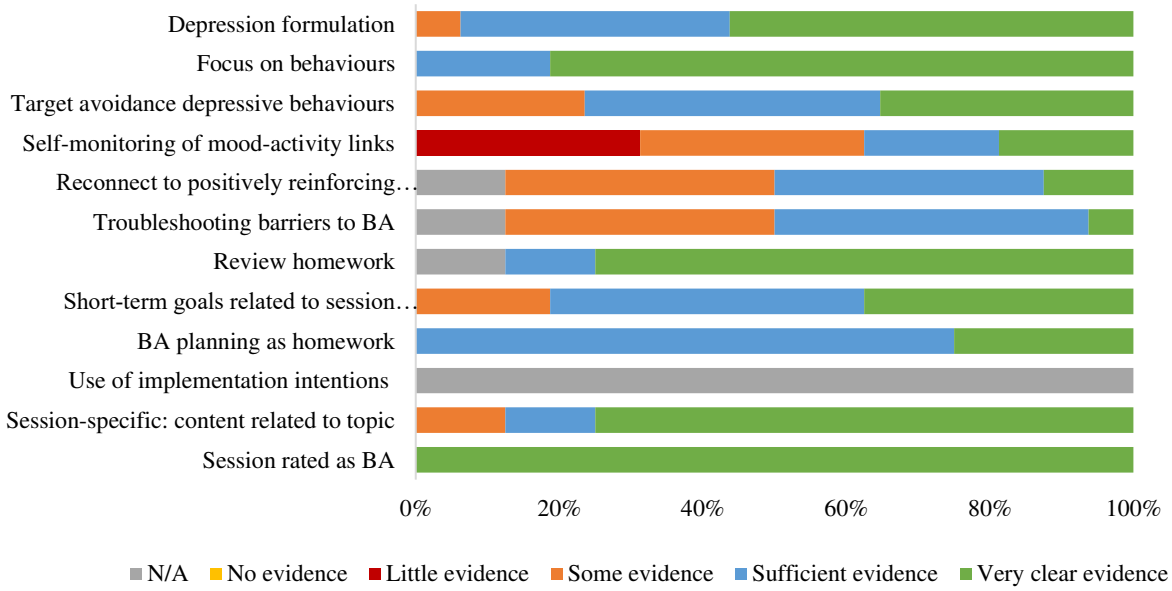


Supplementary Material

Treatment Adherence

Summary of the adherence checks for BAG (Figure S1) and BAG+ (Figure S2) according to each category. The mean rating for the presence of evidence in each category is presented below the legend in each figure.

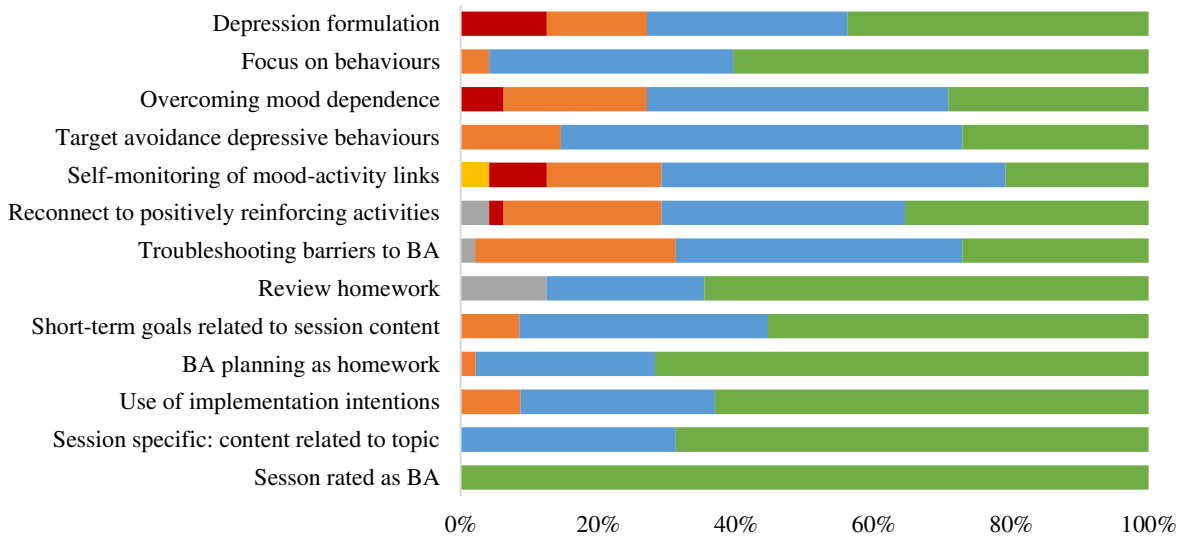
Figure S1 Adherence to the BAG Treatment Protocol (N=3)



	N/A	No evidence	Little evidence	Some evidence	Sufficient evidence	Very clear evidence
Category mean rating	12%*	0%	3%	14%	29%	44%

*Expect use of implementation intentions to be rated N/A

Figure S2 Adherence to the BAG+ Treatment Protocol (N=8).



Category mean rating	1%	0%	2%	11%	34%	51%

Sample Demographics

Table S1 presents the summary of sample demographics and clinical features of the entire sample who received BAG prior to matching and according to BAG/BAG+ cohort.

Table S1 Demographics and Clinical Features of Entire Sample, Unmatched BAG sample and the BAG+ Sample.

	Entire sample	Unmatched BAG	BAG+
Demographics	(*N*=192)	(*n*=161)	(*n*=31)
Mean age (SD)	39.0 (16.1)	38.5 (16.3)	41.8 (14.9)
Gender			
% Female (*n*)	54% (104)	54% (87)	55% (17)
% Male (*n*)	46% (88)	46% (74)	45% (14)
Ethnicity			
% White (*n*)	87% (166)	86% (139)	87% (27)
% Minority group (*n*)	5% (11)	6% (10)	3% (1)
% Missing	8% (15)	8% (12)	10% (3)
Median IMD decile (1-10)	5	4	6
Employment status			
% Employed (*n*)	28 % (54)	26% (41)	42% (13)
% Other (*n*)	71% (136)	73% (118)	58% (18)
% Missing (*n*)	1% (2)	1% (2)	0% (0)
Clinical features			
Mean PHQ-9 score (SD)	17.5 (4.2)	17.3 (4.3)	18.4 (4.0)
Mean GAD-7 score (SD)	12.9 (4.8)	12.9 (4.8)	13.0 (5.0)

Mean WSAS score (SD)	23.1 (7.5)	22.7 (7.3)	24.9 (8.1)

Table S2 Characteristics of the BAG and BAG+ samples.

Pre-treatment characteristic	BAG (*n*=31)	BAG+ (*n*=31)	X^2 (*p* value)
Gender (% female)	55%	55%	0.00 (*p*=1.00)
Ethnicity (% White British)	87%	87%	8.00 (*p*=.156)
IMD deciles 1-10 (median)	5	6	11.46 (*p*=.246)
Antidepressant use	NA	69%	-
Depression severity (PHQ-9)			
Moderate	16%	13%	1.04 (*p*=.596)
Moderately severe	42%	55%	
Severe	42%	32%	
Anxiety severity (GAD-7)			
Minimal	0%	7%	
Mild	16%	19%	2.29 (*p*=.515)
Moderate	39%	36%	
Severe	45%	39%	
Impaired functioning (WSAS)			
Subclinical (<10)	10%	3%	
Significant impairment but less severe symptomology (10-20)	22%	19%	1.27 (*p*=.528)
Moderate/severe psychopathology (>20)	68%	77%	

Note: NA: not available in the historical data accessed for BAG cohort. BAG; Behavioural Activation in Groups (existing intervention cohort), BAG+; Behavioural Activation in Groups (enhanced intervention cohort), IMD; Index of multiple deprivation, PHQ-9; Patient health questionnaire, GAD-7 Generalized anxiety disorder scale, WSAS; Work and social adjustment scale.

Table S3 Treatment overview and session outlines for BAG and BAG+ (Augmentations in Italics)

Session	Common components across BAG and BAG+: Content described using patient wording (Homework activity)	Additional BAG+ components: *Dose-response psychoeducation and implementation intentions augmentations*
1	Learn your patterns and start to change them (Homework: Activity-mood diary)	*Dose-response psychoeducation sheet in group workbook and verbally reiterated at session 1 by facilitators.* *Additional homework task: Read psychoeducation sheet.* *'Achieving your goals' implementation intention information sheet added to group workbook and 'if-then' planning modelled by facilitators at session 1.*
2	Values: the guide to who we are (Homework: Committed-action exercise)	
3	Getting out of the TRAPs and back on TRAC (Homework: Apply TRAP/TRAC handouts to tasks)	
4	Taking action: a problem solving approach (Homework: Problem solving to change unhelpful behaviours)	*At each group session use of specific 'if-then plan' worksheet in group workbook to plan and agree HW consistent with that session content.* *All 'if-then' plan silently repeated 3 times and once out loud to a co-partner in the group*
5	Identifying unhelpful thinking, worry and rumination (Homework: Monitor rumination and use 'two-minute rule')	
6	Developing responses to thinking, rumination and worry (Homework: RCA, mindfulness and self-soothing handouts)	
7	Making changes one step at a time (Homework: 'Short-term goals' planning worksheet)	
8	Building the relationships you want/tying it all together (Homework: Apply ACTION to everyday situations)	

Note: The BAG programme outlined in the table originated in the Sheffield Improving Access to Psychological Therapies (IAPT) service, and was adapted from Martell et al.'s (2001) BA model and supplemented with the Martell et al., (2010) update. *Abbreviations:* HW; Homework task, TRAP: Trigger, Response, Avoidance Pattern; TRAC: Trigger, Response, Alternative Coping., RCA: rumination cues action, ACTION; assess, choose, try, integrate, observe, never give up.

Table S4; *Comparison of baseline covariates and variances in BAG and BAG+ for the unmatched sample and PSM sample.*

Baseline Covariate	BAG Mean (SD)/ Frequency (%)	BAG+ Mean (SD)/ Frequency (%)	Standardized Difference/ Proportion	Variance (BAG+)	Variance (BAG)	Ratio: BAG to BAG+	Ratio diff.
Unmatched sample	(*n*=161)	(*n*=31)					
Age	38.5 (16.3)	41.8 (14.9)	0.20	221.38	264.69	0.84	0.16
PHQ-9 score	17.3 (4.3)	18.4 (4.0)	0.26	16.05	18.19	0.88	0.12
WSAS score	22.7 (7.3)	24.9 (8.1)	0.32	66.17	17.08	0.94	0.06
Employment status			0.37	-	-	-	-
Employed	41 (26%)	13 (42%)					
Other	120 (75%)	18 (58%)					
Matched sample	(*n*=31)	(*n*=31)					
Age	42.6 (14.3)	41.8 (14.9)	**-0.06**	221.38	202.91	1.09	0.09
PHQ-9 score	18.7 (4.1)	18.4 (4.0)	**-0.07**	16.05	52.93	1.25	0.25
WSAS score	24.6 (8.8)	24.9 (8.1)	**0.04**	66.17	77.85	0.85	0.15
Employment status			**0.00**	-	-	-	-
Employed	13 (42%)	13 (42%)					
Other	18 (58%)	18 (58%)					

Note: For continuous covariates mean and SD are presented; for categorical covariate frequencies and percentages are presented. **Bold:** Standardized differences in sample covariates <0.1 deemed representative of minimal difference between groups. Variances are provided for the unmatched and matched BAG+ and BAG samples for continuous covariates. Abbreviations: BAG; Behavioural Activation in Groups (existing intervention cohort), BAG+; Behavioural Activation in Groups (enhanced intervention cohort), PHQ-9; Patient health questionnaire, WSAS; Work and social adjustment scale

Figure S3; *Comparison of baseline continuous covariates in BAG and BAG+ in the overall unmatched sample and after PSM matching procedure*