The Cognitive Behaviour Therapist (2025), vol. 18, e3, page 1 of 16
doi:10.1017/S1754470X24000412

BABCP
www.babcp.com

ORIGINAL RESEARCH

Improving outcomes via treatment augmentations to behavioural activation for depression in routine practice: a cohort comparison study

Melanie Simmonds-Buckley[1,2], Stephen Kellett[1,2], Jennie Hague[2] and Glenn Waller[2]

[1]Rotherham Doncaster and South Humber NHS Foundation Trust, UK and [2]University of Sheffield, Sheffield, UK
Corresponding author: Stephen Kellett; Email: stephen.kellett@nhs.net

(Received 2 November 2023; revised 17 September 2024; accepted 23 October 2024)

Abstract

Whilst behavioural activation (BA) is an empirically supported treatment for depression, some patients do not benefit. The aim of this study was to evaluate the effectiveness of two treatment augmentations to an extant manualized 8-session group version of BA. The two treatment augmentations were (a) dose–response psychoeducation to improve attendance and (b) implementation intentions to improve clinical outcomes. A cohort comparison design in routine practice comparing standard 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). There was no effect of the two treatment augmentations on attendance or in overall mean reductions to depression (mean improvement difference = 2.2), anxiety (mean improvement difference = 1.9) or impaired functioning (mean improvement difference = 1.5). Rates of reliable improvement in depression were significantly higher for augmented BA (odds ratio = 3.21 for BA+ compared with BA). Efforts should be made to still improve outcomes for empirically supported interventions, with any treatment augmentations tested in well-controlled studies.

Key learning aims

(1) To learn about the utility of adapting existing treatments as opposed to developing new treatments.
(2) To learn about the potential of propensity score matching in the analysis of routinely collected datasets.
(3) To learn about delivery of behavioural activation in groups.
(4) To better understand how to enhance and evaluate treatment protocols using theoretically informed and low-cost treatment augmentations.

Keywords: behavioural activation; effectiveness; Improving Access to Psychological Therapies (IAPT) programme

Introduction

When empirically supported interventions are delivered in routine practice there are marked differences in effectiveness compared with clinical trials (Gyani *et al.*, 2013; Hansen *et al.*, 2002), with outcomes up to 12% lower in routine practice, and this occurs particularly when patients do not receive evidence-based interventions (Barkham *et al.*, 2008; Hansen *et al.*, 2002). The recent Gaskell *et al.* (2023) meta-analysis of the outcomes achieved in routinely delivered interventions speculated that such differences were explained by poor therapist attitudes to protocol-delivered interventions. A primary challenge of translational science in psychotherapy is therefore enabling

patients in routine services not to be disadvantaged compared with participants in clinical trials (Strauman *et al.*, 2007). The three main efforts to enable better translation have included (a) clear treatment guidelines to ensure that only empirically supported interventions are recommended and delivered in routine practice (American Psychiatric Association, 2010; National Institute for Health and Clinical Excellence, 2018), (b) routine outcome monitoring tracking sessional outcomes and so enabling timely remedial actions by therapists (Delgadillo *et al.*, 2018; Delgadillo *et al.*, 2022; Lambert, 2017; Wampold, 2015) and (c) enhancing treatment competency through effective training, regular clinical supervision, use of clear treatment manuals and routine auditing of treatment integrity (Bambling *et al.*, 2006; Power *et al.*, 2022; Wilson, 1996).

A lesser used approach to improve translation is the augmentation of existing treatments through enhancements that act on the mechanisms of change in a therapy, or manipulating factors known to be associated with outcome in a therapy. This approach is time efficient as it avoids the need to develop 'new' therapies and rather tries to enhance extant empirically supported therapies via augmentation. Examples of treatment augmentations have been achieved through a variety of practical-technological (e.g. automatic text messaging between sessions to increase engagement, Aguilera *et al.*, 2017; apps to support the delivery of depression interventions, Bae *et al.*, 2023) and theoretical innovations (e.g. see Oldham *et al.*, 2012, for a review). These changes have mostly been implemented as adjuncts to treatment and few studies have tested the effectiveness of integrating augmentations directly into the treatment content and delivery.

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; Pott *et al.*, 2021; Richards *et al.*, 2016; Simmonds-Buckley *et al.*, 2019). Behavioural activation is a time-limited psychotherapeutic approach that aims to change the manner in 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 also (3) understanding and then addressing any apparent blocks to activation (Uphoff *et al.*, 2019). Recovery rates, however, indicate at least 40% of BA patients do not experience a statistically clinically significant and reliable change on depression outcome measures (Hansen *et al.*, 2002; Hopko *et al.*, 2011), indicating the need to test treatment augmentations. It is acknowledged that patients may meet their idiosyncratic treatment goals during BA, whilst not reaching the statistical threshold for recovery 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). Augmentation should target key facilitators of change (van Bokhoven *et al.*, 2003) and in the context of BA *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 drop out of depression treatment due to the debilitating and demotivating impact of low mood, wider systemic factors and poor treatment fit (Barrett *et al.*, 2008). Treatment acceptability also suffers when there is a discrepancy between patient expectations about rate of change and the required number of sessions (Swift and Callahan, 2011). When patient expectations of the duration of therapy differ from the service offer, then the likelihood of drop-out increases (Callahan *et al.*, 2009; Mueller and Pekarik, 2000). Psychoeducation on dose–response evidence has been suggested as a method to align patient expectations about treatment duration to a dose that would most likely invoke meaningful symptom improvement (Swift and Callahan, 2011). However, there are mixed findings for impact of such psychoeducation on eventual treatment attendance. Swift and Callahan (2011) found pre-treatment dose–response leaflets encouraged patients to remain in treatment for longer, but Delgadillo *et al.* (2015) found role induction orientation leaflets had no

https://doi.org/10.1017/S1754470X24000412 Published online by Cambridge University Press

impact on attendance. Both studies delivered the psychoeducation prior to patients attending treatment, but fidelity checks were not completed. Kellett *et al.* (2017) called for efforts to improve the acceptability of group BA. Providing psychoeducation on dose–response to improve attendance was therefore the first treatment augmentation in the current study.

The main change method of BA is the agreement and completion of activation activities as 'homework' (Hopko *et al.*, 2011). Patients' engagement with such activities is crucial in generating a sense of progress and drives positive change (Beck and 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 finds mostly modest effects ($r = .22$ in Kazantzis *et al.*, 2000; $r = .26$ in Mausbach *et al.*, 2010). Whilst homework may be planned, it does not guarantee successful completion, so creating an unhelpful 'intention–behaviour gap' (Sheeran and Webb, 2016) that would maintain depression. Failure to engage in activation has been identified as a contributing factor to non-response during BA (Hopko *et al.*, 2011). Implementation intentions have been shown to close the intention–behaviour gap and so increase goal attainment (Wang *et al.*, 2021). This technique involves the generation of specific plans about how, when and where goals will be acted upon, and these are crystallized using brief 'if–then' formats (Gollwitzer, 1999). Establishing 'if–then' plans link intended actions to environmental cues and in doing so removes the need for unhelpful procrastination (Webb and Sheeran, 2008). Implementation intentions are acceptable to patients (Lucock *et al.*, 2018) and 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 a contingency-management strategy (Kanter *et al.*, 2010). Integrating implementation intentions into planning and completing homework during BA was therefore the second treatment augmentation in the current study.

To summarize, few empirical studies have used translational science approaches to test the effectiveness of within-treatment augmentations to improve depression outcomes in routine practice (Portela *et al.*, 2015). The main aim of this cohort comparison study was therefore to test whether augmented group BA would have better attendance rates and improved clinical outcomes compared with standard group BA.

Method

Design and setting

The study was conducted in an NHS Talking Therapies (NHS TT) for Anxiety and Depression service (i.e. previously called Improving Access to Psychological Therapies). A cohort comparison design compared 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 TT service (see Clark, 2018, for full description of the TT stepped care approach in the English NHS). A sample size analysis using G*Power (Faul *et al.*, 2007) indicated $n = 32$ was needed in each study group (total $N = 64$) to detect a small to medium effect ($f = 0.18$) for differences in mean reductions in depression scores measured by the interaction effect (time × BAG condition) in a repeated measures between-subjects ANOVA with .80 power at $p = .05$. Retrospective anonymized routine outcome data from patients who had previously received standard BAG therefore formed the historical control. Samples were matched using propensity score matching (PSM) as this method enables cohort comparisons to mimic the features of an RCT through balancing pre-treatment covariates (Austin, 2011; see 'Data analysis plan' section below for full details).

https://doi.org/10.1017/S1754470X24000412 Published online by Cambridge University Press

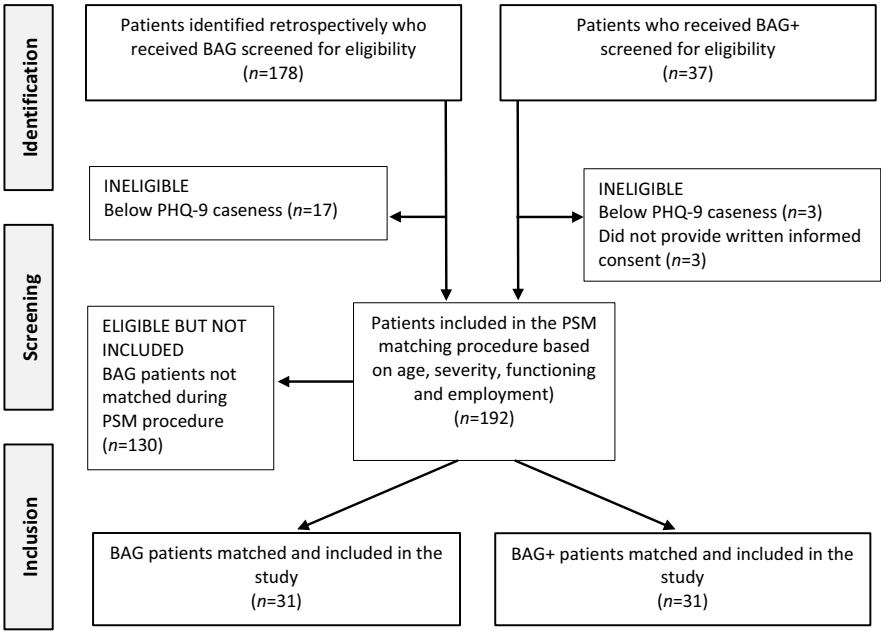

Figure 1. STROBE flow diagram of patient selection.

Participants

The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) diagram summarizing patient flow and sample selection is presented in Fig. 1. Inclusion criteria were: (a) seeking treatment for 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 on the Patient Health Questionnaire-9 (PHQ-9) prior to commencing treatment (i.e. a PHQ-9 score <10). Out of 34 patients who attended BAG+ across three groups, 31 met caseness criteria and so had their outcomes included in the analysis. Out of 178 patients who had attended standard BAG (across 22 groups), 161 met the inclusion criteria. PSM was then used to match 31 patients from the available pool of 161 standard BAG patients to the 31 eligible BAG+ patients in order to ensure the clinical equivalence of baseline assessments (i.e. total $n = 62$). Therefore, the final sample was slightly under the target sample size of $n = 64$.

Outcome measures

Outcome measures consisted of the NHS TT minimum dataset (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. In both cohorts, outcome measures were completed at the start of every group session. Patients receiving BAG+ completed a demographic information sheet capturing age, gender, ethnicity, current anti-depressant medication and previous episodes and associated treatment of depression. Anonymized clinical outcome measures and demographic information for the standard BAG cohort was retrieved and collated from routinely collected service data.

https://doi.org/10.1017/S1754470X24000412 Published online by Cambridge University Press

Group behavioural activation; facilitation, delivery and adherence

BAG or BAG+ groups were facilitated by two British Association for Behavioural and Cognitive Psychotherapies (BABCP) accredited CBT therapists. A total of eight facilitators (i.e. one male and seven female therapists) were in the study and all had delivered BAG before they delivered BAG+. When any group was set up, the same two facilitators then delivered all sessions. All therapists had completed the same BABCP accredited 1-year Postgraduate CBT training programme and this had a 2-day BA workshop. Length of time qualified varied from 2 to 6 years. All therapists attended a quarterly BAG peer supervision group and had 1-hour fortnightly one-to-one clinical supervision. Supervision time was therefore matched between BAG and BAG+. A 1-hour BAG+ training intervention was provided for the study therapists, and this introduced the dose–effect psychoeducation and therapists practised helping patients set implementation intentions. The training evaluation questionnaire 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, 2-hour manualized sessions based on an extant treatment protocol and patient workbook (Houghton *et al.*, 2008; Martell *et al.*, 2010).

All groups were delivered in the same primary care setting. Apart from the treatment augmentations, group interventions were matched in terms of content and time. The Supplementary material summarizes the protocol, and details how and where the treatment augmentations were integrated. The first treatment augmentation was a data-informed psychoeducation enhancement targeted at increasing attendance. The psychoeducation consisted of dose–effect evidence taken from a pilot BAG outcome study (Kellett *et al.*, 2017). The psychoeducation was included in a pre-treatment information pack stating that: (1) attending at least four 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 and use 'implementation intentions' when planning homework at the end of each BAG+ 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 BAG+ session. Patients worked in pairs during BAG+ groups to form implementation intentions using worksheets at each session for their idiosyncratic homework assignments. Patients were instructed to silently repeat their homework implementation intention to themselves three times, then repeated it out loud to their group partner; this is standard practice in forming implementation intentions (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 homework), a *session-specific adherence* section, and an overall assessment of whether the session was BA. An item relating to 'use of implementation intentions' was included to check adherence to BAG+. A customized page of the checklist was adapted for every 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 group therapists then independently completed the adherence checklist. BAG+ adherence was checked throughout the duration of the study and BAG adherence was checked for the delivery of the final two groups of the existing BAG protocol. Full details of adherence check for BAG and BAG+ are provided in Supplementary material (Figs S1 and S2).

All BAG and BAG+ group sessions were rated as representative of BA, indicating that patients were receiving 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%). As expected, adherence checks of the implementation intentions

https://doi.org/10.1017/S1754470X24000412 Published online by Cambridge University Press

augmentation showed they were absent during BAG. Implementation intentions were present during BAG+ and also with sufficient or very clear evidence in over 90% of group 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 and Koch, 1977).

Data analysis plan

The data analysis plan had three stages. First, the eligible BAG+ sample ($n = 31$) was matched to a comparative subsample of BAG patients ($n = 31$) using PSM, where all patients in the existing BAG archived database who met the inclusion requirements were eligible for matching ($N = 161$). Samples were matched on depression baseline severity (i.e. PHQ-9 score) and variables previously identified as predictors of depression outcomes for interventions delivered in NHS TT services (Delgadillo *et al.*, 2016) – age, baseline functioning (WSAS score) and employment status. A one-to-one, nearest neighbour 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 then post-matching. Unmatched and matched sample demographics are reported in the Supplementary material.

Secondly, to assess the impact of clustering in the data, intraclass correlation coefficients (ICCs) estimated the level of variance attributable to group level factors. ICCs and the associated design effect (DE) for all outcome measures were calculated. A DE of >2 indexes significant co-dependence and therefore unsuitability for single-level analysis (i.e. the analysis would need to be a multi-level model; Muthen and Satorra, 1995). Overall, outcomes from 13 groups were analysed (BAG = 10 and BAG+ = 3) and the average cluster size was 4.77. ICCs for PHQ-9 (–0.04), GAD-7 (–0.05) and WSAS outcomes (0.06) produced DEs of 0.85, 0.81 and 1.23, respectively. As all these DEs were <2, single-level analyses were appropriate. Outcomes were analysed using the intention-to-treat (ITT) principle, including all patients who entered group treatment. As outcomes were collected at every session, missing data were accounted for using last observation carried forward (LOCF) imputation, to align with the approach used to compute recovery metrics in routine services.

The final stage evaluated the effect of the BAG+ augmentations on attendance, clinical outcomes and recovery rates. Mean session attendance in BAG and BAG+ was calculated. Given the nature of BA and the samples, the primary outcome was depression scores (PHQ-9), with anxiety (GAD-7) and impaired functioning (WSAS) measures as secondary outcomes. Reliable and clinically significant change criteria were applied to the PHQ-9 outcomes to determine recovery rates (Jacobson and Truax, 1991). Rates were defined as 'recovery' when PHQ-9 scores moved from above to below the clinical cut-off after treatment, 'reliable improvement' when there was a reliable decrease in PHQ-9 scores (i.e. a ≥6 score decrease) and 'reliable recovery' when there was a decrease in PHQ-9 scores of ≥6 in addition to the pre–post score moving from above to below the PHQ-9 clinical cut-off. 'Reliable deterioration' occurred when there was a reliable increase (>6) in PHQ-9 scores. A 'non-response' outcome occurred when no reliable change on the PHQ-9 occurred in either direction (i.e. neither improvement nor deterioration in depression). *Post-hoc* sensitivity analyses of recovery outcomes based on the NHS TT metrics combining PHQ-9 and GAD-7 outcomes (NHS Talking Therapies, 2024) are also included in the Supplementary material. Clinical outcomes, attendance and recovery rates were compared for the BAG and BAG+ cohorts using chi-square and odds ratios for binary variables. A two-way mixed (i.e. pre–post scores via condition) analysis of variance (ANOVA), *t*-tests and Cohen's *d* within- and between-group effect sizes evaluated outcomes for continuous variables. Effect sizes of 0.2, 0.5 and 0.8 are considered small, moderate and large effects (Cohen, 1992). An exploratory longitudinal mixed-model analysis compared the trajectories of PHQ-9 scores over the

https://doi.org/10.1017/S1754470X24000412 Published online by Cambridge University Press

course of treatment for each BAG condition (condition×time interaction; full analysis details and model building are reported in the Supplementary material).

Results

There are three sections to the results: sample matching and description, evaluation of treatment acceptability, and finally the evaluation of clinical outcomes.

Sample matching and sample description

The matched dataset ($n = 62$) was checked to ensure sufficient distribution of covariates across the samples in comparison with the unmatched sample ($n = 192$). The Supplementary material contains comparisons of baseline covariates in BAG and BAG+ in the overall unmatched sample and after PSM matching and the variance and distribution of the continuous covariates before and after matching. Standardized differences evidence that imbalances in all the specified covariates across BAG and BAG+ were reduced to below the specified threshold after matching ($d<0.10$). This suggests that PSM was successful at matching BAG and BAG+ pre-intervention. The Supplementary material also describes the characteristics of the BAG and BAG+ patients included in the final sample. Categories of pre-treatment depression severity were classified as 37% ($n = 23$) severe depression, 48% ($n = 30$) moderately severe depression and 15% ($n = 9$) moderate depression. At pre-treatment, nearly 89% ($n = 55$) also met clinical caseness for anxiety [i.e. 42% ($n = 26$) severe anxiety, 37% ($n = 23$) moderate, 18% ($n = 11$) mild, and 3% ($n = 2$) minimal anxiety]. On the WSAS, 73% met caseness for impairment.

Treatment acceptability outcomes

There was no difference ($t_{60} = 0.92$, $p = .360$) in the number of sessions attended during BAG (mean 4.6 sessions, $SD = 2.6$) compared with BAG+ (mean 5.2 sessions, $SD = 2.4$). Overall, 10% ($n = 3$) of BAG patients and 13% ($n = 4$) of BAG+ patients fully attended all eight sessions; $n = 13$ (42%) during BAG and $n = 9$ (29%) during BAG+ dropped out before receiving at least four sessions (i.e. the minimal recommended dose in the attendance augmentation).

Clinical outcomes

Table 1 presents the primary and secondary outcome means, pre–post effect sizes and between-group effect sizes for BAG and BAG+. Overall, depression symptoms significantly decreased following BAG treatment (pre–post PHQ-9 main effect: $F_{1,60} = 45.22$, $p<.001$), but scores did not differ between BAG conditions (BAG condition main effect: $F_{1,60} = 1.59$, $p = .212$). The time by condition interaction found no significant difference in pre–post changes in depression scores between BAG and BAG+ ($F_{1,60} = 2.91$, $p = .093$). There was a similar pattern for overall impaired functioning, which significantly decreased after BAG treatment (pre–post WSAS main effect: $F_{1,60} = 21.24$, $p<.001$), did not differ between BAG conditions (BAG condition main effect: $F_{1,60} = 0.03$, $p = .871$) or for the time by condition interaction ($F_{1,60} = 0.40$, $p = .529$). Anxiety scores were significantly higher in the BAG condition (BAG condition GAD-7 main effect: $F_{1,60} = 4.32$, $p = .042$) and significantly decreased over time (pre–post main effect: $F_{1,60} = 25.38$, $p<.001$). Again, there were no significant differences in the interaction effect for pre–post anxiety changes between BAG and BAG+ ($F_{1,60} = 1.98$, $p = .164$).

Within-group treatment reductions in depression (PHQ-9) symptoms represented moderate to large and large effects for BAG and BAG+, respectively. The lower post-treatment depression scores for BAG+ compared with BAG were representative of a small between-groups effect ($d = 0.43$). Both BAG and BAG+ both produced small to moderate pre–post reductions in anxiety

https://doi.org/10.1017/S1754470X24000412 Published online by Cambridge University Press

Table 1. Means, standard deviations (*SD*), effect sizes (*d*) and 95% confidence intervals (CI) for BAG and BAG+

	BAG ($n = 31$)	BAG+ ($n = 31$)	Between-group *d* BAG *vs* BAG+ (95% CI)
Primary outcome			
PHQ-9			
Pre-treatment mean (*SD*)	18.7 (4.1)	18.4 (4.0)	0.07 (−0.42 to 0.57)
Post-treatment mean (*SD*)	15.5 (5.1)	13.0 (6.4)	0.43 (−0.07 to 0.94)
Pre–post mean change (*SD*)	−3.2 (4.8)	−5.4 (5.3)	0.44 (−0.07 to 0.94)
Pre–post *d* [correlation] (95% CI)	0.78 [$r = .50$] (0.38 to 1.18)	1.35 [$r = .55$] (0.86 to 1.84)	—
Secondary outcomes			
GAD-7			
Pre-treatment mean (*SD*)	14.6 (4.2)	13.0 (5.0)	0.35 (−0.16 to 0.85)
Post-treatment mean (*SD*)	12.7 (4.9)	9.7 (5.1)	0.60 (0.09 to 1.11)
Pre–post mean change (*SD*)	−1.9 (3.6)	−3.3 (4.5)	0.34 (−0.16 to 0.85)
Pre–post *d* [correlation] (95% CI)	0.45 [$r = .69$] (0.08 to 0.82)	0.66 [$r = .62$] (0.27 to 1.05)	—
WSAS			
Pre-treatment mean (*SD*)	24.6 (8.8)	25.0 (8.1)	−0.04 (−0.55 to 0.45)
Post-treatment mean (*SD*)	20.3 (8.9)	19.2 (10.9)	0.11 (−0.39 to 0.61)
Pre-post mean change (*SD*)	−4.4 (6.2)	−5.9 (10.6)	0.17 (−0.33 to 0.67)
Pre–post *d* [correlation] (95% CI)	0.49 [$r = .75$] (0.12 to 0.86)	0.72 [$r = .42$] (0.32 to 1.11)	—

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 square brackets). BAG, behavioural activation in groups (existing intervention cohort); BAG+, behavioural activation in groups (augmented intervention cohort); PHQ-9, Patient Health Questionnaire; GAD-7, Generalized Anxiety Disorder scale; WSAS, Work and Social Adjustment Scale; CI, confidence interval.

(GAD-7) and impaired functioning (WSAS). Exploratory analyses using longitudinal mixed-models found significant loglinear time trends for reductions in depression scores over the course of group treatment (time main effect $\beta = -1.77$, SE = 0.50, $p = .001$), but no overall differences in PHQ-9 scores between BAG conditions (BAG condition main effect $\beta = 0.33$, SE = 1.17, $p = .778$). The time × condition interaction found a marginally significant difference in trajectories of PHQ-9 scores during group treatment, with BAG+ producing greater improvements after session 2 (time × condition interaction $\beta = -1.45$, SE = 0.69, $p = .042$). Figure 2 presents the fixed effect PHQ-9 trajectory estimates and 95% confidence intervals for BAG compared with BAG+.

Table 2 summarizes case-by-case outcomes for depression showing that BAG+ produced a significantly lower number of 'non-response' PHQ-9 categories. Patients who received BAG+ were three times less likely to have a non-response outcome at the end of group treatment. These reduced non-response outcomes were explained by significantly more patients in BAG+ experiencing reliable improvements in depression symptoms. Recovery and reliable recovery rates were not different in BAG when compared with BAG+. No single patient experienced a reliable deterioration in their depression. *Post-hoc* recovery rate analyses based on combined PHQ-9 and GAD-7 scores did not find any significant differences between BAG and BAG+.

To summarize, the two treatment augmentations did not improve attendance, but did appear to partially improve depression outcomes during BAG+. There was evidence of significantly more patients experiencing improvements and significantly fewer patients experiencing a non-response

https://doi.org/10.1017/S1754470X24000412 Published online by Cambridge University Press

Table 2. Individual depression (PHQ-9) recovery rates for BAG and BAG+

Post-treatment PHQ-9 recovery status	BAG ($n = 31$)	BAG+ ($n = 31$)	Chi-squared (p-value)	Odds ratio BAG:BAG+ (95% CI)
Recovery (clinical caseness)	10% (3)	29% (9)	3.72 ($p = .054$)	3.82 (0.92–15.81)
Reliable improvement	23% (7)	48% (15)	4.51 ($p = .034$)	3.21 (1.07–9.63)
Reliable recovery	10% (3)	29% (9)	3.72 ($p = .054$)	3.82 (0.92–15.81)
Nonresponse	77% (24)	52% (16)	4.51 ($p = .034$)	0.31 (0.10–9.33)
Reliable deterioration	0% (0)	0% (0)	—	—

BAG, behavioural activation in groups (existing intervention cohort); BAG+, behavioural activation in groups (augmented intervention cohort); PHQ-9, Patient Health Questionnaire; CI, confidence interval.

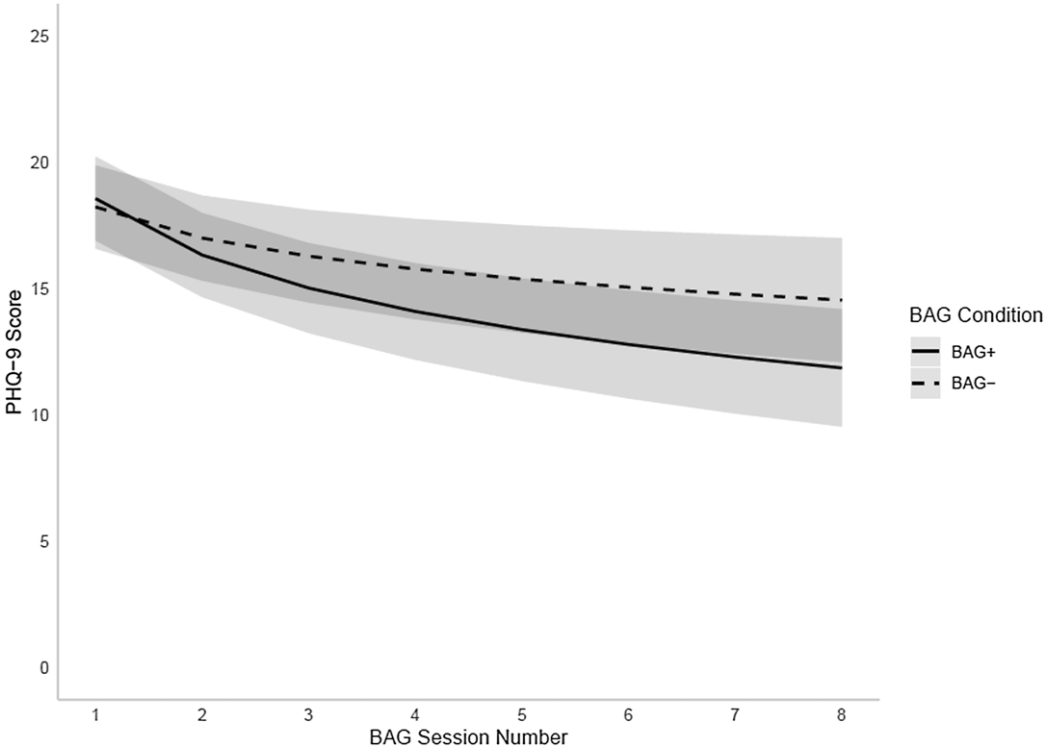

Figure 2. Longitudinal mixed-model fixed effect estimates for trajectories of depression (PHQ-9) scores over eight sessions of treatment for BAG and BAG+ (shaded bars represent 95% confidence intervals).

outcome relative to BAG, with some evidence of differing depression symptom trajectories during group delivery.

Discussion

This study used an implementation science approach to test whether treatment augmentations to an evidence-based depression treatment delivered in routine practice could be used to improve the

https://doi.org/10.1017/S1754470X24000412 Published online by Cambridge University Press

acceptability and effectiveness of the intervention. Both augmentations were able to be integrated into the existing BAG treatment content without changing the duration of the treatment. Augmenting treatment should not necessarily mean lengthening treatment. Fidelity to the BAG and BAG+ treatment protocols were checked and was deemed sufficient and PSM procedures facilitated a balanced comparison of BAG and BAG+ outcomes in the absence of a direct or randomized control. 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 for BAG+ and so did not appear to differentially improve the acceptability of the intervention. The overall attendance rates were poorer than those reported in the general literature (Swift and Greenberg, 2012), and specifically in NHS TT services (Kellett *et al.*, 2021) and may reflect the evidence that group interventions often have poor acceptability for patients (Cuijpers *et al.*, 2008). The effectiveness of the combined treatment augmentations on improving clinical outcomes appeared mixed. BAG+ patients were approximately three times more likely to have improved depression outcomes and appeared to have greater reductions in depression score trajectories from the third group session. However, overall post-treatment reductions in depression symptoms were not significantly different when BAG was compared with BAG+. This was mirrored in no significant differences also being observed in mean reductions in anxiety and impaired functioning scores.

The observed moderate to large within-group reductions in depression, anxiety and impaired functioning across both BAG and BAG+ provide further evidence that BA is clinically effective when delivered in groups in routine practice (Simmonds-Buckley *et al.*, 2019). There were no cases of depression symptom deterioration, suggesting that the treatment augmentations were safe. Improved outcomes were present in the individual depression outcome analyses, with non-response outcomes reduced from approximately 77% (BAG) to 52% (BAG+). Those fewer patients experiencing symptom non-response (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. The small beneficial effect reflected in the individual depression outcomes, but not the overall scores, suggests that the treatment augmentations were only beneficial to a small subsample of patients. It should be noted that partially beneficial effects of BAG+ were seen for depression outcomes only. Recovery rates combining depression and anxiety scores were not significantly different between BAG and BAG+, suggesting treatment additions were ineffective at facilitating greater improvements to wider mental health.

The results provide some tentative evidence for the utility of implementing low-cost treatment augmentations in the effort to improve the effectiveness of BA interventions, although the differences found were small and specific to depression. The mixed findings reflect the inconsistency of the evidence base for low-cost treatment augmentations. For example, Kellett *et al.* (2004) used practice-based evidence to better match patients to group CBT and improved outcomes and implementation intentions have been found to increase attendance at low-intensity group psychoeducational sessions (Avishai *et al.*, 2018). Whilst text-messaging shows moderate benefits in increasing attendance ($d = 0.5$), there are limited effects on clinical outcomes (Aguilera *et al.*, 2017). Similarly, Delgadillo *et al.* (2015) found that 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

https://doi.org/10.1017/S1754470X24000412 Published online by Cambridge University Press

standard BA patients. It is unknown whether patients engaged with the pre-course materials, although therapists did re-visit the information during the first treatment session unlike previous studies which only used pre-treatment leaflets (Delgadillo *et al.*, 2015; Swift and Callahan, 2011). 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, retain and use the psychoeducational information.

Findings build on the clinical use of implementation intentions and demonstrate the potential of integrating 'if–then' plans into existing treatment protocols (Lucock *et al.*, 2018; Toli *et al.*, 2016). The behavioural foundations of BA helped with 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). BA highlights the importance of context in both the maintenance of depression and the breaking of depressive cycles (Martell *et al.*, 2001). Similarly, implementation intentions promote the use of contextual cues to initiate pre-planned actions (Sheeran and Webb, 2016). The mechanisms of 'if–then' planning therefore was relatively easy to integrate into the principles and practice of BA and Fritzsche *et al.* (2016) have also previously illustrated the utility of implementation intentions in managing low mood. The separation observed in depression outcome trajectories during BAG+ from BAG from group session 3 may have been due to the action of the implementation intentions during BAG+ (i.e. potentially via more effective homework completion). However, as homework completion was not monitored, it is difficult to draw particularly firm conclusions here.

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 randomized to treatment, and so the differences found cannot be attributed with true confidence to the treatment augmentations. Future studies should therefore consider randomizing participants into BAG versus BAG+ but with adequate power to detect small effects. Because two augmentations were delivered during BAG+ it is impossible to disaggregate their separate effects. Future studies should therefore consider testing single augmentations iteratively. Potential confounds 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 anti-depressant medication in the BAG historical control cohort and lack of information on any concurrent treatments in both cohorts could have accounted for the differences found. In BAG+ patients signed a consent form to participate, and this may have primed them to respond positively. 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.

A direct association of the treatment augmentations on the intended outcomes cannot be assumed. For example, the dose–response psychoeducational augmentation may have strengthened trust in the efficacy of BAG+ and so created a placebo effect on outcomes. The two quality improvements may have also had a synergistic rather than intended disaggregated effect. Drawing any firm conclusion of a direct connection between implementation intentions and clinical outcome is particularly undermined due to the lack of any check of homework compliance. Monitoring homework compliance 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 et al., 2006). The durability of treatment augmentations was not assessed as no follow-up was conducted and long-term follow-up studies of the durability of treatment augmentations would be welcome.

https://doi.org/10.1017/S1754470X24000412 Published online by Cambridge University Press

Treatment completion rates of the full eight sessions were sub-optimal, so session-by-session scores were utilized to ensure a simplified pre–post score was available for the whole sample, replicating the approach used in routine services. However, the LOCF method used does have acknowledged limitations, as it is overly conservative and can introduce bias (Lachin, 2016). Use of a more sophisticated imputation method (e.g. multiple imputation) would have increased robustness of the findings. However, it should be noted that exploratory longitudinal mixed-model analyses of session-by-session outcomes that was used is well suited to handling missing data, and it found marginal evidence for greater depression reductions during BAG+ compared with BAG. The results are based solely on self-report data with associated risks of social desirability bias (Tourangeau and Yan, 2007). Using a combination of self and independent-assessor rated outcomes would strengthen the internal validity of future BAG studies.

The adherence check relied on BAG+ therapists self-rating 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 adherence to 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+ therapists 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. Finally, although checks were performed to assess the suitably of single-level analyses for the clustered data, it has been argued ICCs as low as 0.01 can still violate dependency assumptions (Baldwin *et al.*, 2011).

Conclusions

These results demonstrate that treatment augmentations to extant evidence-based therapies are feasible and offer a simple and direct means by which services can potentially improve outcomes. This is interventions being 'tweaked and tested' rather than wholescale replacement or being drawn into creating messy theoretical and clinical eclecticism. The use of practice-based data ensures that the results of this study have high external validity. The variety of processes that contribute to positive change during psychotherapy for depression 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 well controlled studies.

Key practice points

(1) Because much of the change work of BA is completed between sessions when activation is put in place, CBT therapists should pay attention to any factors that reduce the likelihood of homework completion (and reward homework completion). Implementation intentions appear to be a brief and useful summary of homework plans.
(2) Attendance in itself is a behaviour, and is the means by which BA is delivered and so therapists need to emphasize the importance of attendance.
(3) The reviewing and mutual design of homework exercises at each session is an effective way of socializing patients to BA 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 terms of understanding the function of their behaviours and how to adopt an 'outside-in' approach to change.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1754470X24000412

Data availability statement. The data are available on reasonable request from the corresponding author.

Acknowledgements. With thanks to the Improving Access to Psychological Therapies (IAPT) service (UK) and all the behavioral activation group facilitators who delivered the augmented interventions.

Author contributions. **Stephen Kellett:** Conceptualization (equal), Methodology (equal), Supervision (lead), Writing - original draft (equal), Writing - review & editing (equal); **Mel Simmonds-Buckley:** Conceptualization (equal), Data curation (lead), Formal analysis (lead), Funding acquisition (lead), Project administration (lead), Writing - original draft (equal), Writing - review & editing (equal); **Jen Hague:** Project administration (equal); **Glenn Waller:** Conceptualization (equal), Methodology (equal), Supervision-Supporting, Writing - original draft (equal), Writing - review & editing (equal).

Financial support. This work was supported by the Howard Morton Trust doctoral studentship awarded to the first author.

Competing interests. The authors declare none.

Ethical standards. The study received ethical and research governance approval from the Leeds East NHS Research Ethics Committee (IRAS project ID: 202197, Research Ethics Committee reference: 16/YH/0324), and was registered with a clinical trial database (ClinicalTrials.gov ID; NCT02970279) and authors have abided by the Ethical Principles of Psychologists and Code of Conduct as set out by the BABCP and BPS. Participants gave informed consent to participate in the study and for publication.

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*. 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 edn.

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

Bae, H., Shin, H., Ji, H. G., Kwon, J. S., Kim, H., & Hur, J. W. (2023). App-based interventions for moderate to severe depression: a systematic review and meta-analysis. *JAMA Network Open*, *6*, e2344120. doi: 10.1001/jamanetworkopen.2023.44120.

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. LʼAbate (Eds.), *Handbook of Homework Assignments in Psychotherapy* (pp. 51–63). Springer. https://doi.org/10.1007/978-0-387-29681-4_4

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*, 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

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

Cuijpers, P., Van Straten, A. & Warmerdam, L. (2008). Are individual and group treatments equally effective in the treatment of depression in adults? A meta-analysis. *European Journal of Psychiatry*, *22*, 38–51. https://doi.org/10.4321/S0213-61632008000100005

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. *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*, *79*, 101–108. https://doi.org/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*, 647–672. https://doi.org/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*, 43–57. https://doi.org/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*, 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*, 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

NHS Talking Therapies (2024). *NHS Talking Therapies for anxiety and depression manual (formerly known as Improving Access to Psychological Therapies)*. The National Collaborating Centre for Mental Health. https://www.england.nhs.uk/wp-content/uploads/2018/06/NHS-talking-therapies-manual-v7-1.pdf

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

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. 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*, 189–202. https://doi.org/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

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

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 Clinician's 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*, 429–438. https://doi.org/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. *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.

Oldham, M., Kellett, S., Miles, E., & Sheeran P. (2012). Interventions to increase attendance at psychotherapy: a meta-analysis of randomized controlled trials. *Journal of Consulting and Clinical Psychology*, *80*, 928–939. doi: 10.1037/a0029630

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*, 427–445. https://doi.org/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. *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*, 79–96. https://doi.org/10.1080/10503307.2019.1566676.

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*, CD013305. https://doi.org/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*, 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

Cite this article: Simmonds-Buckley M, Kellett S, Hague J, and Waller G (2025). Improving outcomes via treatment augmentations to behavioural activation for depression in routine practice: a cohort comparison study. *The Cognitive Behaviour Therapist*. https://doi.org/10.1017/S1754470X24000412