Behavioural and Cognitive Psychotherapy, 2017, 45, 401–418
First published online 13 March 2017 doi:10.1017/S1352465816000540

Effectiveness of Group Behavioural Activation for Depression: A Pilot Study

Stephen Kellett

Centre for Psychological Services Research, University of Sheffield, UK and Sheffield Health and Social Care NHS Foundation Trust, UK

Mel Simmonds-Buckley

University of Sheffield, UK

Paul Bliss

University of Sheffield, UK

Glenn Waller

University of Sheffield, UK

Background: The evidence base for behavioural activation (BA) is mainly grounded in the individual delivery method, with much less known about the impact of group delivery. **Aims:** To conduct a pilot study of behavioural activation in groups (BAG) for depression delivered in a routine service setting, in order to explore acceptability, effectiveness and predictors of outcome. **Methods:** The manualized group treatment format was delivered in a Primary Care mental health setting, at step three of an Improving Access to Psychological Therapies (IAPT) service. BAG was facilitated by cognitive behavioural psychotherapists, and outcome measures (depression, anxiety and functional impairment) were taken at each session. Seventy-three participants were referred and treated within nine groups. **Results:** BAG was an acceptable treatment generating a low drop-out rate (7%). Significant pre–post differences were found across all measures. There was a moderate to large depression effect size ($d_+ = 0.74$), and 20% met the criteria for a reliable recovery in depression. Greater severity of initial depression and attendance of at least four BAG sessions predicted better outcomes. **Conclusions:** BAG appears to be an effective depression treatment option that shows some clinical promise. Further larger and more controlled studies are nevertheless required.

Keywords: depression, behavioural activation, group therapy, pilot, clinical effectiveness, IAPT.

Correspondence to Stephen Kellett, Department of Psychology, University of Sheffield, Sheffield S10 2TN, UK. E-mail: s.kellett@sheffield.ac.uk

Introduction

What people do (i.e. their behaviour) and its function (i.e. the short and long-term consequences of actions) plays a key maintaining role in depression (Martell et al., 2001; Curran et al., 2012). The landmark component analysis of the treatment of depression by Jacobson et al. (1996) found that behavioural activation (BA) had equivalent outcomes to cognitive behavioural therapy (CBT) at both treatment termination and 6-month follow-up. This study re-ignited clinical interest in behaviour therapy as a stand-alone treatment for depression (Martell et al., 2001). Subsequent studies and meta-analyses have consistently demonstrated BA to be a clinically effective stand-alone treatment (Cuijpers et al., 2007; Dimidjian and Davis, 2009; Mazzucchelli et al., 2009; Ekers et al., 2014). Consequently, BA has been delivered in both low and high intensity versions of the model within Improving Access to Psychological Therapies (IAPT) services.

However, when updating the clinical guidelines for depression, the National Institute for Health and Clinical Excellence (NICE) stated that the BA evidence base was not yet compelling and called for more studies to be conducted (NICE, 2009). The Richards et al. (2016) recent large ($n = 221$) non-inferiority trial provided robust evidence of individual BA being no less effective than CBT. As the rates of infrequent adverse incidents were also equally distributed between the arms, BA was promoted as a potential 'frontline' depression intervention. In public services, increasing and amplified demand for psychological treatments and the need for organizationally efficient systems within restricted funding arrangements means that group delivery of treatments is often an attractive option (Kellett et al., 2007). There is still debate, however, as to whether group delivery is as effective as individual delivery for depression (Cuijpers et al., 2008), as any small initial increased effects of individual over group CBT appear to diminish at follow-up (Huntley et al., 2012).

BA is advocated as adaptable to a group delivery method because its parsimonious and teachable behavioural principles are relatively easy for facilitators to disseminate to clients (Jacobson et al., 1996; Sturmey, 2009; Richards et al., 2016). However, in comparison with individual BA, much less is known about the impact of BA delivered in a group format. The extant literature regarding group delivery of BA is based largely in clinical specialities or co-morbid populations, such as drug users undergoing in-patient substance abuse treatment (Daughters et al., 2008), patients with severe obesity (Alfonsson et al., 2015), young adolescents (Chu et al., 2009), opioid replacement therapy (Ross et al., 2012), and combat veterans in an in-patient post-traumatic stress disorder (PTSD) unit (Wright, 2002). The feasibility and effectiveness of group delivery of BA in more mainstream clinical services has been less well researched. Porter et al. (2004) piloted a ten-session group BA intervention ($n = 37$ patients) in Primary Care, resulting in significantly reduced depression. The uncontrolled evaluation by Houghton et al. (2008) of five treatment groups ($n = 42$ patients) in Secondary Care found that ten sessions of BA in groups (BAG) was an effective treatment, with a low drop-out rate.

When delivering treatments in routine service settings, an important consideration is the potential role of co-morbidity. In Primary Care services, rates of co-morbidity between anxiety and depression are typically high (Löwe et al., 2008; Zbozinek et al., 2012). Intervention effectiveness in IAPT is defined by 'change in a person, rather than just a syndrome' (IAPT, 2014, p. 4) and therefore reliable recovery is defined by the co-consideration of both anxiety and depression outcomes. As behavioural activation was

https://doi.org/10.1017/S1352465816000540 Published online by Cambridge University Press

originally designed as an intervention for low mood (Martell et al., 2001), its impact on co-morbid symptoms of anxiety was not considered in the initial group delivery studies (Porter et al., 2004; Houghton et al., 2008). Therefore, the impact that BAG has on anxiety across different clinical populations also needs to be examined in order to support its position as an evidence-based treatment option (NICE, 2009).

The purpose of the current study was to contribute to the BAG evidence base by conducting a pilot study of the implementation of BAG within a routine Primary Care mental health service containing two samples (depressed students and depressed routine referrals). The aims of the study were to: (1) index clinical characteristics of patients accessing BAG and associated drop-out and treatment completion rates, (2) define the clinical effectiveness of BAG both at a group level in terms of effect sizes and at an individual level in terms of applying robust outcome recovery metrics (IAPT, 2014), and (3) identify potential predictors of outcomes. Specific study hypotheses were as follows: (1) that significant reductions in depression and anxiety would occur during BAG, (2) that reductions in depression and anxiety would occur regardless of initial depression severity, and (3) outcomes would be equivalent regardless of clinical population (across sub-samples). Due to the relative scarcity of evidence for group delivery of BA in comparison with individual BA (Ekers et al., 2014), the design of this pilot study is consistent with research carried out within the first stage of the Salkovskis (1995) hourglass model of intervention evaluation.

Method

Participants

A total of 73 participants were initially referred to BAG. There were two sub-samples: (1) students seen in a University counselling centre ($n = 46$; four groups delivered in the student counselling centre) and (2) mainstream IAPT service users ($n = 27$; five groups delivered in GP [general practitioner/family physician] practices). All participants had a primary presenting problem of depression initially identified by a GP and then verified by a clinical assessment at initial screening by psychological wellbeing practitioners (PWPs) in the IAPT service or by staff in the student counselling centre. Self-referral to BAG was not available. The sample was representative of typical referrals within routine practice and therefore not all participants referred to BAG met 'caseness' for depression (scoring above the clinical cut-off pre-treatment). The nine groups ranged in size from 5 to 11, with an average group size of 8.11 (SD = 1.90). The authors assert that all procedures contributing to this work comply with the ethical standards of NHS research (IRAS reference 52311) on human experimentation and with the Helsinki Declaration of 1975, and its most recent revision.

Treatment

BAG was delivered via a manualized treatment protocol by a total of eight British Association for Behavioural and Cognitive Psychotherapies (BABCP) accredited CB psychotherapists. BAG was conducted as part of the step three IAPT 'high intensity' service provision. The treatment manual was adapted for Primary Care and was based on the Houghton et al. (2008) protocol, which was in turn based on Martell et al. (2001). All BAG facilitators attended a facilitator training induction meeting concerning group delivery of BA and were

https://doi.org/10.1017/S1352465816000540 Published online by Cambridge University Press

Table 1. BAG treatment content (patient wording)

Session	Content
1	Learn your patterns and start to change them
2	Getting out of TRAPs and back on TRAC
3	Taking action: a problem-solving approach
4	Values: the guide to who we are
5	Developing responses to thinking, worry and rumination
6	Making changes one step at a time
7	Freeing yourself from mood dependence*
8	Building the relationships you want/tying it all together*

*The final two sessions were amalgamated into one session for the student sample.
TRAP: Trigger, Response, Avoidance Pattern; TRAC: Trigger, Response, Alternative Coping.

already familiar with delivering individual BA within their role as CB psychotherapists. Two therapists facilitated each group. Novice facilitators were paired with an experienced group facilitator, who delivered the session content. The role of the novice facilitator in groups was to review the previous session and between-session work. The same experienced facilitators led the treatment delivery across both clinical sub-samples, but were paired with different novice facilitators in order to train up a pool of experienced facilitators. Treatment fidelity was monitored via (a) co-facilitator debriefing before and after each session and (b) facilitator supervision following each group course. Eight weekly two-hour sessions of BAG were delivered for the non-student sample and seven weekly two-hour sessions for the student sample (this difference was in order to fit within term time). Participants were provided with a welcome pack containing the outline of treatment sessions and their rationale and also a treatment workbook at the first group session. Each group session topic focus was aligned to relevant BA treatment principles (see Table 1). Treatment content was grounded within the following core theoretical principles: (a) mood–activity linkage, (b) a behavioural 'outside in' approach to change, (c) the importance of values-driven and values-consistent behaviour, (d) increasing access to a wide variety of naturally occurring positive reinforcement, (e) reducing avoidance via TRAP (Trigger, Response, Avoidance Pattern) and TRAC (Trigger, Response, Alternative Coping) techniques and (f) rumination-cued action.

Measures

Participants completed the IAPT minimum dataset at the beginning of every BAG session. The measures were as follows:

Patient Health Questionnaire. The PHQ-9 (Kroenke et al., 2001; scored between 0 and 27) was the primary outcome measure. The PHQ-9 was designed to detect depression within Primary Care settings, and items are derived from Diagnostic and Statistical Manual of Mental Disorders (4th edition, DSM-IV; American Psychiatric Association, 2000) classifications pertaining to lowered mood and physiological changes. Sensitivity and specificity for the PHQ-9 have been identified at 92 and 80%, respectively, at the >10 cut-off point (Gilbody et al., 2007). Depressive symptom severity was classified using the following PHQ-9 severity

https://doi.org/10.1017/S1352465816000540 Published online by Cambridge University Press

index scores defined by Kroenke et al. (2001): 'minimal' (0–4), 'mild' (5–9), 'moderate' (10–14), 'moderately severe' (15–19) and 'severe' (20–27).

Generalised Anxiety Disorder Assessment. The GAD-7 (Spitzer et al., 2006; scored between 0 and 21) is a severity measure of anxiety widely used in Primary Care. Applying a threshold score of 8 affords 92% sensitivity and 76% specificity (Swinson, 2006). Anxiety symptom severity was classified using the following GAD-7 severity index scores defined by Spitzer et al. (2006): 'minimal' (0–4), 'mild' (5–9), 'moderate' (11–14) and 'severe' (15–21).

Work and Social Adjustment Scale. The WSAS (Mundt et al., 2002; scored between 0 and 40) offers a valid, reliable measure of disability in terms of functional impairment as a result of mental health problems, and has good internal consistency and test–retest reliability. WSAS scores are sensitive to disorder severity and to treatment effects (Mundt et al., 2002).

Data analyses

Aim 1: index drop-out/completion rates and clinical characteristics. The sample was contextualized by reporting caseness and severity at intake. Independent samples *t*-tests and Pearson's chi square were used to compare demographic, attendance and psychometric differences between the sub-samples at pretreatment. Acceptability of BAG was indexed via session attendance and drop-out rates.

Aim 2: define the group and individual level of clinical effectiveness of BAG. Outcome analysis was conducted using an intention-to-treat (ITT) approach, including all patients who entered treatment (i.e. had intake scores). Due to the use of session-by-session outcome monitoring, last-observation-carried-forward (LOCF) imputation was used to compute post-treatment scores in cases of missing data. This is consistent with IAPT criteria, whereby all patients who attend at least one treatment session (two contacts overall including initial screening) are deemed to be 'cases'.

Firstly, to examine effectiveness at a group level, the pre–post change scores for each measure and associated paired-samples *t*-tests and effect sizes were calculated for the entire BAG sample and for each sub-sample (i.e. students *versus* standard). Paired-samples effect sizes, unweighted by sample size were calculated using the standard deviation (SD) of the change scores as the denominator, taking into account the pre–post score correlation (Morris and DeShon, 2002). Confidence intervals for the effect sizes were calculated using Hedges and Olkin's (1985) formula. Effect sizes were classified using Cohen's *d* (Cohen, 1992), defining $d_+ = 0.20$ as a 'small' effect, $d_+ = 0.50$ as a 'medium' effect, and $d_+ = 0.80$ as a 'large' effect. Graphing of sessional measures was used to illustrate the shape of change. Comparison of change scores between sub-samples and across BAG groups was achieved using analysis of covariance (ANCOVA), controlling for intake scores.

Secondly, individual level effectiveness was examined using recovery rate analysis on the whole sample and the two sub-samples. Categorical outcomes were calculated for each participant on the PHQ-9 and the GAD-7 using the original and updated national IAPT metrics (IAPT, 2014). The following category definitions were used:

(a) *Moving to recovery* counted participants above the clinical cut-off before and below following BAG. A participants was a 'case' when they scored above the clinical threshold on depression and/or anxiety before BAG (i.e. PHQ-9 score ≥10 at

https://doi.org/10.1017/S1352465816000540 Published online by Cambridge University Press

assessment and/or GAD-7 ≥8 at assessment). Moving to recovery occurred when the final outcome score was below the clinical threshold on depression and anxiety (i.e. PHQ-9 score <10 at termination and GAD-7 <8 at termination).

(b) *Reliable improvement* required that any improvement in outcome scores pre- and post-BAG exceeded measurement error of the PHQ-9 and GAD-7 using the reliable change index (RCI; Jacobson and Truax, 1991). Reliable improvement was a reduction of ≥6 points on the PHQ-9 or ≥4 points on the GAD-7 (IAPT, 2014). Conversely, *reliable deterioration* was an increase in scores exceeding measurement error.

(c) *Reliable recovery* required reliable improvement in the PHQ-9 or GAD-7, and that the case had to additionally move below the clinical threshold on both the PHQ-9 and the GAD-7 at the end of BAG.

(d) A *stasis* outcome was recorded when there was no movement to recovery and neither reliable improvement or deterioration.

(e) A *harm* outcome was recorded when a patient had a pre–post reliable deterioration on either the PHQ-9 or GAD-7 and also moved from non-case to case following BAG. The outcome categories of moving to recovery (original criteria), reliable improvement, reliable deterioration and reliable recovery (updated criteria) are standard IAPT terms (IAPT, 2014). Stasis and harm outcome categories were created as additional outcome categories by the research team.

Aim 3: identify predictors of outcome. Paired-samples *t*-tests were used to compare pre- and post-BAG depression scores, and correlations were used to test the role of initial depression severity in predicting pre–post reductions. The BAG dose–effect relationship was investigated by calculating reliable recovery rates by number of sessions attended and comparing group outcomes across different levels of attendance.

Results

Aim 1: Patient demographics, acceptability of treatment and clinical characteristics

Within the total sample ($n = 73$), the gender distribution was 43 females (59%) with an average age of 31.12 years (SD = 14.58), and 30 males (41%) with an average age of 34.97 years (SD = 14.66). The overall age range was 19–77 years. The majority were White British in ethnic origin ($n = 57$; 78%), and 13 (18%) patients were in receipt of state welfare benefits during BAG. The mean number of BAG sessions attended was 5.19 (SD = 1.85). In terms of consistency of attendance, 34 participants (47%) attended all the group sessions or all the group sessions minus one. One participant (1%) dropped out prior to BAG commencing and four participants (6%) dropped out after the first session, resulting in a drop-out rate of 7%. Figure 1 summarizes demographics and attendance rates for each sub-sample. Student and mainstream participants did not differ in terms of initial depression ($t(69) = 1.51$, $p = .14$) or GAD-7 score ($t(69) = 0.22$, $p = .83$) or on subsequent number of sessions attended ($t(71) = 0.81$, $p = .42$). There was a significant difference in terms of age ($t(71) = 12.64$, $p < .001$), but not gender (χ^2 (1, $n = 71$) = 0.29, $p = .59$). Due to the single participant who dropped out prior to BAG commencing and also one participant not completing the initial measures before dropping out after session one, the following ITT analyses relate to data available from the remaining 71 participants. In terms of the severity of depression, the overall pre-BAG PHQ-9 score was 16.07 (SD = 5.36); in categorical terms, seven participants

https://doi.org/10.1017/S1352465816000540 Published online by Cambridge University Press

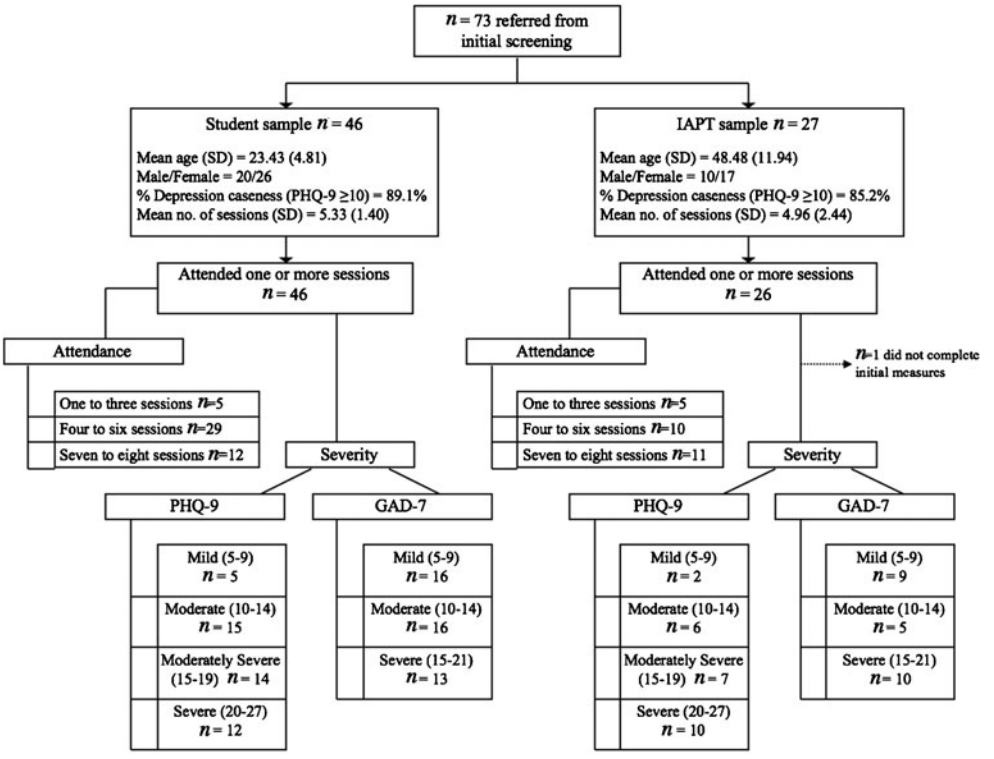

Figure 1. Demographics, severity and uptake of BAG by sub-sample

had 'mild' depressive symptoms, 21 'moderate' depressive symptoms, 21 'moderately severe' depressive symptoms and 22 'severe' depressive symptoms. Therefore, 43/71 (61%) met criteria for either 'moderately severe' or 'severe' depressive symptoms, with 64/71 (90%) reaching caseness prior to BAG. There was also evidence of high co-morbidity with anxiety (mean pre-treatment GAD-7 score = 12.42, SD = 4.99); 44 (62%) scored in the 'moderate' (11–15) or 'severe' (15–21) anxiety symptoms ranges, and 56/71 patients (79%) met overall caseness for anxiety prior to BAG.

Aim 2: Effectiveness of BAG – group and individual level analyses

Table 2 reports overall pre- (session 1) to post- (session 7/8) means, *t*-tests and associated effect sizes. On the primary outcome measure (PHQ-9), there was a statistically significant reduction in symptoms of depression. Significant reductions were also apparent on the secondary outcome measures of anxiety (GAD-7) and functional impairment (WSAS). Moderate to large effect sizes across all the outcome measures were apparent, with the PHQ-9 outcomes nearing a large effect size (Cohen, 1992). The session-by-session PHQ-9 scores for the sub-samples are presented in Figure 2. Outcome comparisons between the sub-samples (controlling for intake scores) show no significant differences in terms of depression ($F(1,68)$

https://doi.org/10.1017/S1352465816000540 Published online by Cambridge University Press

Table 2. BAG outcomes for the total sample and sub-samples

Measures	Pre-mean (SD)	Post-mean (SD)	Mean change (SD)	*t*-score	Correlation between means	*d*-value	95% confidence intervals
Total ($n = 71$)							
PHQ-9	16.07 (5.36)	12.59 (6.38)	3.48 (4.82)	6.08***	0.68	0.74	0.40–1.08
GAD-7	12.42 (4.99)	9.56 (5.39)	2.86 (4.34)	5.55***	0.65	0.66	0.32–1.00
WSAS	21.34 (6.72)	17.21 (8.56)	4.13 (6.57)	5.30***	0.66	0.66	0.32–1.00
Student patients ($n = 46$)							
PHQ-9	15.37 (4.86)	11.72 (5.68)	3.65 (5.00)	4.96***	0.56	0.74	0.32–1.16
GAD-7	12.33 (4.85)	9.00 (5.01)	3.33 (4.42)	5.12***	0.60	0.75	0.33–1.17
WSAS	21.96 (4.84)	17.46 (8.29)	4.50 (6.57)	4.64***	0.61	0.78	0.36–1.20
Non-student patients ($n = 25$)							
PHQ-9	17.36 (6.06)	14.20 (7.35)	3.16 (4.55)	3.47**	0.79	0.73	0.16–1.30
GAD-7	12.60 (5.33)	10.60 (5.99)	2.00 (4.15)	2.41*	0.74	0.49	–0.07–1.05
WSAS	20.20 (9.25)	16.76 (9.19)	3.44 (6.63)	2.59*	0.74	0.52	–0.04–1.08

*$p < .05$, **$p < .01$, ***$p < .001$. The *d*-value is Cohen's *d* effect size (0.2 = small, 0.5 = moderate, 0.8 = large).

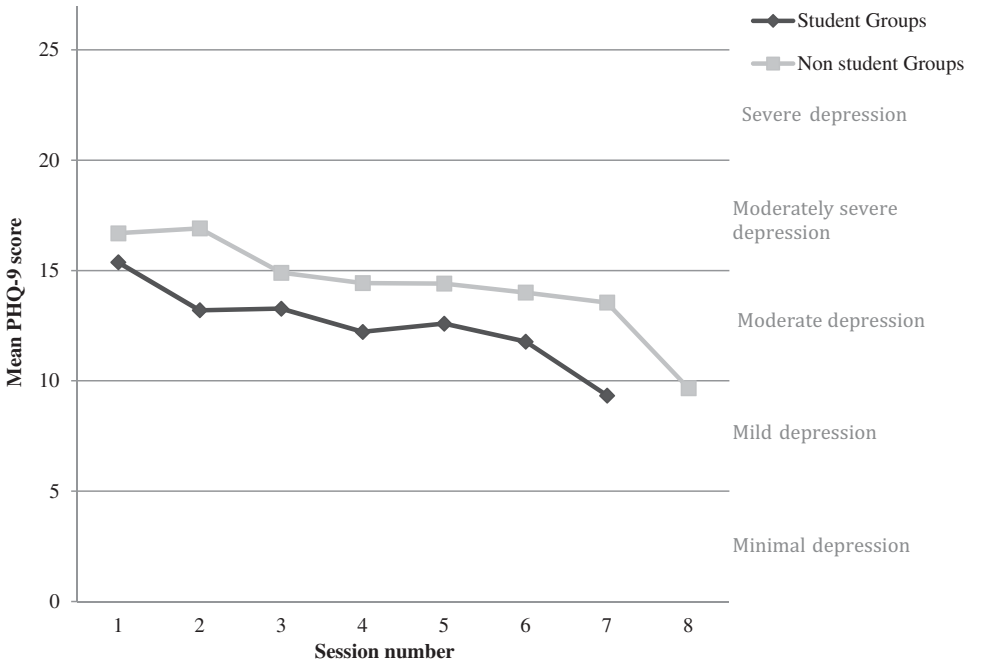

Figure 2. Session-by-session outcomes on the PHQ-9

= 0.58, $p = .45$), anxiety ($F(1,68) = 1.92$, $p = .17$), or functional impairment ($F(1,68) = 0.23$, $p = .64$). Comparisons between individual BAG groups (controlling for intake scores) also found no significant differences in terms of depression ($F(8,61) = 0.60$, $p = .77$), anxiety ($F(8,61) = 0.97$, $p = .47$) or functional impairment ($F(2,61) = 1.52$, $p = .17$).

Table 3 reports the depression and anxiety individual recovery rate analysis for the entire and sub-samples. On the primary outcome measure (PHQ-9), 22 participants made a reliable improvement, with 14 of those participants also meeting the additional criteria for reliable recovery (i.e. 20% showed clinical and reliable change). Thirty-seven participants had a depression stasis outcome following BAG. Change rates were comparable across BAG groups. Reliable improvement rates (including reliable recovery participants) were also evenly distributed across the clinical samples: student participants ($n = 15$, 33%) and standard IAPT participants ($n = 7$, 28%). Similar recovery rates were observed for anxiety, with 21% also showing clinical and reliable change on the GAD-7. When applying IAPT recovery metrics using participants' scores on both measures, 25% showed some form of reliable improvement following BAG, and a further 18% showed reliable improvement as well as reaching non-caseness for both depression and anxiety. Overall, 44% experienced a reliably significant benefit whilst attending BAG.

Aim 3: Predictors of outcome

Initial depression severity. Table 4 reports the rates of depressive symptom severity pre- and post-BAG. On completing BAG, there were almost half as many participants with 'moderate

https://doi.org/10.1017/S1352465816000540 Published online by Cambridge University Press

Table 3. Recovery rate analysis for depression and anxiety during BAG

	Depression (PHQ-9)		Anxiety (GAD-7)		IAPT outcomes: anxiety and depression	
	n	(%)	*n*	(%)	*n*	(%)
Total ($n = 71$)						
Harm	0	(0)	0	(0)	0	(0)
Reliable deterioration*	0	(0)	4	(5.63)	4	(5.63)
Stasis	37	(52.11)	27	(38.03)	33	(46.48)
Moving to recovery**	20	(28.17)	18	(25.35)	16	(22.54)
Reliable improvement*	8	(11.27)	10	(14.09)	18	(25.35)
Reliable recovery*	14	(19.72)	15	(21.13)	13	(18.31)
Student patients ($n = 46$)						
Harm	0	(0)	0	(0)	0	(0)
Reliable deterioration*	0	(0)	2	(4.35)	2	(4.35)
Stasis	23	(50.00)	17	(39.96)	21	(45.65)
Moving to recovery**	14	(30.44)	13	(28.26)	13	(28.26)
Reliable improvement*	5	(10.87)	6	(13.04)	11	(23.91)
Reliable recovery*	10	(21.74)	11	(23.91)	11	(23.91)
Non-student patients ($n = 25$)						
Harm	0	(0)	0	(0)	0	(0)
Reliable deterioration*	0	(0)	2	(8.00)	2	(8.00)
Stasis	14	(56.00)	10	(40.00)	12	(48.00)
Moving to recovery**	6	(24.00)	5	(20.00)	3	(12.00)
Reliable improvement*	3	(12.00)	4	(16.00)	7	(28.00)
Reliable recovery*	4	(16.00)	4	(16.00)	2	(8.00)

*Updated IAPT recovery criteria; **original IAPT recovery criteria. Original and updated IAPT recovery rate categories are not distinct and therefore not mutually exclusive; patients can be classified in both. Therefore, the *n* can total > 100%.

Table 4. Pre- and post-depressive symptom severity rates

	Pre		Post	
Depressive symptom severity	*n*	(%)	*n*	(%)
Minimal (0–4)	1	(1.38)	5	(6.94)
Mild (5–9)	7	(9.72)	21	(29.17)
Moderate (10–14)	21	(29.17)	21	(29.17)
Moderately severe (15–19)	21	(29.17)	12	(16.67)
Severe (20–27)	22	(30.56)	13	(18.06)

to severe' or 'severe' depressive symptoms on the PHQ-9 ($n = 25$, 35%). The number of participants meeting caseness following BAG reduced to $n = 46$ (64%) on the PHQ-9 and $n = 41$ (58%) on the GAD-7. Figure 3 illustrates mean depression outcomes pre- and post-BAG by initial depressive symptom severity, to show that reductions were observed across all depression severities. The pre–post reductions for moderate ($t(20) = 2.79$, $p = .011$),

https://doi.org/10.1017/S1352465816000540 Published online by Cambridge University Press

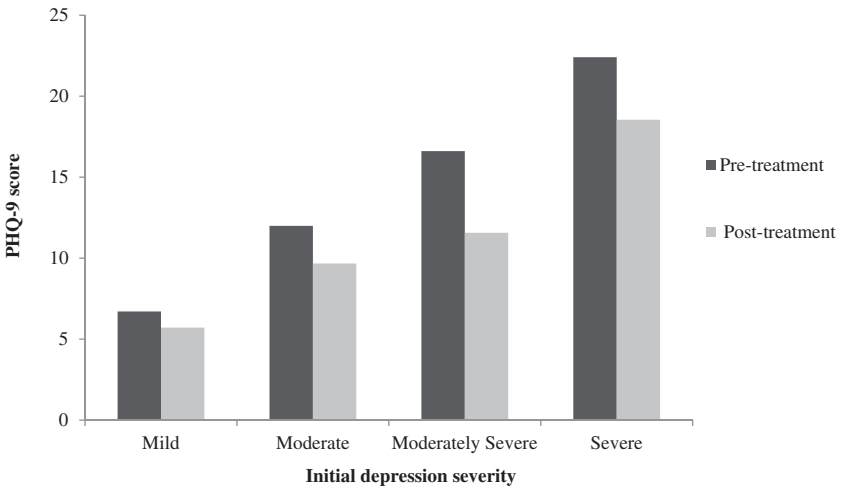

Figure 3. PHQ-9 outcomes pre- and post-BAG according to initial depressive symptom severity

moderately severe ($t(20) = 4.04$, $p = .001$) and severe depressive symptoms ($t(21) = 3.72$, $p = .001$) were significant, whilst the mild depressive symptoms reduction was not significant ($t(6) = 0.90$, $p = .403$). Further investigation of the relationship between initial depressive symptom severity and change scores found a non-significant, small positive correlation nearing significance (PHQ-9; $r(69) = .22$, $p = .07$).

Dose–effect. Figure 4 illustrates the percentage of participants who experienced reliable recovery when grouped according to the number of consecutive BAG sessions attended. The criteria for a reliable recovery were not met by any single participant who attended fewer than four sessions. Recovery criteria tended to be achieved by those who attended a minimum of four sessions. The likelihood of reliable recovery increased according to number of BAG sessions attended.

Figure 5 shows change scores according to number of sessions attended, showing that those patients attending less than four sessions achieved less change. Comparisons of initial depression severity for those who attended three sessions or fewer (mean = 16.22, SD = 1.80), 4–6 sessions (mean = 16.56, SD = 0.86) or 7–8 sessions (mean = 15.13, SD = 1.12) showed no significant differences on initial PHQ-9 scores ($F(2,68) = .53$, $p = .60$). The effect on mean depression (PHQ-9) change of the number of sessions attended (0–3, 4–6 or 7–8) was significant ($F(2,68) = 3.83$, $p = .027$). Those who attended fewer than four BAG sessions experienced significantly less change in depressive symptoms compared with those who attended 4–6 sessions ($p < .05$). Mean change according to attendance at 4–6 and 7–8 group sessions was not significantly different ($p = .86$). *t*-tests on the pre- and post-BAG scores according to number of sessions attended demonstrated that significant reductions in depression were found only for those who attended four or more sessions (4–6 sessions: $t(38) = 5.26$, $p < .001$ and 7–8 sessions: $t(22) = 3.90$, $p = .001$). Participants who attended fewer than three sessions did not experience a significant pre–post reduction in depression ($t(8) = 0.65$, $p = .54$).

https://doi.org/10.1017/S1352465816000540 Published online by Cambridge University Press

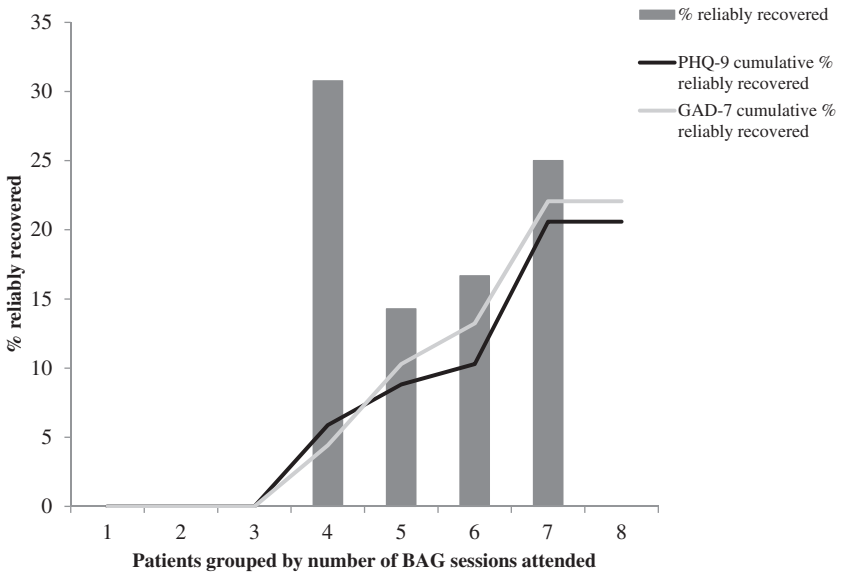

Figure 4. Rate of reliable recovery by number of BAG sessions attended

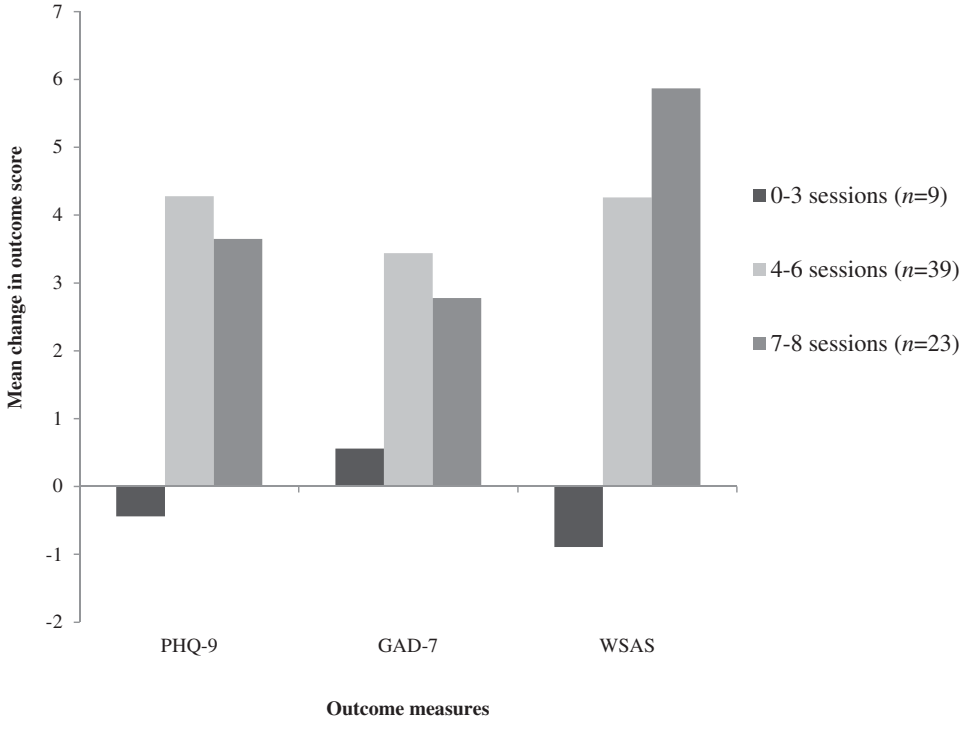

Figure 5. Mean outcome change according to number of BAG sessions attended

https://doi.org/10.1017/S1352465816000540 Published online by Cambridge University Press

Discussion

The evidence base for BA is mainly built around individual delivery, with very few studies investigating group delivery. This pilot study investigated the routine implementation of BAG in order to index acceptability, effectiveness and predictors of outcome. The results supported the hypothesis that significant reductions in depression and anxiety would occur during BAG regardless of clinical population or initial depression severity. BAG appears a feasible primary care treatment at step three of IAPT services, particularly due to the low patient attrition rate observed. Indeed, the more sessions attended, the greater the likelihood of benefit. This pilot study found that therapists already using BA in their individual work can rapidly assimilate BA principles and practice into a group delivery setting. BAG as a treatment option is a clinical innovation in terms of plurality of intervention offered in Primary Care and appears efficient in terms of therapist-to-patient ratios (Kellett et al., 2007).

The depression outcomes mirror the Porter et al. (2004) and Houghton et al. (2008) BAG evidence, albeit demonstrated in a larger and more diverse sample and with a shorter course of group therapy. A third of participants made reliable improvements, with one in five making reliable recovery during the group. BAG appears to be a safe intervention for depression, as no single participant was harmed or reliably deteriorated during treatment. The overall depression effect size found ($d_+ = 0.74$) is categorized as nearing a large effect size and is similar to the between-groups effect sizes reported in the Cuijpers et al. (2007), Mazzucchelli et al. (2009) and Ekers et al. (2014) meta-analyses of BA and also trials of individually delivered BA (Hopko et al., 2003; Dimidjian et al., 2006). It should be noted that as the pre–post score correlation was >0.5, the within-sample effect size reported in this study will appear larger than a between-groups effect size. Facilitator training in BAG was achieved through coupling experienced to novice facilitators, and so enabled the service to quickly acquire a pool of experienced facilitators. No differences were found between groups or sub-samples, indicating consistency of delivery of the manualized treatment approach (Delgadillo et al., 2016).

The presence of co-morbid anxiety symptoms did not appear to hinder BAG depression outcome. Indeed, BAG appears to be beneficial for co-morbid anxiety symptoms (even those reaching caseness), with significant reductions, recovery rates and effect sizes for anxiety that all but matched the depression outcomes. This supports BAG's suitability as an effective intervention in a real world primary mental health care setting, and its appropriateness as a step three intervention in IAPT services. Four participants experienced a reliable deterioration in anxiety during BAG. The reason for this is not clear, but evidence suggests that a relatively small but non-trivial minority of patients do deteriorate during psychotherapy, with estimates ranging from 3 to 10% (Strupp et al., 1977; Mohr, 1995). BAG might not be suitable for patients with higher levels of co-morbid anxiety, but an alternative possibility is that such patients might require further anxiety-specific evidence-based treatment following BAG.

The high uptake and low drop-out rates mirror previous preliminary findings (Porter et al., 2004; Houghton et al., 2008), implying that BAG is well tolerated as an intervention. The drop-out rate for BAG is lower than for individual interventions delivered in IAPT services (NEPHO, 2010; Richards and Borglin, 2011). However, different studies have used various definitions of drop-out and so there is a need for a universally accepted drop-out classification to enable researchers to compare rates with true confidence (Oldham et al., 2012). What it is about BA that is engaging for depressed patients does need researching, but an 'outside-in'

https://doi.org/10.1017/S1352465816000540 Published online by Cambridge University Press

approach (Martell et al., 2001; Curran et al., 2012) appears useful and non-shaming. Ensuring low drop-out rates is vital, as increased attendance at BAG was associated with improved outcome across all measures. Attendance at a minimum of four group sessions was required to achieve significant improvements and to have a chance of meeting recovery criteria. The importance of treatment completion for outcome is well evidenced in individual work (Cahill et al., 2003). BAG was effective across all the clinical threshold severities of depression and across both clinical sub-samples, although it is acknowledged that regression to the mean may have influenced the effect seen for more severely depressed patients.

The main study weakness was the lack of a control group and random allocation (Lilienfeld, 2007), prohibiting understanding whether it was BA itself or participation in a group of similarly depressed people that facilitated the recorded psychological outcomes (Yalom and Leszcz, 2005). Similarly, the lack of follow-up means the durability of BAG is not yet known. Any assumptions about increased organizational efficiency through implementing BAG need to be cautiously framed, as it is not known if stasis patients then return to services to seek further treatment. As with any outcome study in a routine service setting there was loss of data over time, which limits confidence in the results (Ahern and Le Brocque, 2005). The use of the same outcome measure for pre- and post-assessment and session-by-session monitoring also has its own drawbacks, in both the likelihood of testing effects and isolating change inapplicable to other outcome measures (Wampold, 2015). All outcome data were self-reported and therefore subject to recognized validity issues, such as social desirability bias with regards to over- or under-reporting sensitive information (Tourangeau and Yan, 2007). Future studies would benefit from using more mixed methods, such as combining self-report and diagnostic interviews. Further studies should aim to investigate and specify the mechanisms and moderators of change during BAG. There are currently no validated competency measures of BAG. Therefore, treatment fidelity was not accurately assessed and this represents a further future research need.

Although no patients experienced a reliable deterioration in depression or harm outcome, over half of all participants experienced a ‘stasis’ outcome. BAG is clearly not a panacea for all individuals with depression. The importance of stasis as a depression outcome needs further detailed empirical and theoretical attention. Depression has a core feature of hopelessness, and not benefitting from an intervention (despite it being evidence-based) might enhance ‘failure’ schema in depressed patients. Stasis rates should not be overlooked by services and researchers. As the reasons for stasis outcomes are not clear, detailed qualitative and quantitative investigation of stasis outcomes in depression are warranted. Understanding what creates treatment stasis provides many opportunities for redesign and re-testing consistent with the Salkovskis (1995) hourglass model. The considerable rate of stasis observed in this study and other IAPT outcome research (e.g. Chan and Adams, 2014; Green et al., 2014; Firth et al., 2015) demands attention as to why an evidence-based intervention makes little impression on a substantial proportion of patients.

In conclusion, this pilot study has shown that an innovation on the behavioural activation approach (i.e. a manualized brief group delivery format) shows clinical and organizational promise. Clearly, the evidence base for the BAG is dwarfed by the traditional individual approach and much research work on a larger scale is therefore indicated. Patients suitable for a behavioural approach might be offered the choice of group or individual BA and a patient-choice RCT would be useful to test this (Howard and Thornicroft, 2006). Helping people to understand their values and align their behaviour accordingly in order to improve their lives

appears to be a potent and simple approach when treating depression (Curran et al., 2012). Completing such 'outside-in' work in a supportive group environment in which vicarious learning from others can also take place appears a useful avenue for future clinical treatment.

Acknowledgements

Conflicts of interest: The authors have no conflicts of interest with respect to this publication.

Financial support: M.S.B. was funded by the Howard Morton Trust doctoral studentship award.

References

Addis, M. E. and Martell, C. M. (2004). *Overcoming Depression One Step at a Time*. Oakland, Harbinger Publications.

Ahern, K. and Le Brocque, R. (2005). Methodological issues in the effects of attrition: Simple solutions for social scientists. *Field Methods*, *17*, 53–69.

Alfonsson, S., Parling, T. and Ghaderi, A. (2015). Group behavioral activation for patients with severe obesity and binge eating disorder: A randomized controlled trial. *Behavior Modification*, *39*, 270–294. doi:10.1177/0145445514553093

American Psychiatric Association (2000). *Diagnostic and statisitcal manual of mental disorders* (4th edn, text revision). Washington, DC.

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

Chan, S. W. Y. and Adams, M. (2014). Service use, drop-out rate and clinical outcomes: A comparison between high and low intensity treatments in an IAPT service. *Behavioural and Cognitive Psychotherapy*, *42*, 747–759. doi:10.1017/s1352465813000544

Chu, B. C., Colognori, D., Weissman, A. S. and Bannon, K. (2009). An initial description and pilot of group behavioral activation therapy for anxious and depressed youth. *Cognitive and Behavioral Practice*, *16*, 408–419.

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry*, *23*, 318–327. doi:10.3109/09540261.2011.606803

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

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

Cuijpers, P., van Straten, A. and 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.

Curran, J., Ekers, D., McMillan, D. and Houghton, S. (2012). Behavioural activation. In W. Dryden (ed), *Cognitive Behaviour Therapies* (pp. 236–260). London, UK: Sage Publications.

Daughters, S. B., Braun, A. R., Sargeant, M. N., Reynolds, E. K., Hopko, D. R., Blanco, C. and Lejuez, C. W. (2008). Effectiveness of a brief behavioral treatment for inner-city illicit drug users with elevated depressive symptoms: The Life Enhancement Treatment for Substance Use (LETS Act!). *Journal of Clinical Psychiatry*, *69*, 122–129.

Delgadillo, J., Kellett, S., Ali, S., McMillan, D., Barkham, M., Saxon, D. et al. (2016). A multi-service practice research network study of large group psychoeducational

cognitive behavioural therapy. *Behaviour Research and Therapy, 87*, 155–161. Available at: http://dx.doi.org/10.1016/j.brat.2016.09.010

Dimidjian, S. and Davis, K. J. (2009). Newer variations of cognitive behavioral therapy: Behavioral activation and mindfulness-based cognitive therapy. *Current Psychiatry Reports*, *11*, 453–458.

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., Gallop, R., McGlinchey, J. B., Markley, D. K., Gollan, J. K., Atkins, D. C., Dunner, D. L. and Jacobson, N. S. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology*, *74*, 658–670.

Ekers, D., Webster, L., Van Straten, A., Cuijpers, P., Richards, D. and Gilbody, S. (2014). Behavioural activation for depression; An update of meta-analysis of effectiveness and sub group analysis. *Plos One*, *9*, 11. doi:10.1371/journal.pone.0100100

Firth, N., Barkham, M., Kellett, S. and Saxon, D. (2015). Therapist effects and moderators of effectiveness and efficiency in psychological wellbeing practitioners: A multilevel modelling analysis. *Behaviour Research and Therapy*, *69*, 54–62. doi:10.1016/j.brat.2015.04.001

Gilbody, S., Richards, D., Brealey, S. and Hewitt, C. (2007). Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): A diagnostic meta-analysis. *Journal of General Internal Medicine*, *22*, 1596–1602.

Green, H., Barkham, M., Kellett, S. and Saxon, D. (2014). Therapist effects and IAPT Psychological Wellbeing Practitioners (PWPs): A multilevel modelling and mixed methods analysis. *Behaviour Research and Therapy*, *63*, 43–54. doi:10.1016/j.brat.2014.08.009

Hedges, L. and Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. New York, NY: Academic Press.

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

Houghton, S., Curran, J. and Saxon, D. (2008). An uncontrolled evaluation of group behavioural activation for depression. *Behavioural and Cognitive Psychotherapy*, *36*, 235–239. doi:10.1017/s1352465808004207

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

Huntley, A. L., Araya, R. and Salisbury, C. (2012). Group psychological therapies for depression in the community: Systematic review and meta-analysis. *British Journal of Psychiatry*, *200*, 184–190. doi:10.1192/bjp.bp.111.092049

Improving Access to Psychological Therapies (IAPT) (2014). *Measuring improvement and recovery*. Available at: http://webarchive.nationalarchives.gov.uk/20160302154833/http://www.iapt.nhs.uk/silo/files/measuring-recovery-2014.pdf

Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., Gortner, E. and Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology*, *64*, 295–304. doi:10.1037//0022-006x.64.2.295

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

Kellett, S., Clarke, S. and Matthews, L. (2007). Delivering group psychoeducational CBT in Primary Care: Comparing outcomes with individual CBT and individual psychodynamic-interpersonal psychotherapy. *British Journal of Clinical Psychology*, *46*, 211–222. doi:10.1348/014466506x146188

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

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, *2*, 53–70.

Löwe, B., Spitzer, R. L., Williams, J. B., Mussell, M., Schellberg, D. and Kroenke, K. (2008). Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. *General Hospital Psychiatry, 30*, 191–199.

Martell, C. R., Addis, M. E. and Jacobson, N. S. (2001). *Depression in Context: Strategies for Guided Action*. New York, NY: W.W. Norton & Company Inc.

Mazzucchelli, T., Kane, R. and Rees, C. (2009). Behavioral activation treatments for depression in adults: A meta-analysis and review. *Clinical Psychology: Science and Practice*, *16*, 383–411.

Mohr, D. C. (1995). Negative outcome in psychotherapy: A critical review. *Clinical Psychology: Science and Practice*, *2*, 1–27.

Morris, S. B. and DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105–125.

Mundt, J. C., Marks, I. M., Shear, M. K. and Greist, J. H. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *British Journal of Psychiatry*, *180*, 461–464. doi:10.1192/bjp.180.5.461

National Institute for Health and Clinical Excellence (NICE) (2009). *Depression: The Treatment and Management of Depression in Adults* (update). Clinical Guideline 90. London. Available at: www.nice.org.uk/CG90.

North East Public Health Observatory (NEPHO). (2010). *Improving Access to Psychological Therapies: a review of the progress by the sites in the first roll-out year*. Stockton on Tees, UK: NEPHO.

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

Porter, J. F., Spates, C. R. and Smitham, S. (2004). Behavioral activation group therapy in public mental health settings: A pilot investigation. *Professional Psychology: Research and Practice*, *35*, 297–301. doi:10.1037/0735-7028.35.3.297

Richards, D. A. and Borglin, G. (2011). Implementation of psychological therapies for anxiety and depression in routine practice: Two year prospective cohort study. *Journal of Affective Disorders*, *133*, 51–60. doi:10.1016/j.jad.2011.03.024

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. R., Wright, K. A., Hollon, S. D., Reed, N., Rhodes, S., Fletcher, E. and Finning, K. (2016). Cost and outcome of behavioural activation versus cognitive behavioural therapy for depression (COBRA): A randomised, controlled, non-inferiority trial. *The Lancet*, *388*, 871–880. doi:10.1016/S0140-6736(16)31140-0

Ross, J., Kaye, S., Mills, K., Dore, G., Lejuez, C., Deady, M. and Ewer, P. (2012). Group behavioural activation therapy for the treatment of depression among opioid replacement therapy clients: A pilot study. *Drug and Alcohol Review*, *31*, 20–20. doi:10.1111/j.1465-3362.2012.00514.x

Salkovskis, P. M. (1995). Demonstrating specific effects in cognitive and behavioural therapy. In M. Aveline and D. A. Shapiro (eds), *Research Foundations for Psychotherapy Practice* (pp. 191–228). Chichester, UK: Wiley.

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

Strupp, H. H., Hadley, S. W. and Gomes-Schwartz, B. (1977). *Psychotherapy for Better or Worse: The Problem of Negative Effects*. New York, NY: Aronson.

Sturmey, P. (2009). Behavioral activation is an evidence-based treatment for depression. *Behavior Modification*, *33*, 818–829. doi:10.1177/0145445509350094

Swinson, R. P. (2006). The GAD-7 scale was accurate for diagnosing generalised anxiety disorder. *Evidence-Based Medicine*, *11*, 184–184. doi:10.1136/ebm.11.6.184

Tourangeau, R. and Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*, 859–883.

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

Wright, T. (2002). The effectiveness of behavioral activation group therapy: Treating comorbid depression on a specialized inpatient post-traumatic stress disorder unit for combat veterans. *Dissertation Abstracts International*, *64*, 436B (UMI no. 3077387).

Yalom, I. D. and Leszcz, M. (2005). *The Theory and Practice of Group Psychotherapy* (5th edn). New York, NY: Basic Books.

Zbozinek, T. D., Rose, R. D., Wolitzky-Taylor, K. B., Sherbourne, C., Sullivan, G., Stein, M. B., Roy-Byrne, P. P. and Craske, M. G. (2012). Diagnostic overlap of generalized anxiety disorder and major depressive disorder in a primary care sample. *Depression and Anxiety*, *29*, 1065–1071.