The University Of Sheffield.

Group Behavioural Activation for Depression:

Predicting and Reducing 'Stasis' Outcomes

By:

Melanie Simmonds-Buckley

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Science
Department of Psychology

Submission Date

(March 2019)

ABSTRACT

Depression is a widely prevalent, debilitating disorder that causes significant suffering for those affected. Behavioural activation (BA) is an evidence-based psychological treatment for depression. The evidence base for BA is largely grounded in individual delivery, with far less known about BA delivered in groups. Given rising demand for treatment, groups represent an attractive delivery strategy for services. Despite good rates of treatment outcomes for evidenced-based psychological treatments, considerable numbers of patients do not benefit and remain in a state of depression 'stasis' (i.e., their symptoms are relatively unchanged, despite receiving treatment). This thesis sought to investigate the effectiveness and efficacy of group BA, the reasons why stasis might occur and also how stasis can be reduced. First, a meta-analysis (Chapter 2) of group BA depression outcomes in trial (efficacy) and naturalistic (effectiveness) contexts is presented, to clarify the treatment effect of group-based BA interventions. Group BA is shown to be superior to controls (Hedges *g* effect size = 0.72) and equivalent to other active therapies (*g* = 0.14) at treatment completion. Chapter 3 provides an overview of depression stasis after evidence-based treatment. The chapter highlights the difficulties in identifying stasis outcomes, the extent of the problem (up to ~60% of patients treated) and the paucity of evidence about associated factors. As a result, a stasis metric is defined for use in the subsequent Chapters. An analysis of BA treatment response (Chapter 4) then investigates the effect of intervention intensity, format and duration on stasis outcomes. BA is seen to be effective at reducing depression (in 4-9 sessions) regardless of format, with larger effects seen for more intensive versions. 'Stasis patients' are distinguishable from 'improvers' after 2 sessions. Risk of a BA stasis outcome was predicted by attending fewer sessions, greater impaired functioning prior to treatment, and less severe depression. Building on these findings, Chapter 5 then tests

an augmented group BA treatment to determine whether drop-out and stasis outcomes can be reduced. Whilst treatment retention remains stable, significantly fewer patients experience a stasis outcome after the augmented treatment, due to increased rates of improvement. Lastly, a mediation study (Chapter 6) evaluates whether increasing behaviour in accordance with life values (*valued living*) is a BA change mechanism. Discrepancies in *valued-living* were not related to depression severity, nor did *valued-living* increase as a result of group BA. Exploratory analyses showed *valued living* does not mediate reductions in depression symptom clusters (somatic or affective) during group BA therapy. Finally, the overall theoretical, clinical, organisational and research implications are discussed in Chapter 7, with recommendations outlined for future investigation into depression stasis and treatment.

ACKNOWLEDGEMENTS

I have thoroughly enjoyed doing my PhD and cannot believe three years have flown by so quickly. I have learned so much from my experience and have so many people to thank for that. First and foremost, I would like to thank the Howard Morton trust for funding the project. Particular thanks have to go to my supervisors, Steve Kellett and Glenn Waller. A PhD experience is hugely shaped by the relationship you have with your supervisors and I am in no doubt that I have enjoyed it so much in large part to their support and guidance. It is fair to say I would not have even considered a PhD as an option without Steve continually forcing me to do things outside of my comfort zone. I would never have dreamed that taking a Clinical Psychology administrative job seven years ago would result in a doctorate. Steve has been instrumental in shaping my career – and more importantly, the football chat about my beloved Sheffield Wednesday (occasionally interspersed with discussions about work) has kept me sane! I am so grateful for the constant support and insight provided by Glenn. His calmness in the face of any problems always made everything seem immediately more manageable. I have undoubtedly become a better researcher thanks to Glenn's guidance and feedback (fingers crossed I have finally mastered the difference between 'which' and 'that'!).

Enormous thanks must also go to the numerous people who helped ensure the studies were even able to run. My thanks go to Sheffield IAPT for allowing me to conduct the research and providing access to the data, all the BAG facilitators (including Jennie Hague and Jo Priddy as BAG leads) for running the groups and the patients for agreeing to take part. In particular, Jennie's enthusiasm for BAG was pivotal in getting the studies off the ground and I cannot thank her enough for the effort she put in to promote my research. I must also mention Esther, my thesis mentor, who

provided valued support with the arduous task of writing (and kept our meetings going long after she was required to).

One of the highlights of my PhD experience has been the close friends I have made – Harriet, Connor, Anna, Asha and Kerry. I am so grateful to have gone through this experience together. Long may our after-work bottle (or two) of wine tradition continue! I, of course, also need to thank my support network outside of the PhD. My best friends Ness, Ellie, Rosie and Lucy have always been there to provide a most welcome distraction from the PhD. My thanks also go to my family for all their love and support. My dad (who will inevitably try to take the credit for this thesis!) has always been there to look after me with cups of tea and encouragement. Discussions with my older brother Matt, gave me some of the most thought-provoking and useful insights, outcomes of which have had a real, meaningful impact. My younger brother Nathan has provided invaluable advice and support in the tough moments and has shown himself to be the wisest one in the family (who would have thought!). My boyfriend, Jack, has been there through all the ups and the downs, always knowing just what I need without me even needing to ask (and often before I have even figured it out myself!).

Finally, I will dedicate this thesis to my mum, Davina. She inspired me before I even thought of doing a PhD in psychology. Following in her footsteps makes me immensely proud, and I am honoured to be able to put my thesis on the bookshelf next to hers.

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

GLOSSARY OF COMMONLY USED TERMS

'Stasis' – term used in this thesis to refer to post-treatment depression symptoms that have been untouched by therapy (shown neither reliable improvement or deterioration).

Behavioural activation (BA) – brief, time-limited psychological treatment for depression that uses behavioural principles to increase positive reinforcement of non-depressed behaviours and reduce avoidance that maintains depression.

Cognitive behavioural therapy (CBT) – brief, time-limited psychological treatment for depression focused on developing coping strategies through the use of techniques that challenge and change maladaptive thoughts and behaviours.

National Institute for Health and Care Excellence (NICE) – public body that provides guidelines for healthcare provision in the English National Health Service.

Improving Access to Psychological Therapies (IAPT) - nationwide program in England that delivers NICE recommended psychological therapy for common mental health problems via a stepped-care system. Administers the IAPT minimum dataset to collect routine outcomes from every contact session.

Low intensity (LI) – brief guided self-help interventions delivered at Step 2 of IAPT services by trained Primary Care mental health workers called Psychological Wellbeing Practitioners (PWPs; e.g., BA, cognitive restructuring, exposure, sleep hygiene, worry and panic management).

High intensity (HI) – psychotherapies delivered at Step 3 of IAPT services by trained and accredited therapists (e.g., CBT, BA, person centred counselling, couple counselling for depression, counselling for depression, psychodynamic interpersonal therapy, dynamic interpersonal therapy).

Patient Health Questionnaire-9 (PHQ-9) – a nine item self-report questionnaire designed to detect depression within primary care settings (scored between 0-27). Administered as part of the IAPT minimum dataset for routine outcome monitoring.

Generalized Anxiety Disorder-7 (GAD-7) – a seven item self-report questionnaire designed to detect anxiety within primary care settings (scored between 0-21). Administered as part of the IAPT minimum dataset for routine outcome monitoring.

Work and Social Adjustment Scale (WSAS) – a five item self-report measure of functional impairment as a result of mental health problems (scored between 0-40). Administered as part of the IAPT minimum dataset for routine outcome monitoring.

NOTES ON INCLUSION OF PUBLISHED WORK

Work that has contributed to two of the chapters in this thesis has been written up as manuscripts and accepted for publication. The co-authored papers that contain contributions from the present thesis are referenced below. The information from the published works has been presented in the following chapters in a format to fit within this body of work. Therefore they are not identical to the published papers (although there is some replication).

Chapter 2: Simmonds-Buckley, M., Kellett, S., & Waller, G. (2019). Acceptability and efficacy of group behavioural activation for depression among adults: A meta-analysis. *Behavior Therapy.* https://doi.org/10.1016/j.beth.2019.01.003

Chapter 5: 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

CHAPTER 1

Evidence-Based Psychological Treatment of Depression

The objective of this PhD was to develop understanding of when and why patients fail to respond (a phenomenon that will be termed 'stasis') to an evidenced based psychological treatment for depression. Outcomes for group behavioural activation (BA) treatment were investigated over four empirical studies, with the aim of predicting and reducing stasis outcomes. This first chapter will introduce the main topics underpinning the body of research. First, the symptoms and impact depression will be outlined and then the psychological theoretical models of depression will be described. Second, the development of BA as an effective intervention for depression will be summarised and the general background to the psychological treatment of depression in the United Kingdom (UK) will be provided. Finally, the issue of depression treatment nonresponse in real-world services will be introduced. The chapter will conclude by pulling all these topics together to outline the aims of each of the remaining chapters comprising the thesis.

1.1 Depression

1.1.1 Prevalence and symptoms

Depression, described dismissively by Seligman (1973) as the "common cold" of psychological disorders, currently affects over 300 million people worldwide (World Health Organisation [WHO], 2017). Depression is estimated to have an annual prevalence of approximately 7% (1 year) and a lifetime prevalence of 11% (Lim et al., 2018), with women affected almost twice as frequently as men (Seedat et al., 2009). Depression is a debilitating and distressing condition that has become one of the most burdensome diseases and principal causes of global disability (Ferrari et al., 2013; Murray & Lopez, 1996; Harvey A. Whiteford, Ferrari, Degenhardt, Feigin, & Vos,

2015). Depressive disorders are characterised by persistent low mood and/or diminished pleasure in activities (Otte et al., 2016). Additional symptoms include other emotional problems such as feelings of worthlessness or guilt and suicidal ideation, or physical symptoms such as a lack of energy, sleep disturbance, appetite problems, slowed thoughts and movement and difficulty concentrating (Diagnostic and Statistical Manual of Mental Disorders, 5th ed. [DSM-5]; American Psychiatric Association [APA], 2013). As a result, individuals suffering with depression experience significantly impaired social and work functioning (Von Korff, Ormel, Katon, & Lin, 1992).

The development and course of depression is heterogeneous in nature. It can develop at any stage of life, from early childhood to later years, although typically the first episode of depression occurs in early adulthood (Fava & Kendler, 2000). Presentation can vary in terms of severity, ranging from mild to severe. About two-thirds of individuals experience depression as a time-limited episode reaching recovery within 12 months of onset (Melartin et al., 2004), while a small subset seldom experience relief from their symptoms (Walker & Druss, 2015). The likelihood of full recovery is greatest following the first episode of depression, with recent onset a predictor of recovery (Boland & Keller, 2002). However, depression has a chronic nature, meaning recurring episodes are common with subsequent episodes and sub-threshold post-treatment symptoms increasing the risk of future relapse (Wojnarowski, Firth, Finegan, & Delgadillo, 2019).

1.1.2 Depression diagnoses

The heterogenous nature of the disorder is reflected in the use of several different diagnoses for depression. Table 1.1 outlines the different diagnoses of depressive disorders. For the purpose of this thesis, investigation has been focused around the treatment of unipolar depression.

Table 1.1. *Overview of depression diagnoses*

Diagnosis	Overview
Unipolar depression (also called major depressive disorder [MDD])	Persistent presence of five or more diagnostic symptoms (see section 1.1.1 above) for at least two weeks.
Persistent depressive disorder (PDD)	Chronic depression symptoms that last longer than two years and include at least two symptoms of either sleep problems, fatigue, appetite issues, low self-esteem, poor concentration or hopelessness.
Sub-syndromal depression	Enduring sub-threshold depressive symptoms that do not meet the full criteria for major depression.
Treatment-resistant depression	Depression that has not responded to at least two courses of treatment (typically refers to treatment with medication).
Bipolar disorder	Extreme swings in mood, with periods of depression (low and lethargic episodes) and mania (high and overactive episodes), often lasting several weeks at a time.
Psychotic depression	Depressive symptoms with additional psychotic symptoms, including hallucinations, delusions or paranoia.
Seasonal affective disorder	Periods of depression in the winter months when there are less hours of daylight. Symptoms improve during the summer months.

1.1.3 Individual and societal impact

The impact of depression is considerable, both in terms of emotional and physical suffering on an individual level, but also for the wider society. At the individual level, declines in work, social and physical functioning caused by the symptoms of depression are associated with reduced quality of life (Papakostas et al., 2004). Cases of chronic, persistent depression are also linked to increased financial stresses due to the difficulties in engaging with steady work (Maciejewski, Prigerson, & Mazure, 2000). Risk of mortality is high with disability and death resulting from depression (and other common mental health disorders) contributing to the loss of more health life years than cancer or cardiovascular disease (World Health Organisation [WHO], 2004). The burden of disease is further compounded by people with depression

also having an increased risk of developing serious physical health conditions, including cardiovascular disease and diabetes (Whooley & Wong, 2013). In addition, episodes of depression are associated with a heightened risk of suicide, with half of all suicides thought to be committed when an individual is in a depressive episode (Chesney, Goodwin, & Fazel, 2014).

On a wider, societal level, the prevalence of depression has a significant economic impact. The incapacitating nature of depression results in increased sickness absence, reduced productivity and lost employment (Almond & Healey, 2003; Kessler, Greenberg, Mickelson, Meneades, & Wang, 2001). Consequently, there is an indirect cost to society, due to lost taxes and an increased number of people on welfare and benefits (Knapp & Ilson, 2002). In addition, there are the direct healthcare costs to provide effective treatments for depression, as well as the increased healthcare usage due to physical health problems related to depression (Cuijpers et al., 2007). Economic evaluations have shown that the indirect costs from productivity and employment losses greatly exceed the direct healthcare costs (Layard, 2006). By 2026, the overall annual cost of depression in England is expected to reach £12.2 billion (McCrone, Dhanasiri, Patel, Knapp, & Lawton-Smith, 2008).

1.1.4 Theories of depression

The broad and diverse nature of depression has resulted in the development of multiple underpinning theories, with no one model being universally accepted (Ramnerö, Folke, & Kanter, 2016). Table 1.2 summarises the prominent theories of depression. These theoretical frameworks have been used to inform the development of treatments for depression. While the biological model advocates treatment using medication, all the other theoretical perspectives have generated psychological interventions that target the factors believed to be maintaining depression symptoms.

Table 1.2. *Summary of prominent theoretical perspectives of depression*

Theoretical perspective	**Summary**
Biological	Biological models attribute depression to a genetic predisposition and are concerned with functionality of brain areas in relation to depressive symptoms. The role of the biochemical imbalances and structural abnormalities in the brain are investigated to help determine effective pharmacotherapy treatments (Goldstein, Potter, Ciraulo, & Shader, 2011).
Psychodynamic/ analytic	Psychoanalytic theory suggests depression is a reaction to loss (both actual and symbolic). Feelings of loss and repressed anger become internalised, affecting an individual's self-esteem and triggering childhood experiences of loss to be re-lived (Freud, 1917).
Learned helplessness	Seligman (1973) proposed depression is a learned response. Negative thinking evolves from the generalised perception that previous attempts to escape aversive situations had no effect. Feelings of uncontrollability interfere with the ability and motivation to learn new coping skills leading to stymied or limited attempts at change.
Behavioural	Behavioural accounts refer to the role of the environment in shaping depressive behaviour (Ferster, 1973). Onset and maintenance of depression is posited as a function of behaviour, resulting from the removal of positive reinforcement for non-depressive behaviours combined with negative reinforcement of maladaptive behaviours (Lewinsohn & Graf, 1973).
Cognitive	Beck's (1964) cognitive model postulates depression is a product of automatic and negatively biased thinking styles. Faulty beliefs and cognitive bias result in overly negative interpretations of situations (often ignoring evidence to the contrary), inducing feelings of hopelessness and worthlessness.
Interpersonal	Interpersonal perspectives highlight the role of social interactions in the maintenance of depression. Excessive reassurance seeking, as a result of interpersonal stress or uncertainty when depressed, can unintentionally elicit negative responses from support networks. The rejection experienced then exacerbates depressive cycles (Coyne, 2016).

1.2 Evidence-based psychological treatment for depression

Given the extent and impact of depression, treatments need to be effective and efficient. Therefore, considerable efforts have been made to develop an array of empirically validated psychological treatments that can effectively and rapidly alleviate depression symptoms. Depression treatment in the UK is guided by the National Institute for Health and Care Excellence (NICE) guidelines to ensure only suitably evidenced treatments are recommended (NICE, 2016). NICE-recommended psychological treatments for depression include;

- Cognitive behavioural therapy (CBT); intervention focused on developing coping strategies through the use of techniques that challenge and change maladaptive thoughts and behaviours (16-20 sessions)
- Behavioural activation (BA); intervention that promotes behavioural changes designed to increase positive reinforcement of non-depressed behaviours and reduce avoidance that maintains depression (16-20 sessions)
- Interpersonal therapy (IPT); intervention focused on attachment and social functioning with the goal of improving the quality of interpersonal relationships in order to reduce distress (16-20 sessions)
- Behavioural couples therapy; intervention directed at people in a relationship suffering with relational elements of depression. Uses behavioural principles to resolve emotional difficulties arising from the relationship that are maintaining depressive symptoms (15-20 sessions)
- Counselling for depression; manualised person-centred therapy that focuses on emotional difficulties in the context of intrapersonal understanding. Centred around helping patients to make sense of feelings and reflect on new meanings (6-10 sessions)

- Psychodynamic psychotherapy; intervention based around accessing, understanding and resolving deep-rooted unconscious conflicts and facilitating patients to make changes that will improve their decision making and interactions with others (16-20 sessions).

Despite the existence of these effective psychological interventions for depression, people are not always able to access treatment. In 2007, it was estimated that fewer than half of people in Britain with depression received treatment (McManus, Meltzer, Brugha, Bebbington, & Jenkins, 2009). Of those who did get treatment, only 10% were given a talking therapy (the rest were treated with medication only). To address this disparity, major reform of mental health treatment provision in England was undertaken in 2008, resulting in the nationwide roll-out of the Improving Access to Psychological Therapies (IAPT) program (Clark, 2011).

1.2.1 Improving Access to Psychological Therapies program

To enable people suffering with common mental health problems to access recommended evidence-based treatment, the IAPT program was introduced in England in 2008 (Clark, 2011). It was established under the premise that investing in services to enable better nationwide access to psychological therapies would be offset by the economic productivity losses associated with common mental health disorders (Layard & Clark, 2014). IAPT services are based on six criteria; 1) only evidence-based therapies that are NICE-recommended can be administered, 2) all therapists have to be fully trained in the treatments they deliver, 3) outcomes must be measured at every session, 4) all patients undergo an initial assessment to ensure allocation to appropriate treatment, 5) all therapists receive weekly supervision by a suitably-qualified supervisor, and 6) self-referrals are accepted alongside general practitioner (GP) referrals (Layard & Clark, 2014).

Interventions are delivered via a stepped-care system, with treatment options available at different levels of intensity (referred to as 'steps'; see Figure 1.1). Patients are initially offered the least intensive treatment suitable for their clinical presentation. (Bower & Gilbody, 2005). The notion is that for many mild to moderate symptoms a less intense (and therefore cheaper) treatment is sufficient for recovery (Firth, Barkham, & Kellett, 2015). Step one involves support provided by local General Practitioners (GPs) to patients when they initially present for help with their mental health. Support is given around identification, prevention and monitoring of symptoms.

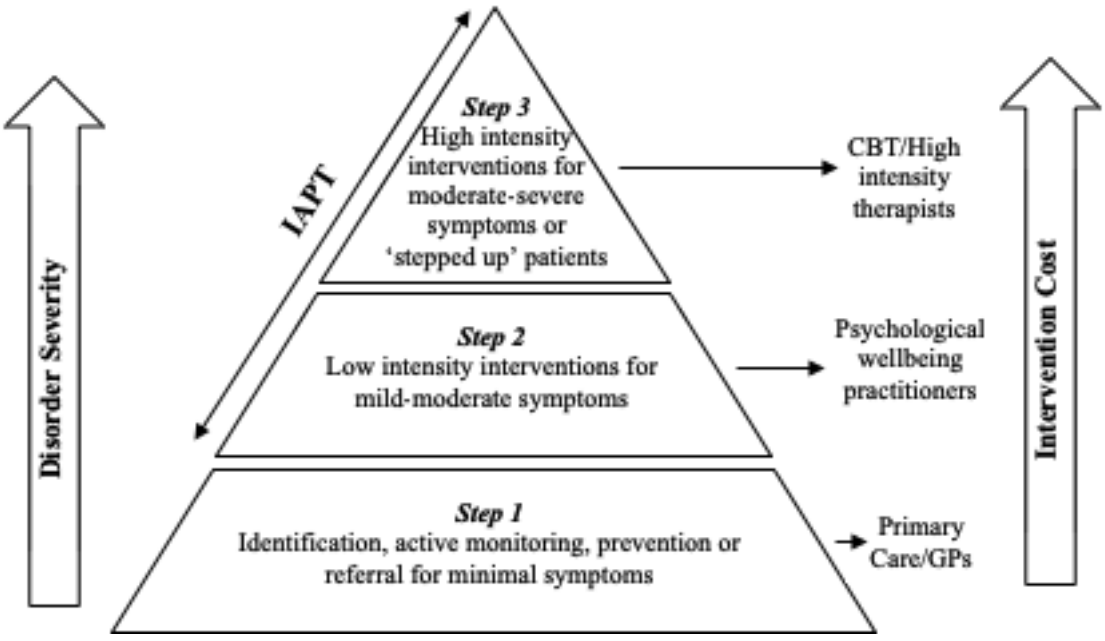


Figure 1.1. IAPT stepped-care model

If further treatment options are required, patients are referred to IAPT services (Step two and three). Low intensity (LI) treatments (Step two) consist of brief guided self-help interventions delivered by trained Primary Care mental health workers (called Psychological Wellbeing Practitioners [PWPs]) (Baguley et al., 2010). If patients do not respond to the low intensity treatment, or they initially present with more severe or complex symptoms, they are 'stepped up' and offered a longer, more intensive psychological treatment. High intensity (HI) treatments (Step three) consist of full

therapeutic treatment models delivered by trained and accredited therapists. Interventions are delivered in a variety of formats, including over the phone and internet (LI versions), one-to-one, and in groups (both LI and HI).

Of the NICE recommended 'talking therapies', CBT has the largest evidence base (in part due to being the most systematically tested intervention) and therefore is the most common treatment type delivered in IAPT services (House of Commons Library, 2018). Based on Beck's (1964) theory, CBT is grounded in the principle that thoughts, feelings and behaviours all interact to maintain depression. CBT for depression adopts a collaborative, time limited and formulation-driven approach to enable clients to recognise and then change depressogenic patterns, that recognises the role of the client's history in terms of being vulnerable to depression (Beck, 2011).

CBT's efficacy at treating depression has been repeatedly demonstrated across various methodologies (both practice-based and randomised controlled trials culminating in meta-analyses), modalities (individual, group, self-help and computerised), and treatment settings (in-patient and community) (DeRubeis et al., 2005; Driessen, Cuijpers, Hollon, & Dekker, 2010; Driessen & Hollon, 2010). CBT is a markedly effective treatment compared to passive controls, and is at least as effective as other psychotherapies and pharmacological medication (Butler, Chapman, Forman, & Beck, 2006; Hofmann, Asnaani, Vonk, Sawyer, & Fang, 2012). However, despite CBT often being referred to as the 'gold-standard' treatment (David, Cristea, & Hofmann, 2018), its predominance as the depression treatment of choice is not without criticism.

Critics argue that CBT is no more efficacious than other available psychotherapies (Cuijpers, 2017), and therefore should not be promoted over and above other treatment options. Treatment acceptability has also been questioned, with some evidence to suggest not all patients are able to engage with the treatment concepts (Hans & Hiller, 2013; Salmoiraghi & Sambhi, 2010). Furthermore, the monopoly of one type

of therapy restricts patient choice. In order to consider patients' preferences, multiple effective frontline treatment options should be readily available (Gelhorn, Sexton, & Classi, 2011). Finally, the considerable symptomatic change that often occur before the introduction of the cognitive components, has been seen as an indication that the cognitive elements of CBT may not be necessary for treatment to be effective (Jacobson et al., 1996; Longmore & Worrell, 2007). If directly changing cognitions is therapeutically redundant, then there may be scope for more parsimonious behaviourally oriented treatment options (especially when training and resources are limited).

1.2.2 Behavioural activation as a standalone treatment

Jacobson et al.'s (1996) landmark component study systematically compared the key components of CBT and demonstrated that BA was an effective standalone treatment. The delivery of exclusively BA content produced outcomes equivalent to the full CBT treatment at end of treatment and at follow-up, with two-thirds of patients experiencing improvement. The findings kick-started renewed interest in BA as a treatment for depression (Addis & Martell, 2004; Lejuez, Hopko, & Hopko, 2001). Martell, Addis and Jacobson (2001) refined the BA component into a manualised treatment for depression based on the principles of behaviour theory (Ferster, 1973). Although the BA intervention emerged as a result of Jacobson's study (1996), it shadowed Lewinsohn's pleasant events focused treatment developed 20 years previously (Lewinsohn & Libet, 1972; Lewinsohn, Sullivan, & Grosscup, 1980).

BA is clinically effective as a treatment in its own right and has been consistently shown to reduce symptoms of depression (Dimidjian et al., 2016; Dimidjian, Barrera, Martell, Muñoz, & Lewinsohn, 2011; Richards et al., 2016). A comprehensive evidence base, consisting mostly of studies of individually-delivered BA, demonstrates that BA is superior to antidepressant medication and exhibits

equivalent effect sizes to those observed in CBT trials (Cuijpers, van Straten, & Warmerdam, 2007; Dimidjian et al., 2006; Ekers et al., 2014; Hopko, Lejuez, LePage, Hopko, & McNeil, 2003; Mazzucchelli, Kane, & Rees, 2009). In response to the NICE review (NICE, 2009) deeming that the existing evidence for BA was not yet sufficient to warrant a frontline treatment recommendation for depression, the Cost and Outcome of Behavioural Activation versus Cognitive Behavioural Therapy for Depression (COBRA) non-inferiority trial was conducted (Richards et al., 2016). Outcomes for BA were matched with CBT, but were produced at a 21% reduced cost. The results of the COBRA trial provide compelling evidence to place BA alongside CBT in the treatment of depression in routine service delivery settings.

The key focus of BA is on the function of behaviours and the role of reinforcement (both positive and negative) in maintaining depression (Martell, Dimidjian, & Herman-Dunn, 2010). Avoidance typically reduces exposure to meaningful and pleasurable events, resulting in less positive reinforcement of non-depressed behaviours. While avoidant coping behaviours allow relief from negative stimuli in the short-term, they act as negative reinforcement for depressive behaviours in the long-term (Curran, Ekers, Mcmillan, & Houghton, 2012). BA adopts a positive, formulation-driven and contextual approach which uses activity scheduling as the primary mechanism of change practiced through the use of between-session work to enable clients reengage with their life (Martell et al., 2001). Increased activity can tackle avoidance and increase engagement with valued living, resulting in increased exposure to positive reinforcement of non-depressed valued behaviours and in turn reduce symptoms of depression (Lewinsohn & Libet, 1972).

1.2.3 Treatment delivery benefits of BA

BA as an independent treatment has additional benefits in terms of treatment provision. The underlying rationale is relatively simple and easy for patients to

understand, and the parsimonious BA techniques are easily disseminated to therapists and patients alike (Sturmey, 2009). As a result, BA remains effective when delivered by non-specialists who have only received brief training (Ekers, Richards, McMillan, Bland, & Gilbody, 2011; Pass, Hodgson, Whitney, & Reynolds, 2018). BA strategies are suited to a structured manualised treatment approach that can be delivered in simple and complex formats with similar outcomes (Ekers et al., 2014). Simple variants are comprised of basic activation strategies, whereas complex variants include additional techniques that contextualise the activation strategies (through functional analysis to target approach/avoidance behaviours in accordance with patient values).

BA can be easily implemented into stepped-care systems. It is delivered in both low intensity and high intensity forms in IAPT services, generally in a one-to-one format. BA is effective at treating all severities of depression (Dimidjian et al., 2006; Kellett, Simmonds-Buckley, Bliss, & Waller, 2017), is suitable in a range of community and in-patient treatment settings (Veale, 2008), and can be adapted for diverse and underrepresented populations or different age groups (Armento, McNulty, & Hopko, 2012; López et al., 2014; Mazzucchelli et al., 2009; Pass, Lejuez, & Reynolds, 2018; Pasterfield et al., 2014).

The content of BA sessions means it is an intervention that also lends itself well to delivery within a group format (Houghton, Curran, & Saxon, 2008; Porter, Spates, & Smitham, 2004). IAPT policy targets services to deliver treatment to 15% of the prevalence for common mental health disorders, with a view to reaching 25% of prevalence by 2025 (NHS England, 2016). However, the demand for depression treatment cannot be met by the workforce and resources available to provide one-to-one treatment (Shidhaye, Lund, & Chisholm, 2015; World Federation for Mental Health, 2012). Establishing effective group delivery of treatments is one strategy that could widen treatment accessibility and help IAPT services achieve their targets. BA

principles are easy to disseminate in a group context, and patients can benefit from additional group learning, normalising and peer support (Yalom & Leszcz, 2005). However, much less is known about group BA, because the evidence base has focused largely on one-to-one treatment. Understanding more about the effectiveness of BA when delivered in an organisationally efficient group format could therefore provide added value to mental health services.

1.3 Nonresponse to psychological treatment

Despite the development of effective psychological treatments for depression and implementation of the IAPT initiative to improve accessibility, interventions are not depression panaceas. Data typically suggest treatment response is approximately 40-60% (Gyani, Shafran, Layard, & Clark, 2013; Hopko, Magidson, & Lejuez, 2011; Lambert, 2011). However, it has been suggested that outcomes in routine practice are even lower (Hansen, Lambert, & Forman, 2002). Taking the inverse of treatment response estimates shows the remaining 40-60% (or more) finish treatment without experiencing clinical benefit. IAPT policy dictates a 50% recovery rate target, inherently reflecting that half of patients may not experience any change. To put that into context, currently 560,000 patients are treated by IAPT services every year (Clark, 2018). That translates to 280,000 patients annually likely to be left still suffering with poor mental health even after accessing treatment. Not much is known about whether or why patients will fail to experience any improvement in their depression symptoms. Perhaps understandably, research has focused on understanding treatment response (i.e., when there has been some change), rather than when it appears nothing has happened (i.e., symptoms have remained in a state of 'stasis'). However, understanding these stasis outcomes is becoming more central in light of psychotherapy research needing to focus more on how existing treatments can be improved.

1.3.1 Value of understanding stasis

Since the 1960s, there has been extensive investigation into psychotherapeutic treatments for depression, resulting in a comprehensive evidence base. The resulting conclusion has been that no one intervention is vastly superior, as all psychotherapies appear to be as effective as each other (often referred to as the Dodo bird verdict; Cuijpers, 2017). There is a growing consensus that as there is an array of equivalently effective interventions now available, the development of new treatments do not significantly improve on what is already available. Rather than directing efforts to developing new treatment, the focus should be on optimising the existing ones (Cuijpers, 2018). Taking that standpoint places a large importance on understanding stasis outcomes, as they provide a useful avenue for improving treatments already offered for depression.

Reducing stasis outcomes would also contribute to alleviating the impact of depression. Treatment non-responders experience significantly greater overall dysfunction and worse quality of life and well-being than those who respond (Mauskopf et al., 2009). Critically, failure to respond to an initial treatment intervention is likely to exacerbate feelings of hopelessness and increases the risk of suicidal ideation, with more suicide attempts among those patients with non-responding depression (Dold et al., 2018; Hawton, Casañas I Comabella, Haw, & Saunders, 2013). Patients who do not respond to the first round of treatment are also more likely to drop-out of treatment altogether or to have less motivation to engage in future treatment attempts (Meltzer et al., 2003; Ten Have et al., 2010). There are also implications for longer-term prognosis, as initial response to treatment is vital for reducing the risk of relapse. Early recovery after the initial onset of depression reduces the risk of later relapse, while longer duration of depressive episodes leaves individuals at greater risk of developing chronic depression (Hölzel, Härter, Reese, & Kriston, 2011). Immediate relapse is predicted by

the presence of residual depression symptoms after treatment completion (Wojnarowski et al., 2019), and patients who then require more treatment phases to reach recovery continue to exhibit higher rates of recurring depression (Rush et al., 2006). Evidently, there is a critical window in the acute-phase treatment of depression, and successful intervention is key to stymieing long-term effects and withdrawal from sources of treatment support.

Finally, fewer stasis outcomes would facilitate economic savings in society. Patients who show no response to treatment have increased usage of healthcare resources, including physical health services, producing higher costs of healthcare (Knoth, Bolge, Kim, & Tran, 2010). Non-response is also associated with lower rates of employment and, for those who are employed, greater rates of lost work productivity and more days of missed work (Knoth et al., 2010). Furthermore, turnover would be enhanced by fewer stasis outcomes, as more people treated successfully would shorten waiting lists. More time spent on a waiting list is associated with a worse treatment outcome for depression (Clark et al., 2017). Therefore, shorter wait times would also give other people a better chance of getting an improved outcome. Therefore, reducing the rate of stasis would not only alleviate individual suffering, but also ease the societal and economic burden associated with depression.

1.3.2 How to approach stasis investigation?

To effectively investigate stasis, there first needs to be a distinction of the type of outcomes that occur in the subset of people who do not benefit from treatment. Treatment failure can include no meaningful change, as well as an active deterioration in symptoms (Lambert, 2011). Clarification of what stasis refers to and a metric on how to distinguish between such outcomes would provide the initial foundation for investigation. After clarifying the concept of stasis, it would be useful to gain an understanding of what differentiates a treatment responder from a patient who

experiences a stasis outcome for specific treatments. Identifying what predicts risk of a stasis outcome could point to strategies that might improve outcomes for existing treatment. Similarly, pinpointing the active ingredient that makes a treatment work would be valuable in targeting treatment at people who are at risk of not benefitting. Finally, the discrepancy in outcomes observed between depression treatments delivered within randomised clinical trials (RCTs) and routine practice indicate stasis investigation should be focused on practice-based contexts. Stasis outcomes are more prevalent in real-world treatment delivery with rates of improvement up to three times lower in routine services (Barkham et al., 2008; Gibbons, Wiltsey Stirman, Derubeis, Newman, & Beck, 2013; Hansen et al., 2002). As IAPT services collect routine outcome data as standard, they provide a practice-based research network primed to enable investigation into stasis outcomes following psychological therapy.

1.4 Conclusion

In summary, BA delivered in groups is a promising treatment option to help meet the demand for depression treatment. However, considerable rates of non-response occur after even the most effective evidence-based treatments. This thesis therefore sought to explore factors relating to stasis outcomes after group BA treatment in routine practice, and to develop an intervention to improve treatment outcomes. The rationale and aims of the following six chapters are outlined below.

1.4.1 Aims of the thesis

- Due to the lack of clarity of the effectiveness for BA in groups, the aim of Chapter two is to first review and synthesise the evidence for BA treatments delivered in a group format. A meta-analysis is conducted on outcomes in both trial (efficacy) and naturalistic (effectiveness) contexts, in order to establish how

effective group BA treatment can be and how outcomes then translate into clinical practice.

- Chapter three provides a more detailed overview of the depression stasis phenomenon. The chapter defines a metric for capturing stasis, establishes the prevalence of stasis identified in the literature, and reviews the current evidence on factors associated with stasis outcomes.
- Chapter four consists of an empirical analysis of IAPT routine outcome data, to investigate treatment response after stepped-care delivery of BA interventions (at low and high intensity; one-to-one and in groups). The effect of intervention intensity, format and duration (treatment 'dose') on stasis outcomes is explored, as well as stasis risk predictors.
- Chapter five describes the development and empirical testing of an intervention to enhance an existing group BA treatment. The effect of the augmented therapy on end-of-treatment stasis outcomes are reported.
- Chapter six explores what mediates change in group BA. Increased behaviour in accordance with life values will be empirically evaluated as a process of change that produces reductions in depression during group BA.
- Finally, Chapter seven concludes by drawing on the findings from the literature review and the four empirical studies. It will discuss the overall implications of the research in terms of the treatment of depression, critique group BA as an intervention, and summarise what can be concluded about depression treatment stasis.

CHAPTER 2

Acceptability, Efficacy and Effectiveness of Group Behavioural Activation for Depression among Adults: A Meta-Analysis

In order to explore treatment non-response after group behavioural activation (BA), there first needs to be clarity about how effective the intervention can be and how those outcomes translate in real-word service delivery. The objective of the first empirical chapter is therefore to review and synthesise the evidence for group BA treatments to establish an aggregated treatment effect. This chapter reports the findings of a meta-analysis of the trial-based evidence, practice-based evidence and acceptability (using drop-out rates) for group BA in comparison to controls and other active psychological therapies.

2.1 Introduction

2.1.1 Behavioural treatment of depression

When a person is depressed, a widely observed symptom is behavioural avoidance and withdrawal, with these behavioural symptoms often contributing to the maintenance of low mood (Curran et al., 2012). Given this behavioural component, behaviour change has long been a treatment target in the psychotherapy of depression. The initial treatment phase of cognitive therapy for depression (Beck, Rush, Shaw, & Emery, 1979) focuses on behavioural techniques (i.e., activity scheduling and behavioural change) in order to initially lift mood, with evidence of associated early change in depressive symptoms (Ilardi & Craighead, 1994). Purely behavioural treatments for depression have existed since the 1970's and can be clustered under four models: Lewinsohn's pleasant events, focusing on increasing access to pleasant events through activity scheduling (Lewinsohn et al., 1980); Rehm's self-control therapy

(SCT), comprising three key elements of self-monitoring, self-evaluation and self-reinforcement (Rehm, 1984); Martell's contextual behavioural activation (BA), derived from the initial BA segment of Beck's cognitive behavioural therapy (CBT) for depression manual (Martell et al., 2001); and Lejuez's behavioural activation treatment for depression (BATD; Lejuez et al., 2001). Early versions of BA applied relatively simple methods (e.g., SCT), whilst more recent developments of BA (e.g., contextual BA) are more complex, due to incorporating functional analysis, problem solving, applying approach behaviours, rumination work, and aligning behaviour to values (Kanter et al., 2010).

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

2.1.2 BA delivered in groups

The importance of understanding the potential of BA as a group therapy relates to its delivery as well as its potential effects. BA works by adopting an 'outside-in' treatment approach, using pragmatic behavioural techniques to increase access to sources of positive reinforcement, that, in turn, reduce associated depressive thoughts

and feelings (Curran et al., 2012). BA is therefore often characterized as a pragmatic and parsimonious treatment for depression (Jacobson et al., 1996; Sturmey, 2009). As fewer treatment competencies are required, therapists can also be trained in a relatively short time (Ekers et al., 2011). The relative simplicity of BA also makes it well suited to group adaptation as behavioural treatment principles can be easily taught, grasped and implemented (Dimidjian et al., 2011). During group treatment, patients can additionally benefit from the peer support, normalizing and the learning opportunities created by group dynamics (Yalom & Leszcz, 2005). Groups are also organizationally efficient, as they optimize scarce therapeutic resources through low therapist to patient ratios (Kellett, Clarke, & Matthews, 2007).

A meta-analysis of group-based BA effectiveness has been conducted recently (Chan, Sun, Tam, Tsoi, & Wong, 2017), but had a broad raft of methodological problems. Only seven randomized controlled trials (RCTs) were identified, which does not represent the full evidence base of clinical trials of group BA (as will be seen below). The seven studies included were actually *individual* BA (Carlbring et al., 2013; Dimidjian et al., 2006; Ekers et al., 2011; Gawrysiak, Nicholas, & Hopko, 2009; Hopko et al., 2003; Moradveisi, Huibers, Renner, Arasteh, & Arntz, 2013; Pagoto et al., 2008). Finally, no mention of treatment acceptability issues was made. Any clinical conclusions concerning group BA drawn from the Chan et al. (2017) meta-analysis are therefore invalid. Indeed, it is noteworthy that the Chan et al. (2017) paper was redacted by the journal Editorial Board in December 2018, for the reason that the choice of data analysed was found to be inaccurate.

2.1.3 Focus of meta-analysis of group BA

This meta-analysis therefore focuses on the acceptability, efficacy and effectiveness of group BA for depression, and addresses four key questions. First, under what conditions is group BA most effective? If differing BA models are not equally

effective, that could suggest that different levels of treatment model complexity moderate outcome, and can identify which models are more suitable to group adaptation. Second, what is the optimum number of group BA sessions? Providing more treatment than required is wasteful of resources, whereas not providing enough treatment risks creating a 'revolving door' for therapy services (Hansen et al., 2002). The dose-response literature suggests a negatively accelerated association between number of sessions and improved outcome, with estimates of 13-18 sessions required to achieve a 50% recovery rate (Hansen et al., 2002; Harnett, O'Donovan, & Lambert, 2010). However, BA has shown significant reductions in depression after much briefer periods of treatment (Armento et al., 2012; Gawrysiak et al., 2009; Hopko, Robertson, & Carvalho, 2009).

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

Finally, to what degree do findings from group BA randomized control trials (RCTs) translate into real-world service settings? Whilst testing the efficacy of group BA using RCTs is of primary importance, it does not necessarily indicate how effective such group therapy is when delivered in naturalistic settings (Rothwell, 2005). The internally valid conditions of an RCT (e.g. patient exclusion, therapist supervision and treatment fidelity) differ widely from the externally valid conditions of routine practice (e.g. the comorbidity of typical patient populations; Seligman, 1995) and both contexts have their advantages and disadvantages. Practice-based studies of group BA conducted in routine practice settings are particularly vulnerable to the influence of confounding variables, such as selection bias (Morris & DeShon, 2002) or spontaneous recovery in the absence of a control condition (H. A. Whiteford et al., 2013). Therefore, this study sought to investigate whether effect sizes from group BA delivered in naturalistic settings (practice-based evidence; PBE) benchmarked as equivalent to the outcomes achieved in controlled clinical trials (trials-based evidence). Whilst some evidence suggests the outcomes achieved during routine practice outcomes are comparable to RCTs (Gibbons et al., 2010; Westbrook & Kirk, 2005), others have found inferior outcomes for naturalistic settings (Barkham et al., 2008; Schindler & Hiller, 2010).

2.1.4 Aim of meta-analysis

To summarize, this meta-analysis of group BA has four interlinked aims: (1) assess the effectiveness of group BA for depression symptomology and rates of recovery when compared to passive and active controls; (2) investigate moderators of group treatment effectiveness in terms of intervention and patient variables; (3) define the acceptability of group BA by calculating drop-out rates in comparison to passive and active controls; and (4) investigate whether outcomes for group BA established in optimal research settings (i.e., trials-based evidence) translate well into routine practice settings (i.e., practice-based evidence).

2.2 Method

2.2.1 Identification and selection of studies

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

2.2.2 Inclusion and exclusion criteria

Eligible studies were identified based on the following six criteria:

2.2.2.1 Participants

Adults aged 18 and over with a depressive disorder or elevated symptoms of depression. There was no limitation in terms of co-morbidity, as long as depression was a primary presenting problem. Studies containing child and adolescent participants, individuals with intellectual disability and participants with sub-clinical symptoms of depression were excluded.

2.2.2.2 Study design

Randomized controlled trials (RCTs), controlled clinical studies and uncontrolled (pre-post) clinical studies were included. In order to investigate the effect of BAG outcomes for controlled research trials (i.e. optimal delivery settings) versus routine practice (real world delivery), RCT studies were classified as trials-based evidence of BA in groups, whilst quasi-experimental and pre-post outcome studies in clinical service settings were classed as practice-based evidence of group BA.

2.2.2.3 Interventions

Studies were included if they used BA group treatment for depression. The methods of studies were analysed, and the intervention was labelled BA if, and only if, the study delivered a purely behavioural treatment. Therefore, studies were labelled BA when the treatment focused on the functional analysis of behaviour (in the absence of changing cognitions) and resultant behavioural change, in the pursuit of increasing positive mood. Therefore, mood-activity monitoring, activity scheduling and behavioural activation comprised the behavioural treatment components. The Mazzuchelli et al. (2009) BA treatment definitions were used for this review; i) *pleasant events* (Lewinsohn et al., 1980); ii) *self-control* (Rehm, 1984); iii) *contextual* (Martell et al., 2001); and iv) *BATD* (Lejuez et al., 2001). Minimum group size was defined as three or more participants in a group in a study. There was no limit on treatment duration or the setting of the intervention.

2.2.2.4 Comparators

Presence of a treatment comparator was not an inclusion requirement, thus allowing for the inclusion of uncontrolled studies. Controlled studies compared group BA with a range of control or active treatments. *Control* comparators provided patients with a waitlist, placebo or time-matched control. *Treatment as usual (TAU)* comparators provided standard treatment in routine care settings, such as hospital or Primary Care Physicians/General Practitioner care. *Active treatment* comparators were other active

psychotherapies, including cognitive therapy (CT), cognitive behaviour therapy (CBT), supportive therapy, psychodynamic therapy, problem solving therapy, assertiveness training and non-specific psychotherapy.

2.2.2.5 Accessibility

No language restrictions were applied, but a publicly available English language translation of the paper was an inclusion criteria. Unpublished studies and dissertations were included if available. Those studies that did not provide sufficient data to calculate effect sizes were excluded.

2.2.3 Outcome measures

2.2.3.1 Primary outcome

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

2.2.3.2 Secondary outcomes

When available, information on drop-out and recovery rates was extracted as dichotomous data. Drop-out rates were used as a proxy for treatment acceptability. This was defined as the percentage of non-completers during group BA and (where

applicable) control conditions. Non-completers were determined by the original study authors' definition. Recovery rates were the percentage of patients at end of treatment and/or follow-up that scored below the specified clinical threshold on the primary outcome measure. Recovery definition was determined by the original study authors' definition.

2.2.4 Quality assessment

To enable confident interpretation of the included studies' contributions to the overall conclusion, methodological quality of included studies was assessed. Study quality was rated using the Downs and Black (1998) tool, designed for both randomized and non-randomized studies. This tool comprises a 27-item checklist divided into five subscales assessing methodological quality - reporting (10 items), external validity (3 items), internal validity - bias (7 items), internal validity - confounding (6 items), and power (1 item). For the purpose of this review, the final item assessing power was modified to the same yes/no scale used for the other items, in order to indicate whether a power calculation had been performed. This decision was due to the commonly agreed uncertainty about how this item should be calculated (O'Connor et al., 2015). The quality assessment scale therefore ranged from 0-28; with higher scores indicating higher study quality. The primary author assessed all studies and two independent raters (graduate students) each assessed 20%. Studies were allocated to raters by sampling within each of the quartiles of the primary author's ratings. Inter-rater reliability was calculated using Cohen's kappa (J. Cohen, 1960) (where .21-.40 = fair agreement; .41-.60 = moderate agreement; .61-.80 = substantial agreement; .81-1.0 = almost perfect agreement; Landis & Koch, 1977). The kappas between the primary rater and the two independent raters were k=.67 and k=.76 respectively, indicating substantial agreement. Discrepancies in ratings were resolved through discussion to produce a final quality rating for each study.

2.2.5 Data extraction

Data were extracted by the primary reviewer and studies were coded on the following variables; methodological characteristics (study design/type, control group/type, active comparator, quality, analysis method [completers/intention to treat] publication date/status), intervention characteristics (number of sessions, length of sessions, group size, BA treatment type and treatment setting), and participant characteristics (population, age, gender, initial depression severity). Where data were available, outcomes for depression, recovery and drop-out rates were extracted at post-treatment and follow-up (8-weeks or the closest possible time point).

2.2.6 Effect sizes

Effect sizes and standard error terms were calculated according to study design using either the between-groups post-treatment method (for controlled studies) or the within-groups pre-post method (for uncontrolled studies). To aid comparability of effect sizes between designs, both methods were standardized using raw scores (Morris & DeShon, 2002). This method was chosen over standardization using the change score, due to one aim of the study being the comparison of group BA to other psychotherapies for depression. Effect sizes were interpreted according to Cohen's criteria, where 0.2 is indicative of a small effect, 0.5 a moderate effect, and 0.8 a large effect (Cohen, 1992).

2.2.6.1 Controlled between-group effect sizes

Standardized mean differences (SMDs) were computed for the difference between conditions at treatment completion for each comparison between BA and a comparator condition or psychotherapy. SMDs (Cohens *d*) were calculated by subtracting the mean post-treatment score of the comparator condition or psychotherapy from the mean post-treatment score of the BA intervention and dividing the result by the pooled standard deviation (SD) of both conditions at treatment completion. Due to

the risk of small-sample bias, the *J* correction was applied to convert SMDs to Hedges *g* (Hedges & Olkin, 1985).

2.2.6.2 Uncontrolled pre-post effect sizes

SMDs were computed for the difference between the pre-treatment and post-treatment scores, calculated by subtracting the mean post-treatment score from the mean pre-treatment score and dividing by the pre-test SD. The Hedges *g* correction was again applied to adjust for potential small-sample size bias. Within-groups variance calculations require knowledge of the correlation between pre-and-post scores to be able to account for the lack of independence. Only one within-groups study reported a pre-post correlation (Kellett et al., 2017), so a value of .7 was imputed for other studies. In the absence of reported values, this applied a conservative estimate and matched the one available correlation value reported. To account for spontaneous recovery in uncontrolled BA studies, a time effect was estimated from pre-post effect sizes for those studies with a waitlist control condition (Becker, 1988; Morris & DeShon, 2002). The aggregated effect was subtracted from the uncontrolled effect sizes to account for the waitlist time effect bias, to allow comparison of unbiased treatment effect estimates across study designs.

2.2.6.3 Dichotomous outcomes

Dichotomous data for recovery and drop-out rates were calculated as odds ratios (OR); percentage of recovery or drop-out from group BA in relation to the comparator condition or psychotherapy. Odds ratios were transformed into the logit for aggregation and back to the original unit for interpretation.

A hierarchical procedure was applied to effect size calculations - means and SDs were used wherever possible, followed by effect size data, dichotomous data, and finally *t* or *F*-scores. Controlled studies with sub-groups or multiple arms that were comparable were collapsed into one group (Fereidooni, Gharaei, Birashk, Sahraeian, &

Hoseini, 2015; Gallagher-Thompson et al., 2000; Kornblith, Rehm, O'Hara, & Lamparski, 1983; Rehm et al., 1981) using Cochrane's recommended method (Cochrane Collaboration, 2011). Studies with multiple comparators within one comparison which could not be collapsed were included separately, with the number of participants in the shared intervention group split evenly across comparisons (Besyner, 1978; Rehm, Kaslow, & Rabin, 1987; Shaw, 1977).

2.2.7 Meta-analysis

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

2.2.8 Subgroup and moderator analysis

Sources of heterogeneity within comparisons were investigated using planned subgroup and moderator analyses. Random-effects analyses were conducted using the restricted information maximum likelihood (REML) model from MetaF and MetaReg SPSS Macros (Wilson, 2010). Analogue to ANOVA was used to investigate eight

categorical variables: methodological design (within/between-group); study type (trials-based/practice-based evidence); control/therapy type (waitlist/TAU and CBT/other psychotherapy); quality (high/low); publication status; recruitment setting; type of BA; and population. Meta-regression was used to investigate five continuous variables: initial depression severity (standardized *Z*-scores); gender (proportion of males); number of group sessions; group size; and publication date. The beta-coefficient significance threshold was adjusted to $p < 0.01$ to account for multiple testing (Thompson & Higgins, 2002), and a minimum of 10 studies was required to investigate moderators within comparisons (Cochrane Collaboration, 2011).

2.2.9 Publication bias

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

2.3 Results

2.3.1 Study selection

After the removal of duplicates, searches identified 5335 records to be screened (Figure 2.1). Title and abstract screening identified 78 articles to be retrieved for full-text review. Upon review, 50 were excluded (reasons outlined in Figure 2.1) leaving a total of 28 studies meeting the inclusion criteria. One remaining study was identified as an outlier (Zemestani, Davoodi, Honarmand, Zargar, & Ottaviani, 2016) and excluded from the quantitative synthesis. This was due to a very large effect size ($d = 5.76$) in favour of group BA compared to waitlist. Removal of this single study was conservative

and favoured the null hypothesis; this was deemed appropriate to reduce the risk of over-estimation of overall effect of BA. (*Note*: numbered square brackets refer to study numbers in Table 2.1.)

Study details and quality ratings are presented in Table 2.1. Of the N=27 studies included, 18 were RCTs and nine were PBE studies (including five uncontrolled studies). Study quality ranged from 8-22. Overall study quality was sub-optimal; in particular, nearly all studies were poorly rated due to presence of confounding variables and lack of a power analysis (see Appendix B for full quality ratings). RCTs (M=15.21; SD=3.52) generated a higher methodological quality mean score than the PBE studies (M=13.22; SD=3.77). A median split of 14.5 then categorized high (N = 13) and low-quality (N=14) studies.

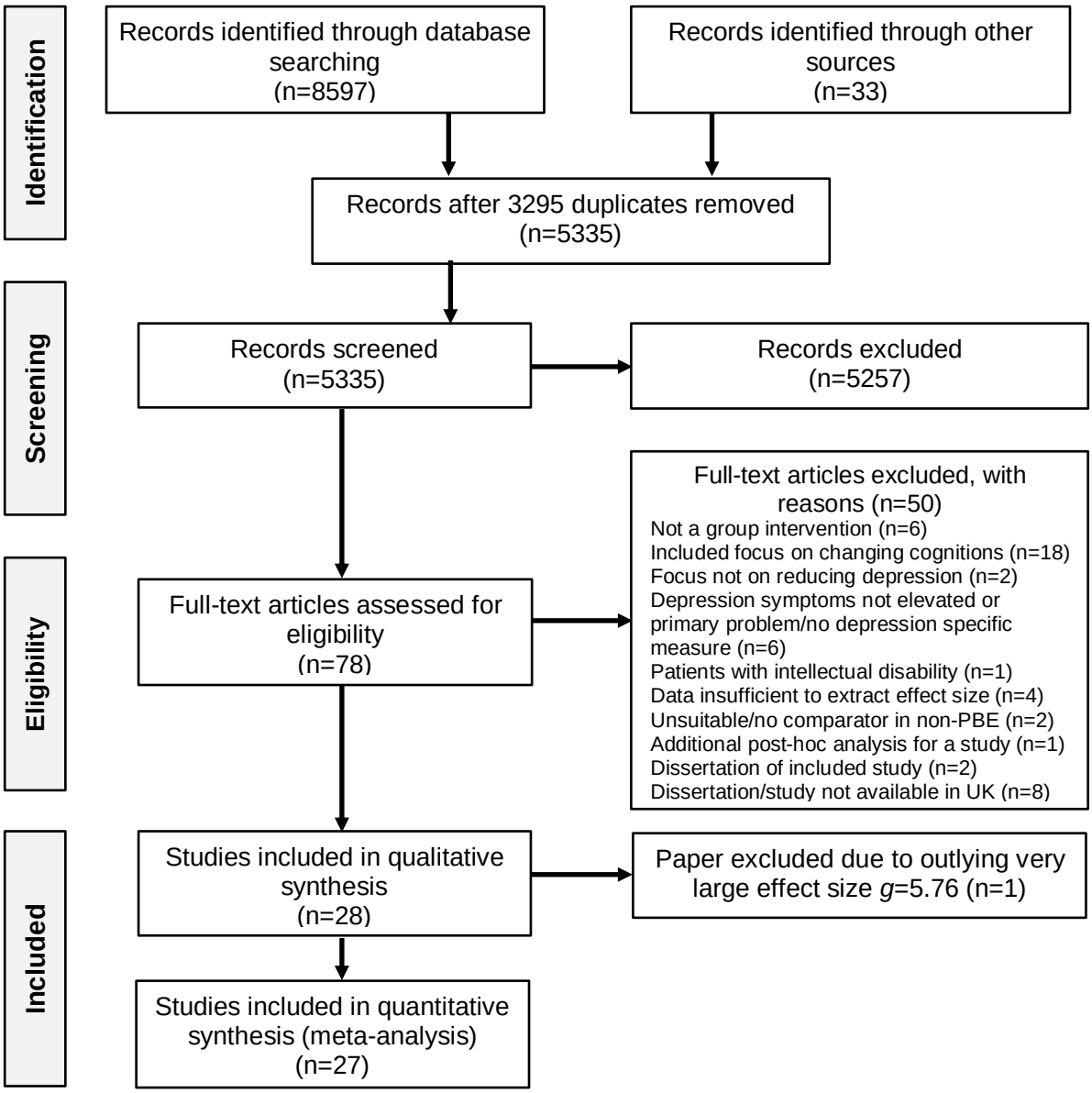


Figure 2.1. PRISMA flowchart of study selection.

Table 2.1. *Characteristics and quality ratings of the included studies*

Study First Author	Year	RCT/ PBE	Recruitment Setting	Population Age in years (mean) [range]	Sex (% male)	Interventions [*type of BA*]	Cell size at baseline	No. of sessions (duration in mins)	Measures	Initial depression severity	Follow-up (weeks)	BA Drop-out rate	Recovery Rate BA (response definition)	Quality Score
[1] Fuchs	(1977)	RCT	Community	Adults (28.8) [18-48]	0	1. Self-control therapy [*self-control*] 2. Non-directive group 1. Waitlist (8 weeks)	12 12 12	6 (120)	BDI, MMPI-D	Moderate	6	33%	100% (<11 BDI)	14
[2] Shaw	(1977)	RCT	University	Young adults (20.1) [18-26]	31	1. Behaviour modification [*pleasant events*] 2. Cognitive therapy 3. Non-directive group 4. Waitlist (4 weeks)	8 8 8 8	8 (120)	BDI, HRSD	Moderate	4	NR	25% (<10 BDI)	14
[3] Besyner	(1978)	RCT	Community	Adults (42.3) [NR]	29	2. Behaviour therapy [*pleasant events*] 3. Cognitive therapy 4. Non-specific therapy 3. Waitlist (4 weeks)	14 10 10 16	4 (120)	BDI	Moderate	4	NR	NR	18
[4] Barrera	(1979)	RCT	Community	Adults (36) [NR]	50	1. Activity scheduling [*pleasant events*] 2. Waitlist (4 weeks)	10 10	8 (120)	BDI, MMPI-D	NR	4 & 28	NR	NR	8
[5] Catanese	(1979)	RCT	University	Young adults (NR) [NR]	~27	1. Overt reward [*pleasant events*] 2. Covert reward 3. Overt punishment 4. Covert punishment 5. Social influence 6. Waitlist (4 weeks)	26 25 25 21 26 32	2 (30)	BDI, SRDS	Mild	2	12%	NR	10
[6] Rehm	(1979)	RCT	Community	Adults (NR) [21-60]	0	1. Self-control therapy [*self-control*] 2. Behaviour assertion skills	14 10	6 (120)	BDI, MMPI-D	Moderate	6	0%	79% (<11 BDI)	14
[7] Comas-Diaz	(1981)	RCT	Community	Low SES unemployed women (38) [NR]	0	1. Activity scheduling [*pleasant events*] 2. Cognitive therapy 3. Waitlist (4 weeks)	8 8 10	5 (90)	BDI, HRSD	Severe	5	NR	NR	13
[8] Gallagher	(1981)	RCT	Clinical (Outpatient)	Older adults (67.7) [NR]	45	1. Behaviour therapy [*pleasant events*] 2. Supportive therapy	14 14	10 (90)	BDI, MMPI-D, SRDS	Moderate	5	14%	NR	14
[9] Rehm	(1981)	RCT	Community	Adults	0	1. Self-control therapy	41	7	BDI,	Moderate	No F-U	13%	NR	15

				(39.2) [20-58]		(combined) [*self-control]* 2. Waitlist (7 weeks)	15	(NR)	MMPI-D, HRSD					
[10] Kornblith	(1983)	RCT	Community	Adults (37.9) [19-59]	0	1. Self-control therapy (combined) [*self-control]* 2. Psychodynamic therapy	34 5	12 (90)	BDI, MMPI-D, HRSD	Moderate	12	21%	65% (< BDI cut-off)	17
[11] Thompson	(1983 a) (1983 b)	PBE	Community	Older adults (68.4) [60-82]	2	1. Coping with depression class [*pleasant events*]	41	6 (120)	BDI	Mild	8	29%	NR	9
[12] Rehm	(1987)	RCT	Community	Adults (38.6) [NR]	0	1. Self-control therapy (behavioural target) [*self-control*] 2. Self-control therapy (cognitive target) 3. Self-control (cognitive-behavioural target)	35 35 34	10 (90)	BDI, MMPI-D, HRSD	Severe	24	NR	89% (no longer meet SADS/ RDC criteria for MDD)	17
[13] Lovett	(1988)	RCT	Community	Adults (59.3) [NR]	17	1. Increasing life satisfaction class [*pleasant events*] 2. Problem solving class 3. Waitlist (10 week)	23 20 19	10 (120)	BDI; RDC/SADS	Mild	No F-U	NR	NR	11
[14] Gallagher-Thompson	(2000)	RCT	Community	Adults (59.7) [31-81]	17	1. Increasing life satisfaction class [*pleasant events*] 2. Problem solving class 3. Waitlist (10 week)	56 59 46	10 (120)	RDC/SADS	Mild	No F-U	12%	79% (MDD RDC criteria)	18
[15] Brand	(1992)	PBE	Clinical (Inpatient)	Older adults (71.8) [NR]	13	1. Standard treatment plus behaviour therapy [*pleasant events*] 2. Standard treatment	27 26	8 (90)	BDI; HRSD; NOSIE-D	Moderate	No F-U	NR	44% (<9 BDI)	16
[16] van den Hout	(1995)	RCT	Clinical (Outpatient)	Adults (34) [20-59]	39	1. Standard treatment plus self-control therapy [*self-control*] 2. Standard treatment	15 14	12 (90)	SDS; VROPSOM	Mild	13	NR	NR	15
[17] Wright	(2003)	PBE	Clinical (Inpatient)	Adults Combat Veterans (NR) [40-60]	100	1. Standard treatment plus behavioural activation [*contextual*] 2. Standard treatment	24 21	8 (60-90)	BDI-II; HRSD; BHS	Severe	8 & 12	20%	88% (<16 HRSD)	18
[18] Porter	(2004)	PBE	Clinical (Outpatient)	Adults (44) [NR]	11	1. Behavioural activation [*contextual*] 2. Waitlist (4-6 weeks)	12 22	10 (95)	BDI-II; HRSD-R	Severe	12	13%	73% (no DSM diagnosis)	12

[19] Houghton	(2008)	PBE	Clinical (Outpatient)	Adults (42.5) [NR]	43	1. Behavioural activation [*contextual*]	42	10 (105)	BDI-II; CORE-OM	Severe	No F-U	19%	33% (RCSC)	11
[20] Daughters	(2008)	RCT	Clinical (Inpatient)	Adults (42.1) [NR]	63	1. LETS Act! plus standard treatment for substance abuse [*BATD*] 2. Standard treatment for substance abuse	20 19	6 (30-60)	BDI-II; HRSD; MINI	Moderate	2	5%	NR	22
[21] Norton	(2010)	PBE	Clinical (Inpatient)	Older adults 65+ (72) [65-81]	50	1. BATD plus standard hospital treatment [*BATD*] 2. Standard hospital treatment	24 25	8 (NS)	GDS	Mild	No F-U	NR	NR	18
[22] Magidson	(2011)	RCT	Clinical (Inpatient)	Adults (44.8) [NR]	66	1. LETS Act! [*BATD*] 2. Supportive counselling	29 29	5 (60)	BDI-II; HRSD-7	Mild	No F-U	4%	NR	20
[23] Magidson	(2014)	PBE	Clinical (Outpatient)	Adults (50.8) [NR]	54	1. Act Healthy [*BATD*]	4	8 (60)	HRSD-7	Mild	No F-U	NR	NR	10
[24] Wesson	(2014)	PBE	Clinical (Outpatient)	Adults (NR) [NR]	NR	1. MBARC [*contextual*]	37	10 (120)	PHQ-9	Moderate	12	7%	80% (<10 PHQ-9)	9
[25] Soleimani	(2015)	RCT	University	Young adults (22.9) [NR]	26	1. Group BA [*contextual*] 2. Cognitive therapy	14 13	8 (90)	DASS-42 D & A sub-scale	Severe	No F-U	12.5%	71% (< DASS cut-off)	15
[26] Fereidooni	(2015)	RCT	Clinical (Inpatient)	Adults (32.2) [NR]	100	1. Group BA plus standard treatment [*contextual*] 2. Standard treatment and placebo (combined)	8 16	7 (NS)	BDI-II	Severe	8 & 32	NR	NR	14
[27] Kellett	(2017)	PBE	Clinical (Outpatient)	Adults (33.1) [19-77]	41	1. Group BA [*contextual*]	71	8 (120)	PHQ-9	Moderate	No F-U	15%	28% (<10 PHQ-9)	16
Study not included in the meta-analysis														
[28] Zemestani	(2016)	RCT	University	Young Adults (24) [18-30]	36	1. Group BA [*contextual*] 2. Meta-cognitive group therapy 2. Waitlist (8 weeks)	15 15 15	8 (90)	BDI-II	Severe	12	0%	NR	20

Note: Abbreviations: NR: not reported; LETS Act!: Life Enhancement Treatment for Substance Abuse; BATD: behavioural activation treatment for depression; BDI-I/II: Beck Depression Inventory; MMPI-D: Minnesota Multiphasic Personality Inventory – Depression Scale; HRSD: Hamilton Rating Scale for Depression; SRDS: Zung Self-rating Depression Scale; RDC/SADS: Research Diagnostic Criteria/Schedule for Affective Disorders and Schizophrenia; VROPSOM: Dutch version of Depression Adjective Checklist; GDS: Geriatric Depression Scale; PHQ-9: Patient Health Questionnaire; DASS-42: Depression Anxiety and Stress Scale; RCSC: reliable and clinically significant change; F-U: follow-up; MDD: Major Depressive Disorder.

2.3.2 Meta-analysis of group BA

Twenty-seven studies were included across two meta-analytic comparisons (Table 2.2).

Table 2.2. *Meta-analyses of studies investigating the effect of group BA*

Comparison	No. of comparisons	No. of patients	SMD/ OR	95% CI	I^2 (%)	Q	NNT
Group BA vs. Waitlist/TAU							
Symptom level post-treatment	21	830	0.72***	0.46 to 0.98	83%	117.64***	2.6
Symptom level follow-up	6	183	0.78***	0.48 to 1.08	0%	4.93	2.4
Recovery rate[a]	4	215	2.74***	1.47 to 5.08	0%	0.45	-
Drop-out[a]	6	387	0.62	0.33 to 1.17	9%	5.51	-
Group BA vs. Active Therapy							
Symptom level post-treatment	15	526	0.14	-0.18 to 0.46	63%	38.22***	12.7
Symptom level follow-up	10	240	0.32	-0.10 to 0.74	57%	21.00**	6.2
Recovery rate[a]	7	351	1.30	0.41 to 4.07	61%	20.42**	-
Drop-out[a]	7	370	0.71	0.37 to 1.34	0%	5.25	-

Note: $^{**}p < .01$; $^{***}p < .001$ [a] Indicates odds ratio. Positive effect size indicates in favour of group BA for continuous outcomes; Odds ratio >1.0 indicates in favour of group BA for dichotomous outcomes. TAU: treatment as usual; SMD: standardized mean difference; OR: odds ratio; CI: confidence interval; NNT: Numbers needed to treat.

2.3.2.1 Comparison 1: Group BA versus waitlist/TAU control comparators

2.3.2.1.1 Study characteristics

The control comparison included 21 studies of Group BA. Five studies were uncontrolled (adjusted for time effects using the Becker method, see section 2.2.6.2), while the remaining 16 studies compared group BA directly with control conditions - six studies used TAU and 10 used a waitlist control. TAU consisted of inpatient (N=5) and outpatient (N=1) standard treatment, with varying levels of daily to weekly contact during the study period. Nine studies were practice-based and 12 were RCTs. Participants were recruited from the community (N=8), Universities (N=2), clinical

services (N=11; outpatient (N=6) and inpatient (N=5)). Depression symptomology was assessed via self-report (N=11), clinician report (N=2), or a combination (N=8). The most commonly employed self-report outcome measure was the BDI or BDI-II (N=15), and the most commonly employed clinician-rated outcome measure was the HRSD (N=8). Follow-up duration ranged from 2-32 weeks across N=14 studies. The mean follow-up period was 6 weeks.

BA group studies were conducted on adults in the general population (N=16), students (N=2) and older adults (N=3). Mean depression severity at intake ranged between mild (N=7), moderate (N=8) and severe (N=5). One study did not report sufficient information to establish baseline severity. Three studies focused on treating a primary problem of depression in conjunction with co-morbid disorders (substance abuse, PTSD and HIV). BA treatment type included pleasant events (N=9), self-control (N=3), contextual (N=6) and BATD (N=3). Group sizes ranged from 4-10 participants, treatment duration ranged from 2-12 sessions, with session duration ranging from 30-120 minutes. Drop-out rates ranged between 5-33%, but were unreported in ten studies. Recovery rates ranged from 25-100%. Recovery was defined by use of clinical cut-offs on measures (N=6), MDD diagnosis (N=2), and reliable and clinically significant change rates (N=1). Intent-to-treat analysis was used in N=2 studies, with the remaining 19 studies using completers analyses.

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

Post-treatment outcomes from 21 studies contributed to this analysis, totalling N=830 participants (group BA N=537; control N=293). A time-effect estimate calculated from the pre-post effect sizes of waitlist controls (SMD = 0.1 in favour of symptom improvement) was subtracted from the five uncontrolled effect sizes before aggregation. The overall aggregated SMD was 0.72 (95% CI 0.41 to 1.03; Z = 4.91; $p<0.0001$) in favour of group BA, suggesting a significant moderate to large effect

(Figure 2.2). Group BA was effective at reducing depressive symptoms at treatment completion, when compared to waitlist and TAU controls. The NNT for group BA was 2.57; one out of every three patients experiences additional benefit from group BA when compared to controls at treatment completion. There was significant between-study heterogeneity contributing to large variation in effect size (I^2 = 83%; Q = 117.64, p=<0.0001).

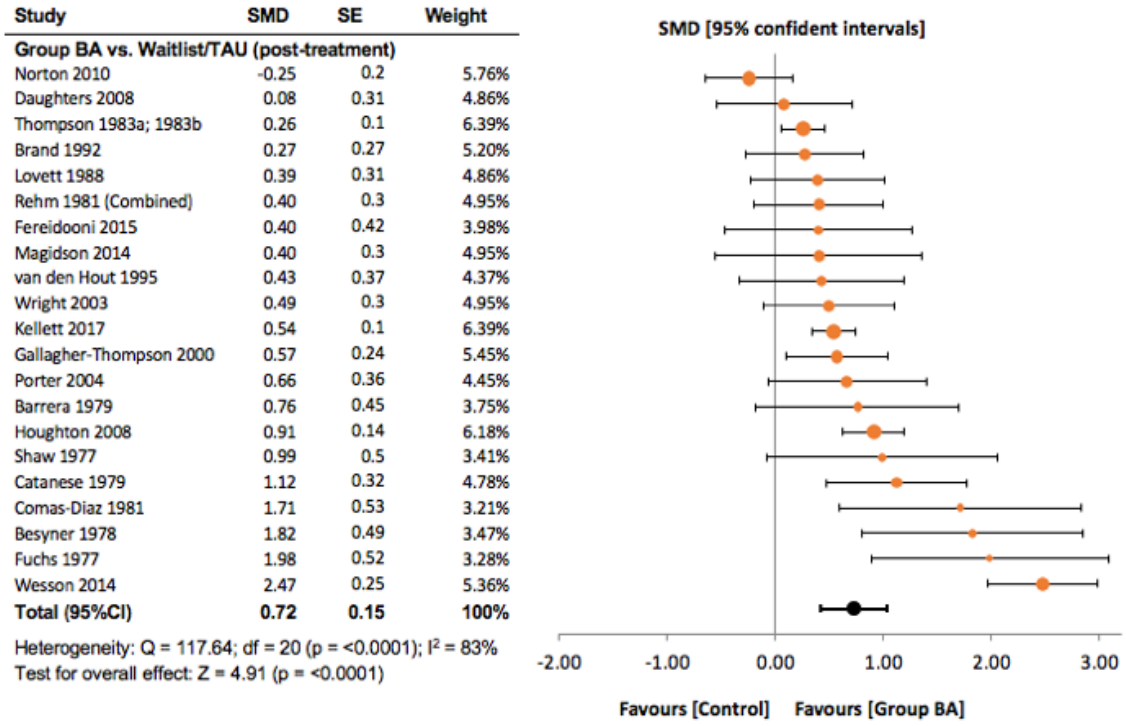


Figure 2.2. Forest plot of post-treatment depression symptom effect sizes for group BA versus waitlist/TAU.

Subgroup analysis and meta-regression results are displayed in Table 2.3. Methodological design did not produce significantly different effect sizes, indicating that study designs could be appropriately combined. Significant variation in effect size was associated with type of control condition. A large effect was observed for waitlist controls, but the effect for group BA was small and non-significant when compared to TAU. Trial-based studies were associated with large treatment effects and practice-

based studies were associated with moderate treatment effects – the comparison was non-significant. Low quality studies produced larger effects than high quality studies. Treatment effects were not significantly affected by publication status, study setting, type of BA or sample population. Substantial heterogeneity was evident in the majority of sub-groups. Meta-regression analyses found initial depression severity, gender, number of sessions, group size and publication date were not associated with improved treatment outcomes. Interpretation of moderator variables was limited by the low number of studies in subgroup arms and potentially confounded by the high correlation with TAU studies. For example, the significant association with study quality became non-significant when controlling for type of comparison.

Table 2.3. *Subgroup and meta-regression analysis of Group BA versus controls (post-treatment)*

Subgroup analysis		**No. of comparisons**	**SMD (*g*)**	**95% CI**	**I^2(%)[b]**	**P (between subgroups)**	**NNT**
Methodological design	Within-group (time-adjusted)	5	0.89	0.12 to 1.66	95***	0.48	-
	Between-group	16	0.66	0.36 to 0.95	65***		-
Control type	Waitlist	15	0.93*	0.62 to 1.23	85***	0.01*	2.04
	TAU	6	0.21	-0.26 to 0.69	22		8.47
Study type	RCT	12	0.82*	0.42 to 1.23	55*	0.52	2.28
	PBE	9	0.63*	0.22 to 1.05	91***		2.91
Quality	High (>14.5)	9	0.43*	0.04 to 0.83	65**	0.05*[a]	-
	Low (<14.5)	12	0.96*	0.61 to 1.32	87***		-
Publication status	Published	18	0.76*	0.45 to 1.07	82***	0.61	-
	Unpublished	3	0.54	-0.22 to 1.31	99***		-
Recruitment setting	Community	8	0.85*	0.40 to 1.29	74***	0.14	2.21
	Outpatient (clinical)	6	0.92*	0.45 to 1.39	91***		2.06
	Inpatient (clinical)	5	0.18	-0.36 to 0.71	29		9.87
	University	2	1.07*	0.47 to 1.99	0		1.82
BA type	Pleasant events	9	0.78*	0.36 to 1.20	66**	0.21	2.39
	Self-control	3	0.83*	0.08 to 1.58	74*		2.26
	Contextual	6	0.93*	0.45 to 1.42	91***		2.04
	BATD	3	0.06	-0.62 to 0.74	41		29.55
Population	Adults general	16	0.83*	0.53 to 1.13	81***	0.09	2.26
	Young adults	2	1.07*	0.15 to 1.98	63		1.82
	Older adults	3	0.09	-0.53 to 0.72	0		19.71
Meta-regression analysis		**No. of comparisons**	**B-coefficient**	**95% CI**	**SE**	**P**	**NNT**
Initial depression severity	(z scores)	19	0.02	-0.27 to 0.30	0.13	0.92	-
Gender	(% of males)	20	-0.01	-0.01 to 0.00	0.00	0.24	-
Number of sessions	(2-12 sessions)	21	-0.04	-0.18 to 0.09	0.06	0.51	-
Group size	(4-10 patients)	20	0.04	-0.13 to 0.21	0.07	0.64	-
Publication date	1977 – 2017	21	-0.01	-0.03 to 0.01	0.01	0.34	-

Note: *significant at $p < .05$ threshold; **significant at Bonferroni adjusted $p < .01$ threshold; ***significant at $p < .0001$. [a] Effect non-significant when controlling for control type; [b] P value of Q-statistic. Abbreviations: TAU: treatment as usual; SMD: standardized mean difference; SE: standard error; NNT: Numbers needed to treat; RCT: randomized controlled trial; PBE: practice-based evidence; BATD: BA treatment for depression.

Funnel plot inspection gave a slight suggestion of asymmetry. This indicates that smaller studies may have tended to produce larger effects in favour of group BA (Figure 2.3). However, the adjusted effect size produced by Trim and Fill imputation of missing data did not differ (0.72, 95% CI 0.41 to 1.03). Testing the extent of asymmetry via funnel plot regression showed sufficient symmetry of study distribution (B = -0.004, t(20) = -0.70, p=0.50). The 'small study influence' appeared limited, as the overall effect estimate using the 11 studies (50%) with the largest samples produced a similar SMD of 0.64 (95% CI 0.18 to 1.11).

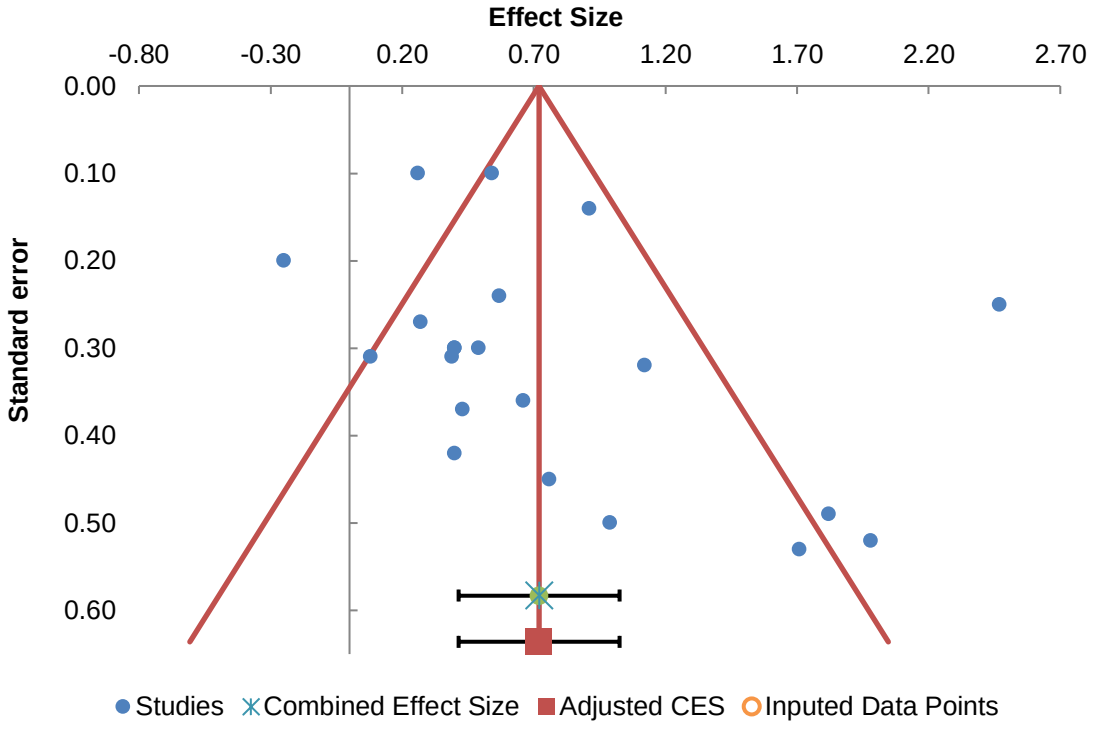


Figure 2.3. Funnel plot for group BA versus waitlist/TAU control post-treatment symptom level.

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

Six studies had follow-up comparisons with a total of N=183 participants (group BA N=89; control N=94). There was a large pooled SMD of 0.78 (95% CI 0.48 to 1.08; Z = 5.06; p<0.0001) in favour of the maintained effects of group BA at follow-up.

Group BA therefore appeared effective at sustaining improvement at follow-up compared to controls. The NNT was 2.39, indicating that at follow-up one out of just over every two participants experienced additional benefit from group BA compared to controls. Studies were statistically homogeneous (I^2 = 0%; Q = 4.93, p=0.42), even when taking a higher significance level threshold (p<0.1) to account for low power from the small number of studies. Limited variance between studies negated the need for further heterogeneity analysis. There were an inadequate number of studies (*k*<10) to test for publication bias.

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

Four studies reported recovery rates for 215 participants (group BA N =115; control N=100). Recovery rates were significantly higher following group BA than waitlist or TAU (group BA 59%, control 38%), producing a significant odds ratio of 2.74 (95% CI 1.47 to 5.08; Z = 3.19; p = 0.001). More participants recovered after receiving group BA than those allocated to a waitlist condition or receiving TAU in the service. All studies were statistically homogeneous (I^2 = 0%; Q = 0.45, p = 0.93).

Six studies reported drop-out rates for 387 participants (group BA N=215; control N=172). There was no difference in drop-out rates between group BA (16%) versus control conditions (20%), with a non-significant odds ratio of 0.62 (95% CI 0.33 to 1.17; Z = 1.48; p = 0.14). Patient drop-out rates were matched across group BA (16%), waitlist (18%) and TAU (24%). Between-study variance was minimal and not significant (I^2 = 9%; Q = 5.51, p = 0.36). Limited heterogeneity and the small number of studies reporting recovery and drop-out outcomes constrained further investigation into sources of variation in effect sizes. The number of studies of group BA reporting recovery and dropout rates were insufficient to perform any publication bias tests.

2.3.2.2 Comparison 2: Group BA versus other active psychotherapies

2.3.2.2.1 Study characteristics

Group BA was compared to other active psychotherapies in 12 studies across 15 comparisons. CBT/CT was the most common comparison psychotherapy (N=5). The treatment comparators included supportive psychotherapy, psychodynamic psychotherapy, non-directive psychotherapy, problem-solving and assertiveness training. All 12 studies were RCTs - eight recruited community participants, two University students, one out-patients, and one inpatients. Depressive symptoms were assessed via self-report (N=5), clinician-report (N=1) or in combination (N=6). The most commonly used self-report outcome measures were the BDI and BDI-II (N= 10) and the HRSD was the most commonly used clinician-rated measures (N=5). Follow-up duration ranged from 4-24 weeks (N=8), with follow-up offered at 8-weeks on average.

Samples included nine studies with adults, two on young adults and one on older adults. Initial depression severity at the groups ranged from mild (N=3), moderate (N=6) and severe (N=3). Two studies included co-morbid disorders (anxiety and substance abuse). Group treatments in these clinical trials consisted of pleasant events BA (N=6), self-control focus (N=4), contextual BA (N=1) and BATD (N=1). Treatment duration ranged from 4-12 sessions, with sessions lasting between 60-120 minutes. Group size ranged from 4-10 participants and dropout rates varied from 0-33%. Five studies did not report their drop-out rates. Recovery rates varied from 25-100%. The definition of recovery was clinical cut-offs (N=5) or MDD diagnosis (N=2). Only one trial used intent-to-treat analysis and completers analysis was used in the remaining studies (N=11).

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

Post-treatment outcomes from 15 comparisons contributed to this analysis, totalling N=526 participants (group BA N=254; active psychotherapies N=272). There

was no difference in the effect of group BA when compared to other psychotherapies, with a non-significant SMD of 0.14 (95% CI -0.18 to 0.46; Z = 0.87; p = 0.38) (Figure 2.4). Group BA was as effective at reducing depressive symptoms as other active psychotherapies. The NNT for group BA was 12.68. This indicates one out of every 13 participants would experience additional benefit post-treatment from being in a group BA treatment, when compared to other psychotherapies. Between-study heterogeneity was moderate and significant (I^2 = 63%; Q = 38.22, p=0.0005).

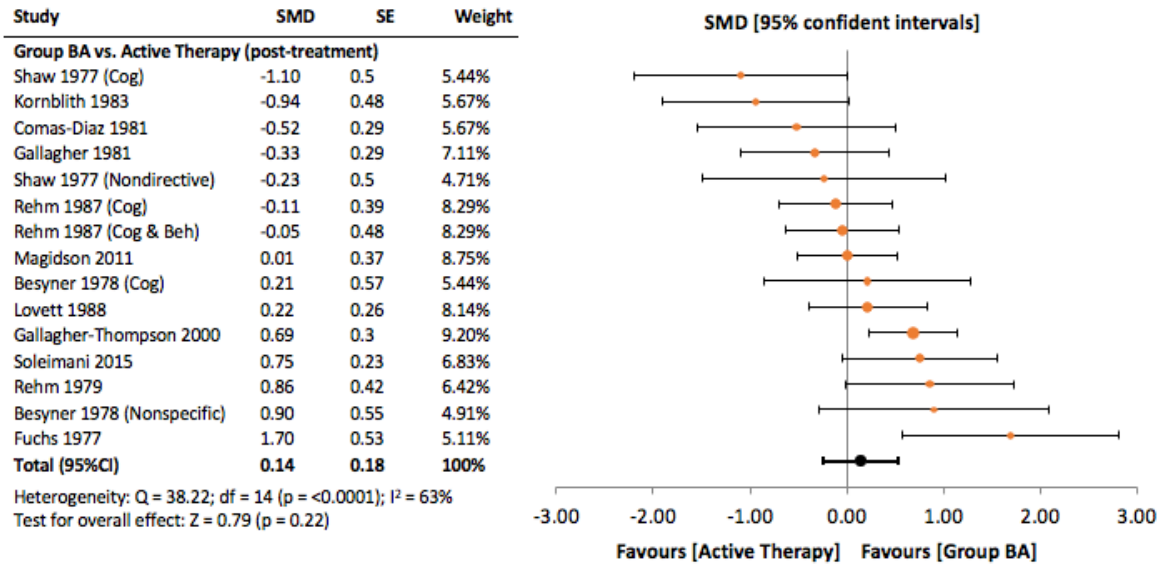


Figure 2.4. Forest plot of post-treatment depression symptom effect sizes for group BA versus active therapy.

Further investigation with subgroup analysis and meta-regression is displayed in Table 2.4. Subgroup analyses of different psychotherapies found that group BA compared to CBT/CT therapies resulted in a minimal non-significant effect. When compared to other psychotherapies, group BA resulted in a small effect in the direction of favouring BA, though it did not meet significance. Significantly differing effect sizes were not evident when comparing high versus low-quality studies, published versus unpublished studies, different recruitment settings, types of BA or the sample populations. There was moderate heterogeneity present in most of the subgroups. Meta-

regression analyses found no evidence of variation in effect sizes according to initial depression severity, gender, number of sessions, group size or publication date. However, low power from the small number of studies available indicates that caution should be applied to these moderator interpretations.

Funnel plot inspection did not suggest evidence of asymmetry (Figure 2.5), with funnel plot regression providing evidence of a symmetrical study distribution (B = 0.005, t(14) = 1.09, p=0.30). Trim and Fill imputation estimated one study was missing and produced an adjusted overall effect estimate of 0.21 (95% CI -0.18 to 0.61), representing a slight increase in favour of group BA, albeit still not reaching significance. The removal of the smallest studies reduced the overall effect estimate to 0.08 (95% CI -0.33 to 0.50), indicating minimal influence of a small study effect. These observations indicate a minimal effect of publication bias.

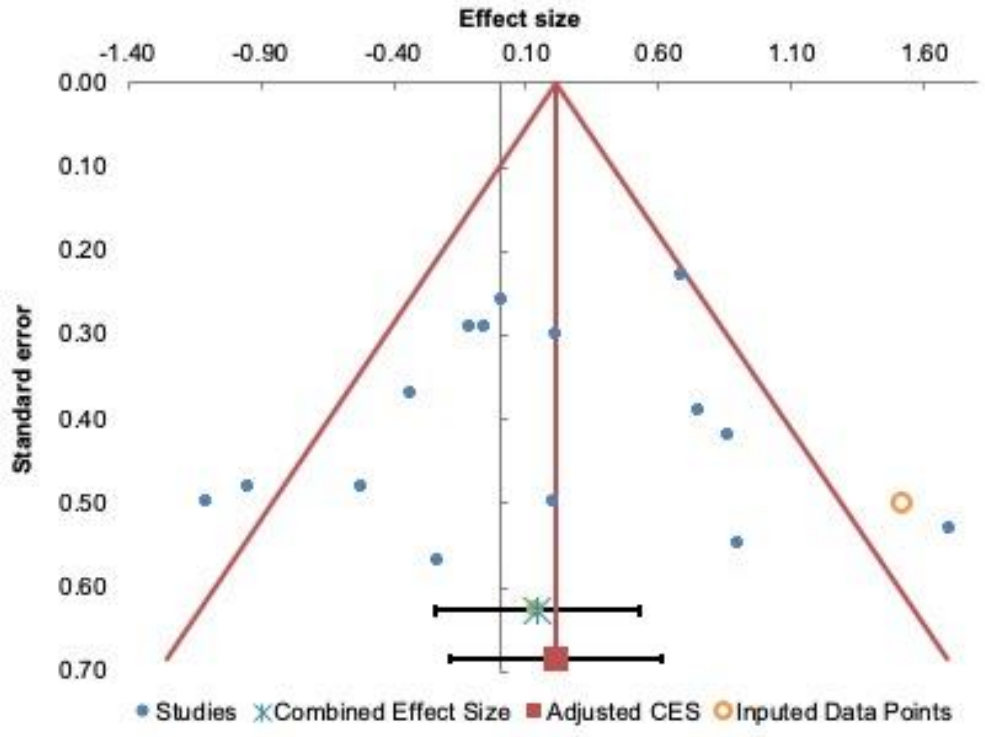


Figure 2.5. Funnel plot for group BA versus active therapy post-treatment symptom level.

Table 2.4. *Subgroup and meta-regression analysis of Group BA versus active therapy (post-treatment)*

Subgroup analysis		**No. of comparisons**	**SMD (*g*)**	**95% CI**	**I^2(%)[b]**	**P (between subgroups)**	**NNT**
Therapy type	CBT/CT	6	-0.10	-0.59 to 0.39	49	0.22	-17.74
	Other therapies	9	0.30	-0.10 to 0.70	67**		5.95
Quality	High (>14.5)	8	0.18	-0.26 to 0.62	57*	0.78	-
	Low (<14.5)	7	0.09	-0.42 to 0.59	72**		-
Publication status	Published	13	0.09	-0.25 to 0.44	67***	0.41	-
	Unpublished	2	0.54	-0.47 to 1.54	0		-
Recruitment setting	Community	10	0.27	-0.13 to 0.68	65**	0.70	6.61
	Outpatient (clinical)	1	-0.33	-1.56 to 0.90	-		-5.42
	Inpatient (clinical)	1	0.01	-1.11 to 1.13	-		177.24
	University	3	-0.12	-0.91 to 0.67	77*		-14.79
BA type	Pleasant events	8	0.02	-0.45 to 0.49	61*	0.50	88.62
	Self-control	5	0.23	-0.34 to 0.80	78**		7.74
	Contextual	1	0.75	-0.53 to 2.03	-		2.48
	BATD	1	0.01	-1.13 to 1.15	-		177.24
Population	Adults general	11	0.24	-0.13 to 0.62	63**	0.53	7.42
	Young adults	3	-0.12	-0.90 to 0.67	77*		-14.79
	Older adults	1	-0.33	-1.55 to 0.89	-		-5.42
Meta-regression analysis		**No. of comparisons**	**B-coefficient**	**95% CI**	**SE**	**P**	**NNT**
Initial depression severity	(z scores)	14	-0.43	-1.01 to 0.16	0.30	0.15	-
Gender	(% of males)	15	0.00	-0.02 to 0.01	0.01	0.62	-
Number of sessions	(4-12 sessions)	15	-0.09	-0.22 to 0.04	0.07	0.17	-
Group size	(4-10 patients)	15	-0.09	-0.31 to 0.13	0.11	0.43	-
Publication date	1977 - 2015	15	0.01	-0.02 to 0.04	0.01	0.44	-

Note: *significant at $p < .05$ threshold; **significant at Bonferroni adjusted $p < .01$ threshold. Positive effect size indicates in favour of group BA. Abbreviations: CBT/CT: cognitive behavioural therapy/cognitive therapy; SMD: standardized mean difference; CI: confidence interval; SE: standard error; NNT: Numbers needed to treat; BATD: behavioural activation treatment for depression.

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

Eight studies performed 10 follow-up comparisons with a total of 240 participants (group BA N=122; active psychotherapies N=118). There was a small SMD of 0.32 favouring group BA, but this was not significant (95% CI -0.10 to 0.74; Z = 1.50; p = 0.13). Group BA and the other active psychotherapies therefore produced similar maintained treatment effects at follow-up. The NNT was 6.16, indicating that by follow-up one out of every six patients experienced additional benefit from group BA. Significant between-study heterogeneity was observed representing a moderate level of variance (I^2 = 57%; Q = 21.00, p=0.01). Five comparisons of group BA versus CBT/CT produced similar effects at follow-up (SMD = 0.07; 95% CI -0.41 to 0.55; Z = 0.27; p = 0.78). BA was compared to other psychotherapies in the remaining five studies at follow-up and showed a moderate (but non-significant) effect in favour of group BA (SMD = 0.59; 95% CI -0.09 to 1.69; Z = 0.27; p = 0.09). The small number of studies prevented any further exploration of moderating variables and publication bias.

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

Seven studies with nine comparisons reported recovery rates for 351 participants (group BA N=169; other psychotherapies N=182). There was no difference in recovery rates following group BA compared to other psychotherapies (69% during group BA versus 61% during other active psychotherapies) with a non-significant odds ratio of 1.30 (95% CI 0.41 to 4.07; Z = 0.44; p = 0.66). The recovery rate for group BA was comparable to that of other active psychotherapies. Group BA versus CBT/CT had a non-significant OR of 0.39 in favour of CBT/CT (95% CI 0.04 to 4.15; Z = 0.77; p = 0.44). Group BA versus all other psychotherapies had a non-significant OR of 2.72 in favour of group BA (95% CI 0.83 to 8.85; Z = 1.66; p = 0.10). The studies were

significantly heterogeneous (I^2 = 61%; Q = 20.42, p=0.009), but there were insufficient studies to examine moderators of variation in effect size or to test publication bias.

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

2.4 Discussion

The objective of the present meta-analysis was to quantify the effectiveness of group BA for depression when compared to passive controls, when compared to other active psychotherapies, and when delivered in routine practice settings. This review was conducted in order to provide guidance to clinicians in terms of offering a choice of evidence-based treatments for depression. Particularly, this meta-analysis has provided a scientifically credible quantitative study of the evidence base for group BA, in contrast to the review conducted by Chan et al. (2017).

2.4.1 Summary of group BA outcomes

In relation to the first aim, the results provide support for the efficacy and effectiveness of group BA in the treatment of depression across trial and routine service contexts. Compared to waitlist and TAU comparators, group BA facilitated significantly reduced depressive symptoms at treatment completion and at follow-up, improved recovery rates and equivalent drop-out rates. One out of every three participants would

expect to experience additional benefit from receiving group BA, when compared to waitlist or TAU. Compared to other routinely used psychotherapies for depression (including CBT), group BA produced equivalent outcomes at treatment completion and at follow-up, with matched recovery and dropout rates. The results therefore indicate that group BA offers an acceptable, equivalent and useful treatment option in the treatment of depression, both in the short and medium-term.

The moderate to large effects in the reduction of depressive symptoms and increased clinical recovery rates suggests that BA principles translate well into group format settings. The translation of BA theory to group delivery supports the notion that the principles of BA remain simple and parsimonious to deliver, regardless of context (Jacobson et al., 1996; Sturmey, 2009). The magnitude of the group BA treatment effect compared to controls is similar to the effect observed (SMD 0.70-0.87) for individually-delivered BA (Cuijpers, van Straten, et al., 2007; Ekers et al., 2014; Mazzucchelli et al., 2009). Likewise, the group BA treatment effect is comparable to the individual BA versus other treatments effect (SMD 0.13; Cuijpers, van Straten, et al., 2007). Furthermore, benefits of group BA were still evident at follow-up, suggesting durability of outcomes for this behavioural intervention. Therefore, it is reasonable to conclude that allocating to group BA is not detrimental to participant outcome and that participants are as likely to engage in group treatment as individual work.

2.4.2 Moderators of group BA effectiveness

Analysis of the variation between studies enabled investigation of moderators of group BA effectiveness in order to explore which patients find group BA beneficial and to define the conditions in which group BA works best. Whilst such moderator analyses highlight the magnitude of treatment effect associated with certain patients, treatments and methodological factors, they do not infer causality (Cochrane Collaboration, 2011). Due to the broad inclusion criteria adopted for this review, there was a lot of variation in

study characteristics. Consequently, interpretation needs to be undertaken with caution, as some subgroup arms only had a small number of studies and the high correlation of some variables (e.g., TAU and inpatient settings) potentially confounds the observed effects.

Group BA was used in studies with a range of participants and varied clinical presentations, though the treatment effect when compared to controls or active therapies was not related to gender, initial depression severity or population. The finding that there was no association between the size of treatment effect and initial depression severity is in line with extant evidence (Driessen et al., 2010; Weitz et al., 2015), and contradicts original conclusions that psychotherapy effects are larger for less severe depression (Elkin et al., 1995). The current results imply that, regardless of baseline severity of depression, participants can experience benefit from group BA. Behavioural techniques are easily grasped and implemented by patients, even when (for example) cognitive functioning is impaired during depressed episodes (Lam, Kennedy, McIntyre, & Khullar, 2014).

Differences between age population subgroups were not significant, but two of the subgroup arms were very small for control and the active psychotherapy comparisons. Inspection of the size of the effects suggested some variation; group BA was very effective for young adults and adults (versus controls), but much less effective in older adults. It may be the case that BA in groups with older adult participants needs to have relevant treatment adaptations applied, in order to retain clinical effectiveness (Pasterfield et al., 2014).

Various treatment delivery factors (group size, different settings, type of BA or number of sessions) were not associated with differences in effectiveness, when compared against controls or active therapy comparisons. Again, statistical interpretation may have been hampered by confounding variables and insufficient

comparisons in the subgroup arms for setting and types of BA. Non-significant variation in effect sizes for different types of BA was evident - contextual versions seem to produce the largest treatment effects, but without being statistically superior. This finding mirrors the results produced by Mazzuchelli et al. (2009). The contextual BA protocol has additional components (e.g., rumination work) which might explain the larger treatment effects. However, the lack of a definitive advantage of one version of BA highlights that the behavioural treatment model will need further refining and testing to discern the optimal conditions for group delivery.

Number of sessions was not associated with the size of the treatment effect – a greater dose of therapy did not produce better outcomes. This supports the argument that group BA interventions only need to be brief. However, the range of sessions was relatively small, so any definitive statements concerning dose-response for group BA is premature at this stage. Control type did produce differences in treatment effects; waitlist comparisons resulted in a large effect, but TAU comparisons only had a small beneficial effect in favour of group BA. Similar effects have been seen for other types of psychotherapy (Cuijpers et al., 2013; Cuijpers, Van Straten, Bohlmeijer, Hollon, & Andersson, 2010), highlighting the importance of the type of comparator in determining a relevant estimate of effect.

2.4.3 Acceptability

The low drop-out rate for group BA found in this study implies BA delivered in a group can be well tolerated by patients. Treatment completion is fundamental to ensure the full benefit of treatment is received which is especially pertinent as early termination of psychotherapy is related to poorer outcomes (Cahill et al., 2003; Hansen et al., 2002). Any claims of the organizational efficiency benefits of group delivery are offset when group attendance is poor, with the dropout rates observed for group BA suggesting that group delivery does not suppress attendance. The equivalence of drop-

out rates for group BA in comparison to passive and active controls supports the notion that it is an acceptable treatment and matches the meta-analytic findings for individual BA (Ekers, Richards, & Gilbody, 2008).

2.4.4 Trial-based versus practice-based evidence

Critics of meta-analytic approaches argue the effects reported are often unattainable in routine practice due to being based solely on clinical trial evidence, and therefore suffer from limited generalizability (Rothwell, 2005). To alleviate this concern, this meta-analysis also estimated the effect of group BA achieved in naturalistic routine-practice settings. All comparisons with other active psychotherapies were within RCTs, but practice-based evidence was available for control or uncontrolled time-effect adjusted comparisons.

Delivery of group BA in clinical trials produced large treatment effects, but when group BA was provided in routine practice, treatment effects were more modest. Whilst trial-based and practice-based settings produced different effect size classifications, they were not significantly different. The origins of this difference might lie in different levels of study rigor, with clinical trials closely and routinely monitoring treatment delivery, patient selection and therapist supervision. The more complex and heterogeneous participant samples used in the PBE studies possibly contributed to the somewhat reduced effects observed. This study suggests that group BA has the potential to produce attainable real-world treatment effects that are not significantly less favourable than those produced in clinical trials.

2.4.5 Clinical and organisational implications

Access to clinically effective group interventions generates a range of organizational benefits, in relation to efficient use of facilities, high therapist to patient ratios and potential reductions to treatment wait-times (Piper, 2008). Recent evidence (Richards et al., 2016) also noted the health economic advantage of BA when delivered

on a one-to-one basis. Demand for psychotherapeutic treatment for depression is consistently high, and services can struggle to meet this demand whilst simultaneously ensuring high quality care (Kazdin & Blase, 2011). Frontline depression treatments in clinical services should balance the evidence of clinical effectiveness with issues relating to ease of access, acceptability and efficient use of scarce resources (i.e., balancing both effectiveness and reach). When evaluating a treatment, it is also recommended that it should be compared to the current gold-standard treatment (Spielmans, Pasek, & McFall, 2007). Compared to CBT, BA has an advantage of a potentially simpler, shorter training for therapists (or even non-specialists; Ekers et al., 2011). There were no differences in subgroup clinical outcomes or drop-out rates when group BA was compared to individual or group CBT (or CT variations) at post-treatment and follow-up. As originally highlighted by Jacobson et al. (1996), this meta-analysis echoes that therapy focused on changing depressogenic cognitions directly might be therapeutically redundant during the treatment of depression.

2.4.6 Limitations

There is a debate about what evidence is appropriate to include in a meta-analysis of a healthcare intervention, with RCTs often considered the only reliable data suitable for aggregation (Cochrane Collaboration, 2011). For the purpose of this meta-analysis, it was deemed appropriate to combine study designs to allow investigation of trial-based versus practice-based evidence, in order to increase the potential generalizability of results back to routine practice. It is acknowledged nevertheless that the inclusion of non-randomized and uncontrolled studies did introduce a risk of bias due to, for example, the role of patient selection as a confounding variable. RCT and observational evidence often produce similar estimates when aggregated (Shrier et al., 2007), and steps were taken to justify their combination. Effect sizes were calculated to ensure they were in the same metric regardless of study design (Morris & DeShon,

2002), correlations were used to account for non-independence of pre-post scores, and uncontrolled studies were adjusted using a time-effect estimate to account for spontaneous recovery. Empirical analysis of the influence of methodological design and study type found no statistical difference in effect estimates, suggesting they could be appropriately combined. Nevertheless, appropriate caution should be applied with overall interpretations.

There are a range of limitations to consider for this meta-analysis. The number of BA group studies was somewhat limited, with the majority of studies also having relatively small sample sizes (Turner, Bird, & Higgins, 2013). For primary outcomes, the number of comparisons was suboptimal for most subgroup analyses of post-treatment outcomes and as discussed above, the resulting moderator interpretations were somewhat restricted. Even fewer studies conducted follow-up depression assessments. The follow-up periods that were reported were generally short and so were too brief to provide a truly valid assessment of the durability of group BA. The measurement periods for follow-up assessments were typically between 4-12 weeks and this should be increased to at least one year in future group BA outcome research. As depression has a chronically relapsing nature, whether the effects of group BA compared to controls or active therapies can be retained in the long-term is still unclear (Steinert, Hofmann, Kruse, & Leichsenring, 2014). Longitudinal tracking of outcomes following group BA, relapse rates and any need for further intervention (e.g., behavioural 'top-up' sessions) would supplement the durability evidence base for group BA.

Recovery and drop-out data were not widely reported, meaning investigations of moderators and publication bias were not possible for those outcomes. Future group BA outcome studies should report core information on recovery and drop-out rates as standard and also report average session attendance. In terms of future controlled

research, a randomized patient preference trial (Howard & Thornicroft, 2006) of individual versus group BA would be a valuable addition to the evidence base.

The treatment effect reported for group BA in this meta-analysis may be subject to risk of some over-estimation and imprecision. First, study quality was varied and was in general largely sub-optimal. As lower quality studies were shown to produce higher estimates of treatment effect (in some instances), the degree of sub-optimal study quality may have contributed to an overstated overall treatment effect. Although the quality of group BA studies has improved over time, researchers should ensure the highest methodological quality wherever possible, according to context. Second, very few studies analysed outcomes using the intention-to-treat method and observed effects were mostly based on per protocol analyses. Such 'completer samples' are again at risk of over estimating treatment effects (Heritier, Gebski, & Keech, 2003).

Third, the distribution of comparator types across studies was not ideal. With regards to control comparisons, the majority were waitlist conditions. Waitlist controls are prone to overestimating treatment effects in active comparators (Cuijpers et al., 2013). The large difference in the group BA treatment effect compared to waitlists and TAU potentially reflects an overstated waitlist effect. It was also noted that the reporting of what TAU entailed was often vague, which makes generalizability of the effect of TAU and group BA similarly difficult to interpret. During the active therapy comparisons, the types of other psychotherapies were very varied, which might have diluted their effect in comparison to group BA. Only CBT or CT treatments were compared in enough studies to allow comparisons by treatment type. However, as CBT is the frontline treatment for depression, this allowed subgroup comparison of group BA with the current gold-standard (Spielmans et al., 2007).

Fourth, the broad inclusion criteria enabled data from a range of populations and settings to be analysed, but resulted in high levels of heterogeneity amongst studies, not

accounted for by the use of a random-effects model or moderator effects. The use of practice-based effectiveness studies undoubtedly contributed to increased heterogeneity, although significant variation was still evident across the BA clinical trials. Results give an indication of the effectiveness of group BA, but it is acknowledged that the variability increases the statistical imprecision of the effect estimate. Finally, fewer than half the included studies included a treatment integrity check. This means that group BA might not have been delivered in a protocol-adherent way.

2.4.7 Future research direction

This evidence shows that group BA is an effective treatment. However, there is no single version of BA. Although contextual versions may have the greatest benefit, the data are not firm enough to conclude this firmly. Direct comparisons in clinical trials of the different versions of group BA are needed to establish the most effective behavioural approach. BA is promoted for its simplicity – therefore, adding complexity or extending treatment without improving outcomes is counterintuitive and needs testing if it is to be justified. Hence, the focus going forward in the group BA evidence base should be on identifying the most clinically effective and organizationally efficient model for BA to be delivered in a group setting. This research could also embed longitudinal measures in the method, to allow analysis of what mediates the relationship between BA and outcome. Similarly, the evidence regarding older adults having a poor response to group BA indicates that moderators of group BA outcomes, such as age, need further investigation.

2.4.8 Conclusion

This review again provides support for BA as a standalone treatment for depression, but has shown that group delivery can be adopted. Group BA appears to be acceptable to patients and works for a broad population of participants, regardless of depression severity. The outcomes for group BA produced in controlled research trials

translate well into routine practice, albeit with a slightly smaller effect. Furthermore, group BA appears as clinically effective and acceptable as CBT, the frontline treatment for depression (National Institute for Health and Clinical Excellence [NICE], 2016). In light of the high and increasing demand for depression treatment in mental health services, BA should be considered a frontline intervention, on a par with CBT. Future research should focus on establishing the optimal delivery, mediators, moderators and long-terms effects of group BA, based on high quality efficacy and effectiveness studies.

CHAPTER 3

Depression 'Stasis' Following Evidence-Based Psychotherapy: A Review of Commonly Used Definitions

The previous chapter focused on the positive outcomes of group behavioural activation (BA), demonstrating its overall efficacy and effectiveness. However, such evidenced-based treatments are not always and unanimously effective. A considerable number of patients do not experience any benefit, instead remaining in a state of 'stasis' with regards to their depression symptoms. Processes of measuring change in psychotherapy have an emphasis on those who respond, often leaving these patients with minimal change overlooked. The objective of this chapter is to provide an overview of the current status of depression 'stasis' outcomes following psychotherapy treatments by 1) first, defining and discriminating depression stasis outcomes in the context of all suboptimal outcomes, 2) second, outlining the prevalence of stasis, and 3) third, examining potential factors associated with stasis.

3.1 Defining and discriminating 'stasis' outcomes

3.1.1 Distinguishing treatment failures

Lack of a positive outcome for some patients following psychotherapy for depression is a common occurrence, even for empirically evidenced treatments (Hansen et al., 2002). Treatment outcome research has tended to focus on treatment responders, with relatively limited research directed towards people who do not experience benefit. Part of the difficulty in understanding when and why therapy does not work, is because any outcome other than adequate treatment response tend to be grouped together as 'treatment failure.' An insufficient outcome can encompass multiple scenarios - premature treatment termination, failure to show any change in symptoms, a worsening

of depression symptoms, or even harm and inability to prevent relapse or recurrence of depression (see Table 3.1; Dimidjian & Hollon, 2011). The heterogeneous concept is reflected in the numerous terms used in the literature including 'treatment failure', 'suboptimal outcome', 'partial response', 'deterioration', 'harm', 'negative outcomes', 'unwanted events', 'extreme non-response', 'treatment refractory/resistant', and 'relapse' (Al-Harbi, 2012; Coffman, Martell, Dimidjian, Gallop, & Hollon, 2007; Hardy et al., 2017; Mauskopf et al., 2009; Parry, Crawford, & Duggan, 2016). Although all these terms all represent negative treatment effects, they are often measured in different ways and denote distinctive phenomena. The general grouping of all insufficient outcomes together makes identifying and quantifying separate phenomena challenging. Consequently, it is hard to draw conclusions about insufficient treatment outcomes in general, as there are likely to be different causes and interpretations (Lambert, 2011).

Table 3.1. *Negative outcome definitions associated with treatment failure*

Outcome/term	**Definition**
No change/nonresponse*	Symptoms that remain the same, with no change in either direction
Deterioration	An active worsening of symptoms following treatment
Harm	Sustained and significant worsening of symptoms as a direct result of treatment
Drop-out	Premature treatment termination
Relapse	A return of depression symptoms after showing an initial improvement after treatment
Recurrence	Another episode of depression after previously achieving recovery
Treatment-resistant depression	Lack of an adequate clinical response following two courses of appropriate treatment for depression (usually medication)

*Nonresponse term can often include deterioration outcomes

The broad nature of what can be deemed treatment failure has meant some areas have received more attention, leaving others unaddressed. The resulting research (in an already neglected area) has tended to focus on the more easily conceptualised aspects or attention-grabbing outcomes, with the likes of deterioration and harm being investigated since the 1960s (Bergin, 1963). The area which has arguably received the least attention is lack of response to treatment. The reason for this has probably been influenced by the unattractive nature of results, where nothing appears to have happened (i.e. the patient has neither improved, deteriorated nor being harmed) and termination outcomes are similar to assessment levels For therapists, there may be concerns about how minimal response reflects on their own competence (Lambert, 2011). For researchers, there are the pressures of wanting publishable results synonymous with evidence of significant effects. Depression symptom 'stasis' has subsequently been neglected by both therapists and researchers for years.

The definitions of minimal response used in the literature vary and are vague in terms of being able to quantify what is considered a lack of change (Linden, 2013; Mauskopf et al., 2009). Nonresponse to treatment is often used interchangeably with outcomes that include deterioration (Hiller, Schindler, & Lambert, 2012). However, latent class growth modelling has shown that nonresponse during psychotherapy has a distinct trajectory from other classes of outcome (Thibodeau et al., 2015). These outcomes should therefore be distinguished and further investigated in their own right. People who exhibit minimal change in their depression symptoms after treatment (neither improvement nor deterioration) experience what can be termed a 'stasis' (i.e., essentially unchanged) clinical outcome. It has been suggested that the historical lack of attention on this subset of patients has, in part, been determined by the processes used to measure change and define outcomes following psychological treatment. It is therefore

important to ensure that there is an adequate way to measure lack of change in order to capture the phenomenon of depression 'stasis'.

3.1.2 Processes for defining outcomes and measuring change

From the outcome measures to the paradigms for producing evidence-based practice, the process for measuring change places the emphasis on treatment responders. Standard outcome measures are developed to assess change in symptoms, rather than lack of change (Barlow, 2010). Change becomes the default focus of interest, priming people to notice change. Therapists often over-estimate how effective they are at delivering therapy, implying they are not very good at noticing when there has been a lack of change (Parker & Waller, 2015). Furthermore, evidence-based practice has emerged as the gold-standard paradigm for evaluating the efficacy of psychotherapy treatments (Barkham & Mellor-Clark, 2003). Randomised controlled trials (RCTs) are the foundation stone of subsequent evidence-based practice and involve the analysis of group mean outcomes for comparative conditions. However, by grouping patient outcomes together, the outcomes for individuals who do not experience any worthwhile gains are masked (Hiller et al., 2012).

In clinical practice, metrics for defining individual outcomes are more commonplace, enabling more scope for the breakdown of different types of outcomes. There is no clear consensus on which response metric to use for psychotherapy interventions with clinically significant change, percent improvement and reliable change index some of the most prominently employed techniques (Hiller et al., 2012). Each of these metrics have the single patient as the unit of analysis enabling percentages of types of clinical outcomes to be calculated. Although these individual outcome metrics do acknowledge treatment failure more than group statistics, the extent to which stasis outcomes are captured in their own right is not fully accomplished by any of

them. The ability of the current metrics to capture and define the phenomenon of stasis is outlined below.

3.1.2.1 Clinically significant change

Clinically significant change (CSC) was proposed by Jacobson, Follette and Revenstorf (1984) to assess treatment response by determining the ‘‘class’ a patient ends treatment in. CSC is defined as moving below a specified clinical cut-off deemed to class the range and breadth of symptoms as indistinguishable from the general population. It is undoubtedly important for patients to end treatment below a clinical range and is rightly classed as a successful outcome. However, when used in isolation, CSC has several drawbacks with regards to capturing change produced by the intervention. It does not recognise different forms of change that might represent a worthwhile gain (e.g., large reductions in scores for patients with high initial severity that do not reach the cut-off point), nor adequately distinguish between patient outcomes (e.g., outcomes including stasis, deterioration and large reductions short of the cut-off are grouped together as not clinically improved). There is also uncertainty about the extent of change evident that can be attributed to the intervention when initial severity is close to the cut-off (e.g., initial low severity patients only require marginal change in scores that does not exceed measurement error to be labelled a responder when they themselves may feel there has been little change in their symptoms). Little can be learned about stasis outcomes when they are grouped under definitions which leave them obscured by other types of outcomes.

3.1.2.2 Percent improvement

In the percent improvement (PI) method, treatment response (or partial response) is often defined as at least 50% improvement on a symptom rating scale. It is more commonly applied to pharmacological treatments (Hiller et al., 2012). In an improvement on the cut-off method, non-response is occasionally acknowledged

(although less commonly reported) and quantified (although not uniformly) as less than a 25-50% reduction (Mauskopf et al., 2009). By taking into account relative amounts of change in relation to initial depression severity (e.g., higher initial severity can be classified as a responder without reaching the normal range), responders are better distinguished from non-responders. However, at the lower end of the scale the PI method is more liberal. Therefore, 50% improvement for mild depression may not correspond to a change in excess of measurement error for the measure or represent a meaningful change for the patient (Hiller et al., 2012). Although limited change is recognised and to some extent quantified, there does not appear to be a consensus on the magnitude of change that can be construed as a lack of improvement. Cut-offs are arbitrary and not based on any psychometric properties (Hiller et al., 2012). Consequently, the non-response classification captures the concept of stasis to some extent, but the issue of emphasising responders still remains. Typically, responder percentages are reported, resulting in all other outcomes being grouped into the analogous non-responder's percentage. Within non-response, stasis outcomes are indistinguishable from deterioration.

3.1.2.3 Reliable change index and reliable and clinically significant improvement

Jacobson & Truax's (1991) reliable change index (RCI) is used to evaluate the degree of pre-post change at the patient level. It takes into account psychometric properties of symptom measures to calculate change thresholds for reliable change, exceeding what could be attributed to measurement error. Reliable change thresholds can then also be combined with clinically significant change (determined by the absence of clinical symptoms) to produce the reliable and clinically significant improvement (RCSI) metric. Reliable change can be in the form of *improvement* (reliable decrease in scores) or *deterioration* (reliable increase in scores), whereas *clinically significant change* is movement to below the clinical cut-off. When both criteria are applied in

conjunction, *reliable recovery* is represented by reliable improvement and movement into the normal range. Alternatively, *harm* is represented by reliable deterioration and movement into the clinical range (from the normal range). Reliable improvement or deterioration without clinical change provide mid-point response criteria, representing a continuum of sorts. Reliable recovery is the ultimate goal following therapy, but the benefit of RSCI is that it recognises and quantifies other forms of change, both beneficial and negative. Criteria for reliable change also enable marginal change to be distinguished from more relevant treatment responses (Lunnen & Ogles, 1998). However, there is still no explicit recognition of outcomes that remain unchanged, meaning the RSCI has been shown to be poor at distinguishing stasis from deterioration outcomes (Lunnen & Ogles, 1998).

3.1.3 Clarification of concept for stasis outcomes

As outlined above, the existing methods for defining change from individual outcomes fail to sufficiently capture and quantify the concept of lack of change. They have vague and undiscriminating definitions of stasis as an outcome in its own right, often including cases of deterioration. The reasons and subsequent implications behind deterioration following therapy are going to be different to those for stasis outcomes. Therefore. there needs to be a metric for explicitly capturing the concept of stasis. The RCSI method seems best placed to incorporate a distinguishable and quantifiable stasis metric and recent studies of routine practice outcomes have that started using the RCSI criteria with a nonresponse classification (Delgadillo, Moreea, & Lutz, 2016; Kellett et al., 2017; Pybis, Saxon, Hill, & Barkham, 2017). Following this example, for the purpose of the investigation in this thesis, an additional classification will be adopted to define *stasis* outcomes. Stasis will be recorded in cases where neither reliable improvement nor reliable deterioration occurs. *Stasis outcomes* can then be distinguished from *reliable/clinical improvement/recovery* (patients who reliably benefit

from treatment) and *reliable deterioration* (a reliable worsening of symptoms) and *harm* (a reliable worsening of symptoms coupled with movement from below to above the clinical cut-off).

3.2 Prevalence of depression stasis following psychological treatment

The literature available on depression nonresponse following treatment predominately concerns response to medication (Corey-Lisle, Nash, Stang, & Swindle, 2004; Knoth et al., 2010; Limosin et al., 2004; Rush et al., 2006). Investigation into stasis as a categorical outcome following psychotherapy pales into comparison compared to the studies of summary-level or positive treatment response studies. In theory, recovery rates from general outcome studies can be used to help give a sense of the scale of people not benefitting. However, due to the problems of distinguishing stasis outcomes from other unwanted responses, they can only provide estimates. On the other hand, the stasis outcome research that is available is sporadic in reference to different psychotherapies and how lack of response has been defined and measured. Table 3.2 presents an overview of stasis rates from clinical trial and practice-based psychotherapy depression outcomes that have used a defined classification for nonresponse.

The rates of stasis that are reported are quite varied, ranging from 14% to 57%. The variation could be due to use of a mixture of nonresponse classifications with different stringency, differences in patients populations (severity, presentation), differences in interventions (type, duration, frequency) and differences in the conditions of treatment delivery (clinical trial versus practice-based) (Dimidjian & Hollon, 2011). While it is difficult to draw conclusions about the effect of different patient populations and interventions, it is evident from Table 3.2 that some variation is caused by treatment conditions. Stasis is more prevalent in real-world settings, with rates up to 57% reported

in comparison to the maximum of 42% seen in clinical trials. Research trials are characterised by stringent internal validity. They use highly selected patient samples, all variables are controlled, and treatment quality and adherence is closely monitored (Rothwell, 2005). However, naturalistic settings of everyday practice do not operate under the same controlled conditions. Patient populations are far more varied and complex, treatment delivery does not always follow recommended guidelines and amount of treatment received is often suboptimal (Hansen et al., 2002). Nonetheless, even when treatments are delivered in optimum trial conditions, there is still variation in the rates of stasis. Other factors that might explain this variance may therefore come from the differences in patient populations or the interventions delivered.

Table 3.2. *Rates of depression stasis reported in clinical trial and practice-based studies*

Study	Nonresponse classification	Rate of stasis	Clinical presentation	Intervention(s)
Clinical trial outcomes				
Mohr et al. (1990)	Post-Tx scores within ±1 standard error measurement of week 1 scores	29%	MDD	Focused expressive psychotherapy, CT, supportive/self-directed therapy
Jacobson et al. (1996)	>8 on BDI and still met MDD diagnosis criteria	32-42%	MDD	BA, AT, CBT
Hansen et al. (2002)*	No clinically significant change	42%	Common mental health disorders	Psychological treatment
	No reliable change	33%		
Coffman et al. (2007)	Extreme nonresponse Post-Tx score >30 on BDI	22%	MDD	CT
Van et al. (2008)	<25% response on HAMD-17	39%	MDD	Short-term psychodynamic supportive therapy
Thibodeau et al. (2015)	Latent non-responder classification identified with growth mixture modelling	15%	Moderate or severe MDD	CBT, supportive therapy, psychodynamic therapy, antidepressants
Gollan et al. (2016)	No response or remission	42%	MDD	BA
Practice-based outcomes				
Hansen et al. (2002)*	NO RCI change in either direction on OQ-45	57%	Common mental health disorders	Psychological treatment

Lorentzen et al. (2011)	Change in psychosocial functioning <RCI	14%	Axis 1 affective and anxiety disorders	Long-term dynamic group psychotherapy
Schindler et al. (2013)	No reliable improvement on BDI	36%	MDD or dysthymic disorder	CBT
Delgadillo et al. (2016)	No reliable or clinically significant change on PHQ-9	54%	Depression in Primary Care	IAPT stepped care interventions
Kellett et al. (2017)	No reliable or clinically significant change on PHQ-9	52%	Depression in Primary Care	Group BA
Pybis et al. (2017)	No reliable or clinically significant change on PHQ-9	46%	Depression and/or anxiety in Primary Care	CBT, counselling in primary care
Schlagert & Hiller (2017)	Post-Tx BDI score not meeting criteria for response or deterioration	46%	MDD or dysthymic disorder	CBT

Note: *Hansen (2002) rates are a result of a review of treatment outcomes. Abbreviations; Post-Tx; post-treatment, BA; behavioural activation, AT; automatic thoughts, CBT; cognitive behaviour therapy, CT; cognitive therapy, MDD; major depressive disorder, BDI; Beck Depression Inventory, PHQ-9; Patient Health Questionnaire-9, OQ-45; Outcome Questionnaire

3.3 Factors associated with depression stasis

The amount of literature on factors associated with stasis does not reflect the proportion of stasis outcomes that occur in practice. Unsurprisingly, as the majority of the treatment outcome evidence base focuses on positive effects of therapy, the understanding about factors associated with outcomes is also largely focused on predicting symptom improvement (as a continuous variable). Using a continuous scale of therapy improvement makes it difficult to discern factors that predict no change, as opposed to less change than those who respond well. Far fewer studies have looked at predicting stasis as a binary outcome (i.e., responder versus non-responder). The following sections provide an overview of the evidence for proposed associations between stasis outcomes and patient characteristics (sociodemographic, clinical), treatment variables (type, format, duration) and process factors (attendance, engagement, therapist factors).

3.3.1.1 Patient factors

3.3.1.1.1 Sociodemographic characteristics

Sociodemographic factors represent an attractive and easily applied method for predicting treatment outcomes, as they are commonly identified in the process of patient information gathering. The prognostic abilities of patient variables including age, gender, disability, marital status, ethnicity, education, employment and socioeconomic status (SES) have mostly been investigated as predictors of treatment response. In general, findings have been inconclusive in terms of predicting post-treatment stasis, with inconsistent or minimal associations for most sociodemographic factors (Bohart & Wade, 2013; Reuter et al., 2016; Vittengl et al., 2016). The few factors that have been implicated have differed across studies. Greater rates of stasis has been linked to older age for psychodynamic supportive therapy (Van et al., 2008) and younger age for cognitive behavioural interventions (Delgadillo, Moreea, et al., 2016). Higher levels of stasis after cognitive-behavioural therapy (CBT) has also been linked to unemployment, lower SES and disability (Delgadillo, Moreea, et al., 2016; Falconnier, 2009; Thase, Simons, & Reynolds, 1993). Meanwhile, a review of response to cognitive therapy (CT) determined the only sociodemographic factor associated with poor response was marital status (Hamilton & Dobson, 2002).

3.3.1.1.2 Clinical characteristics

Clinical factors relating to symptom severity, functional impairment, chronicity, age of onset, prior episodes and co-morbidity are also conceivable outcome predictors for poor treatment response. Although clinical factors appear to have a stronger association with treatment response than sociodemographic characteristics, the relationship with risk of stasis is still relatively under-investigated. Initial depression severity has produced inconsistent findings, with studies finding both higher levels of severity (Hamilton & Dobson, 2002; Reuter et al., 2016; Thibodeau et al., 2015; Vittengl et al., 2016) and milder levels of depression related to nonresponse (Lorentzen et al., 2011; Van et al., 2008). Different methods for measuring nonresponse might

contribute to contrasting findings, as outlined previously different methods are more liberal at different ends of the severity scale. Factors relating to development of depression such as chronic depression (Hamilton & Dobson, 2002; Thase et al., 1993), multiple previous episodes (Lorenzo-Luaces, Derubeis, & Webb, 2014) and a younger age of onset are thought to increase risk of stasis after CT/CBT (Hamilton & Dobson, 2002). Furthermore, previous non-response to treatment has been shown to increase risk of subsequent treatment nonresponse across a variety of inpatient treatment modalities (Reuter et al., 2016). Level of functional impairment at the start of treatment has been shown to be associated with a lack of post-treatment change after CBT (Coffman et al., 2007; Delgadillo, Moreea, et al., 2016) and interpersonal therapy (Frank et al., 2011). Finally, comorbidity of depression with other mental health disorders impacts on treatment response, with evidence for comorbid personality disorders (Reuter et al., 2016) and comorbid anxiety disorders resulting in greater nonresponse (Gelhart & King, 2001) after depression treatment.

3.3.1.2 Treatment factors

Treatment factors refer to treatment types or components, format (one-to-one, group, online, self-help), or the duration and frequency of sessions. Very few studies have looked at how these factors affect nonresponse directly. Outcomes for the average patient have been widely compared across treatment models, with all interventions generally being as effective as each other (Cuijpers, 2017). However, individual outcomes highlight outcomes vary greatly between patients, so it remains to be seen if different treatment models differ in terms of stasis rates. Again, effect of modality of treatment on stasis outcomes has received little attention due to nonresponse rarely being reported as an outcome. General treatment outcome research tends to suggest treatment modality does not have an overly clinically relevant effect on outcomes. Some studies have found one-to-one treatment produce slightly bigger treatment effects

(Cuijpers & Straten, 2008), whereas other have shown modality of treatment has little effect on the amount of change produced (Saxon, Firth, & Barkham, 2017). Duration of treatment has a well-known relationship to outcome (Hansen et al., 2002). Initially response increases as treatment dose increases, until response plateaus and additional sessions do not add any significant benefit. 'Dose-response' analyses attempt to pinpoint the optimal dose that enables between 50-95% of patients to respond. One of the contributing factors to the higher rates of stasis seen in routine practice compared to clinical trials is thought to be due to insufficient doses of therapy being delivered in the real world. Furthermore, direct comparisons of responders versus non-responders show treatment non-responders have shorter treatment durations (Lorentzen et al., 2011), suggesting they fall short of receiving the optimal dose.

3.3.1.3 Treatment process factors

Treatment process factors refer to variables relating to how therapy is implemented and produces change. Several key factors relating to the treatment process have been associated with outcome in the literature and represent feasible predictors of stasis outcomes (Dimidjian & Hollon, 2011). First, the more time patients have to spend on a waiting list to start treatment, the less likely they are to have a good outcome (Clark et al., 2017). Second, attendance at treatment has well known link to outcome (related to dose-response in the previous section). Drop-out reduces the effectiveness of treatments for depression and has been shown to be associated with poor treatment outcomes and nonresponse across a variety of therapies (Barrett, Chua, Crits-Christoph, Gibbons, & Thompson, 2008; Cahill et al., 2003; Cooper & Conklin, 2015; Hans & Hiller, 2013; Swift, Greenberg, Whipple, & Kominiak, 2012; Van et al., 2008). Third, failure to engage with and complete homework tasks is related to less benefit (Addis & Jacobson, 2000). Therefore, being unable or unwilling to engage with treatment components is likely to increase the likelihood of a stasis outcome, especially for

therapies that involve between-session work. Fourth, early symptom change has been implicated as one of the most reliable predictors of depression treatment outcome, regardless of therapy type or duration (Delgadillo et al., 2014; Fowler et al., 2015; Lutz, Stulz, & Köck, 2009; Schlagert & Hiller, 2017; Tadić et al., 2010). It follows that patients who do not show early change have a heightened chance of finishing treatment with a stasis outcome. Finally, therapist factors have been shown to be influential in terms of post-treatment outcomes and could be potential predictors of stasis. Poor therapeutic alliance (the clinician-patient bond) and reduced therapeutic adherence (delivery of interventions as prescribed in the protocol) have been shown to predict nonresponse after CBT (Vittengl et al., 2016; Weck, Grikscheit, Jakob, Höfling, & Stangier, 2015). Interestingly, comparisons of treatment responders versus non-responders have suggested the level of therapist competence (the skilfulness with which the interventions are implemented by the therapist) is less significant.

3.4 Conclusion

In summary, there is a dearth of understanding about what predicts a depression stasis outcome. Treatment failure outcomes are often all grouped together making it difficult to differentiate stasis from other outcomes. Stasis needs to be defined as an outcome in its own right and an appropriate metric needs to be applied to capture nonresponse in the same way rates of response are recorded. A review of the current outcome metrics identified the use of a stasis classification (lack of reliable change in either direction) within the reliable and clinically significant improvement criteria would be best placed to record the variation in individual outcomes after psychotherapy for depression.

Stasis outcomes appear more prevalent in routine practice suggesting there are contributing factors that are specific to treatment delivery in real world services.

Potential sources of variation in individual outcomes could come from patient, treatment or process factors. Although patient factors represent an attractive method for predicting stasis outcome, findings have been inconclusive with regards to which factors are useful in identifying those at risk of stasis. Whether the inconclusive outcomes are attributable to the paucity of evidence, due to heterogeneous predictors of stasis for different therapies or indeed reflective of no association is still to be determined with further stasis investigation. The effect of treatment factors on stasis outcome is not clear. There are not enough studies or direct comparisons to get a clear picture of how separate interventions or treatment formats affect stasis rates within routine practice. There may be differential predictive effects for different treatment models. Several promising treatment process factors have been identified; however they largely are specific to CBT interventions. Expanding stasis investigation to other interventions could increase the clinical relevance for specific treatments. The utility of patient and treatment factors that predict stasis would pave the way for methods to match patients to the appropriate treatment for their presentation. On the other hand, treatment process factors that predict stasis could be used to develop feedback systems and targeted strategies that can improve treatments, and hopefully increase the benefit patients can experience. Evaluations of outcomes after large-scale routine practice treatment delivery are needed to establish rates of stasis for specific treatments (delivered in different formats and doses) and explore associated predictors of stasis.

CHAPTER 4

Effect of Intervention Intensity, Format and Dose on Treatment Response following Behavioural Activation for Depression in Routine Practice

The previous chapter outlined the issue of stasis after depression treatment and the inconsistency in the identification of factors that affect stasis outcomes. It was suggested that understanding treatment response and stasis predictors for separate treatments may be more clinically relevant. A metric incorporated into the reliable and clinically significant change method was proposed to aid further stasis investigation in this thesis. Therefore, the objective of this second empirical chapter was to investigate treatment response after different intensities and formats of BA therapy. First, summary and patient-level effects of BA modes are evaluated to establish rates of treatment response (including the stasis metric). Second, the dose-response effect for symptom improvement across BA modes is calculated and plotted for 'stasis' versus 'improver' patients. Third, predictors of 'stasis' versus 'improvement' after BA interventions are explored. Finally, characteristics of 'stasis' versus 'improver' patients are compared.

4.1 Introduction

4.1.1 Establishing treatment response to behavioural activation

There is a growing evidence base supporting behavioural activation (BA) as an effective psychological treatment for depression, when delivered one-to-one (Ekers et al., 2014; Richards et al., 2016) and in groups (see Chapter 2; Simmonds-Buckley, Kellett, & Waller, 2019). In line with the concept of evidence-based practice, BA treatment efficacy has been demonstrated through the use of clinical trial paradigms (Barkham, Hardy, & Mellor-Clark, 2010; Reynolds, 2000). Highly selected patient

samples are randomised to treatments delivered in optimal conditions and summary post-treatment mean outcomes are compared to determine treatment effects. However, as outlined in Chapter 3, while these methods allow confident interpretations of the efficacy of treatment at the level of the clinical population, they often mask what happens at the patient level and when the intervention in then delivered in routine care settings (Hiller et al., 2012).

Recovery rates highlight that treatments for depression are not panaceas and therefore not beneficial for everyone. Estimates of reliable symptom improvement (reductions that exceed measurement error) range from 64% in trial contexts to 35% in naturalistic settings (Hansen et al., 2002). It is therefore apparent that a significant proportion of patients may have an poor therapy outcome. While a small proportion (~10%) might reliably deteriorate, the majority fail to experience any meaningful change in their depression symptoms. Remaining in this state of symptomatic 'stasis' can leave patients demoralised and at risk of failing to seek future help (Meltzer et al., 2003; Ten Have et al., 2010). Understanding what is associated with having a stasis outcome following BA could identify ways to change the manner in which BA is delivered or highlight 'at risk' groups. However, as treatment outcome research has predominantly focused on summary outcomes and treatment responders, the frequency and predictors of stasis outcomes after BA is unclear. Hence, alternative methods are needed to explore BA treatment response further.

4.1.2 The Improving Access to Psychological Treatment (IAPT) programme

Given that observed rates of deterioration and stasis are greatest within routine care settings, practice-based approaches seem most suited to gain an understanding on non-responsivity to psychological interventions and also have greater external validity. The Improving Access to Psychological Therapies (IAPT) model, set up in England in 2008 (refer to Chapter 1 – section 1.2.1), offers a practice-based paradigm for

investigating variability in treatment outcomes (Layard & Clark, 2014). Nationwide IAPT services deliver evidence-based talking therapies for common mental health problems informed by the National Institute for Health and Care Excellence (NICE) guidelines (Clark, 2011). Routine outcome monitoring (ROM) of session-by-session outcomes is utilised as standard to allow a focus on patient-level recovery rates. Services are tasked with achieving an overall 50% rate of recovery to match the outcome rates achieved in the trials that form the evidence base for the NICE guidelines for depression and anxiety (IAPT, 2011).

Treatments are delivered in a stepped-care service delivery system on the premise that for many mild to moderate clinical presentations, a brief and less intense (i.e., and also cheaper), but nevertheless effective treatment is sufficient for recovery (Firth, Barkham, & Kellett, 2015). Low intensity (LI) treatments are guided self-help interventions (e.g., LI BA, cognitive restructuring, exposure, sleep hygiene, worry management and panic management) delivered by Psychological Wellbeing Practitioners (PWPs) at Step 2 of IAPT services. PWPs are trained according to a national curriculum to adopt a role as a 'coach', supporting patients to engage with LI interventions and signposting to additional specific support that may be required from external sources. If patients do not respond to the low intensity treatment, or they initially present with more severe or complex symptoms, they are offered a longer, more intensive psychological treatment (Bower & Gilbody, 2005). High intensity (HI) treatments are traditional psychotherapies delivered at Step 3 of IAPT services (e.g., HI BA, cognitive behavioural therapy [CBT], person centred counselling, couple counselling for depression, counselling for depression, psychodynamic interpersonal therapy, dynamic interpersonal therapy) by trained and accredited therapists. Interventions are delivered in a variety of formats, including over the phone and internet (LI versions), one-to-one and in groups (both LI and HI). Critics of IAPT services are

concerned that lack of independent evaluation has resulted in inflated recovery rates, arguing far fewer patients actually recover (Scott, 2018). Critiques have questioned whether IAPT is able to deliver what it promises in terms of treatment accessibility and effectiveness, calling for independent studies to assess the suitability of IAPT services (Marks, 2018).

Routine IAPT data therefore contain a wealth of information on outcomes from BA interventions based on the same theoretical rationale, but delivered via different modes and in varying durations (i.e., number of sessions). BA is promoted as a parsimonious treatment that offers a valuable treatment option for clinical services (Kanter & Puspitasari, 2016). In light of equivalent outcomes with CBT, the frontline treatment for depression (David et al., 2018), it is argued the benefit of BA comes from the simple and flexible implementation of treatment (simple or complex versions/one-to-one or in groups/by novices or experts). The IAPT context therefore provides a unique opportunity to explore how factors relating to the intensity, format and dose of BA interventions impact treatment response at both the summary and patient level.

4.1.3 Exploring variability in BA treatment response

Given the variability in treatment response, there is a need to identify what predicts risk of having a stasis outcome after differential modes of BA. As outlined in Chapter 3 (section 3.3), the majority of psychotherapy research has focused on predictors of treatment response, grouping all unwanted outcomes together (stasis, deterioration, relapse). The current understanding about lack of treatment response is therefore largely derived from these estimates, rather than specific investigation. There is variable evidence that patient characteristics predict poor psychotherapy treatment outcomes in general. Demographic factors (such as age, gender, ethnicity) appear to have an inconsistent association with negative outcomes (see Chapter 3, section 3.3.1.1.1 for overview; Bohart & Wade, 2013; Reuter et al., 2016; Vittengl et al., 2016),

whereas pre-treatment clinical factors, such as baseline symptom severity and functioning have stronger evidence for predicting improvement (see Chapter 3, section 3.3.1.1.2 for overview; Delgadillo, Moreea, et al., 2016; Hamilton & Dobson, 2002; Thibodeau et al., 2015). However, research into clinical predictors of IAPT populations is still in the early stages (Hepgul et al., 2016). Analyses of large treatment outcome datasets will facilitate the identification of predictors of treatment response. To date, no studies have investigated predictors of stasis across stepped-care delivery of BA. Identifying pre-treatment patient characteristics (demographics and clinical markers) that distinguish improver versus stasis patients after BA modes may be useful for selecting treatment paths that have the greatest probability of producing a positive outcome.

However, in terms of developing interventions or adaptations to the BA model to reduce stasis outcomes, establishing BA treatment-related associations with outcome may be more practical (e.g., a patient's demographics cannot be changed). First, how does the intensity of BA interventions affect treatment response? The structure of the stepped care IAPT model means BA is delivered via both *simple* LI and *complex* HI protocols. The LI version is based on the brief behavioural activation for depression treatment manual (BATD; Lejuez et al., 2001). The protocol comprises the basic activation principles of BA (i.e., widening activity levels to increase response-contingent positive reinforcement and hence improve mood). Meanwhile, the HI version is based on Addis and Martell's BA method (2004). In addition to the basic activation principles of BA, HI treatment components initially analyse the function of the patient's behaviours to then target activation and positive reinforcement to the most valued areas of the patient's life. Complexity is added to the BA protocol to help patients gain a deeper insight into their problem and how it could be resolved, hence providing an additional therapeutic benefit. It would therefore be expected that more

complexity may result in improved outcomes. However, as BA's strength is in the simplicity of the model, adding complexity may be counterproductive if benefits are not reflected in the treatment response. Furthermore, differential frequencies of stasis outcomes after LI and HI versions would improve our understanding of the crucial treatment elements in BA.

Second, how does the format of BA delivery affect treatment response? The literature demonstrates that one-to-one and group treatments can each be effective at treating depression (McDermut, Miller, & Brown, 2002), with some evidence for potentially larger effects for individual therapies (Cuijpers & Straten, 2008). There also does tend to be a preference towards one-to-one methods within clinicians and patients (Brown et al., 2011). However, group therapies offer potential additional therapeutic benefits (normalising problems, peer support, learning opportunities; Yalom & Leszcz, 2005) and organisational benefits (efficient therapist to patient ratios; Kellett, Clarke, & Matthews, 2007). The simplicity of the BA method suggests that BA translates well in a group context (Simmonds-Buckley et al., 2019), yet there are limited extant direct comparisons of one-to-one and group BA delivery. Comparability of treatment response after each BA format would strengthen the evidence base for group BA and could encourage it being offered more widely as a viable treatment option.

4.1.4 BA dose-response across modes of delivery

A final treatment-related factor that needs to be taken into consideration is amount of treatment delivered. An association between the 'dose' of treatment (i.e., number of sessions received) and positive treatment response has been broadly supported in psychotherapy literature, termed the *dose-response effect* (Hansen & Lambert, 2003; Hansen et al., 2002; Howard, Kopta, Krause, & Orlinsky, 1986). Optimal dose estimates for at least 50% of patients to experience recovery in routine care range between 4-26 sessions, with a diminishing probability of improvement in

subsequent sessions (Robinson, Delgadillo, & Kellett, 2019). Dose-response effects are often a reflection of widely heterogenous clinical populations and interventions, meaning the wide variation in estimates is perhaps not surprising. Studies based on particular interventions with specific clinical presentations have produced more specific estimates (Robinson et al., 2019). It is important for clinicians to have predictions specific to the intervention they are delivering, so that they are able to effectively utilise ROM and identify when a patient is not on track for a positive treatment response.

With that in mind, the optimal dose specific for BA interventions within a stepped-care model has yet to be established. The dose-response for all anxiety and depression LI interventions suggest 4-7 sessions are sufficient for a 50% probability of response, whereas HI interventions require 5-14 sessions (Delgadillo et al., 2014; Robinson et al., 2019). These findings suggest LI interventions produce faster improvements in comparison to HI interventions, but it remains to be seen whether the same pattern is observed within a more pragmatic and parsimonious BA intervention or whether improvement patterns differ according to treatment format. Comparisons of dose-response in HI and LI treatments are confined to the boundaries of NICE guidelines and the extant treatment protocols. The maximum number of sessions offered sets a natural upper limit (LI = 6-8 sessions, HI = 14-16 sessions). However, group BA protocols are delivered in less sessions (8 sessions), which will enable a comparison of a HI therapy with a natural upper limit comparable to the LI version. It would be useful to establish the dose required to identify treatment improvers across different modes of BA within the confinements of routine delivery. This information could then be used to pinpoint where treatment improvers can be distinguished from those that are at risk of a stasis outcome, and would provide clinicians with valuable information to supplement ROM data during treatment.

4.1.5 Study aims and hypotheses

In summary, BA is emerging as an effective and valuable treatment option for depression and is commonly delivered as a treatment for depression in routinely delivered public service settings. In general, treatment response research has masked outcomes that leave individual patients in a state of depression stasis outcome, and has meant that there is lack of understanding about the frequency and predictors of these patient-level outcomes. Within IAPT services, BA is delivered as either a LI intervention (simple, briefer protocol) and a HI intervention (longer, more complex protocol). It can also be delivered in a one-to-one format or as part of a group (Kellett et al., 2017; Simmonds-Buckley et al., 2019). IAPT services therefore provide a useful practice-based setting to explore how patient and treatment factors affect treatment response at the summary and patient level. Frequencies and predictors of stasis outcomes could help identify ways to adapt routine care to increase the number of patients who get a benefit from treatment. Meanwhile, comparison of dose-response effects specific to BA interventions in a stepped-care system has not previously been established. A better understanding of both these concepts could provide researchers and clinicians with useful information to identify patients at risk of stasis outcomes and aid with outcome monitoring via feedback systems.

Using IAPT outcomes from BA delivered via different modes, the present study aimed to (1) compare the summary and patient level effects of low and high intensity versions of the BA model, (2) establish the number of sessions required to enable a treatment response across BA models, (3) explore predictors of end of treatment depression stasis outcomes, and (4) compare characteristics of stasis and improver BA patients. It was hypothesised that;

- Complex versions of BA (i.e., HI BA) will produce better end of treatment summary-level and patient-level outcomes compared to simple BA (i.e., LI BA)

- However, LI BA will require fewer sessions to identify treatment improvers (i.e., produce faster improvements in depression symptoms)
- Clinical and treatment characteristics will be stronger predictors of end of BA treatment stasis outcomes than patient characteristics (e.g., demographics)
- Stasis patients will be differentiated from improvers on clinical and treatment characteristics

4.2 Method

4.2.1 Design

Secondary analysis was conducted on session-by-session single service IAPT routine outcome data collected between April 2009 and April 2018. Longitudinal treatment outcomes were compared for three modes of BA therapy delivered within a stepped-care model (LI one-to-one BA, HI one-to-one BA and HI group BA; see Figure 1.1 in Chapter one, section 1.2.1 for description of IAPT stepped-care model). The study received ethical and research governance approval from the Leeds East NHS Research Ethics Committee (*IRAS project ID: 202197, REC reference: 16/YH/0324*). Information and evidence about ethical approval can be found in Appendix C (protocol, ethical approval confirmation, minor amendment documentation and approval).

4.2.2 Study sample

The study sample was obtained from an anonymised dataset of existing routine demographic, service usage and outcome data collected in a UK IAPT service. Following the stepped-care model, patients requiring psychological therapy are referred to IAPT services by general practitioners (GPs) or self-refer. Patients are assigned to an appropriate treatment following an initial assessment of the type and severity of their problems by a Psychological Wellbeing Practitioner (PWP). BA is one of the interventions offered as a treatment option for patients with depression. Patients with

mild to moderate symptoms are typically referred to treatment at step two and offered a low-intensity guided self-help intervention with a PWP (e.g., one-to-one low intensity BA). If patients have a limited response to the step two intervention, they are stepped up to step three and provided with a 'stronger dose' of high intensity therapy with a qualified and accredited therapist (e.g. high intensity BA). Alternatively, patients with problems that are determined to be moderate to severe or more complex at the initial assessment are stepped straight to treatment at step three.

The IAPT service used in the present study delivered BA therapy in three different modes. A low-intensity guided self-help BA intervention delivered one-to-one was available at step two. On the other hand, patients that after assessment were deemed appropriate for BA treatment at step three were offered the option of one-to-one or group BA therapy. The sample of patients who received HI group BA were self-selecting as they had to opt-in or could directly self-refer to attend a group. Otherwise, patients were placed on a waiting list for one-to-one HI BA treatment when a HI therapist became available. Decisions to opt-in to a group may have been in part influenced by wait times as group BA typically enabled patients to access treatment quicker. Group BA involved a shorter waiting time as a new group was delivered every 8 weeks, whereas wait times for one-to-one high-intensity therapy could be more variable. The accessed dataset consisted of data for *N*=1968 patients who had accessed the IAPT service within the dataset timeframe (LI one-to-one BA & HI one-to-one BA; data collected between 2014-2018; HI group BA; data collected between 2009-2017) and been referred to one of these modes of behavioural activation therapy.

4.2.2.1 Eligibility criteria

Inclusion criteria were: (a) seeking treatment for a primary presenting problem of depression, b) received a BA intervention (LI one-to-one BA, HI one-to-one BA or HI group BA), c) have data available from at least two sets of routine measures (i.e.,

attended two or more sessions to be able to calculate a change score), and d) meet criteria of caseness for clinical depression (Patient Health Questionnaire; PHQ-9 score ≥10). As it was a study of routine practice, exclusion criteria were minimal to ensure a representative sample. Exclusion criterion applied were a) a PHQ-9 score <10 prior to commencing treatment (to ensure that calculation of stasis rates were not unduly influenced by a floor effect), and b) only having one available set of routine outcomes, as session one measures reflect symptom-severity during the two-week period prior to the treatment sessions, therefore it cannot be known if session one alone leads to symptom changes. The final study sample consisted of 843 patients (LI one-to-one BA: *N*=609; HI one-to-one BA: *N*=65, HI group BA: *N*=169; see Figure 4.1).

4.2.2.2 Sample size estimate

A sample-size analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) indicated detection of a small interaction effect size (*f*=0.15) between three conditions, providing data across two time points, with .80 power at a significance level of p =.05 would require a total sample of *N*=37 patients per mode of BA (= total sample size of 111). Use of logistic regression to detect a small effect size (odd ratio = 1.4), with .80 power at a significance level of p =.05 would require a total sample of *N*=344 for continuous predictors and *N*=1309 for categorical predictors. Therefore, although the total sample of *N*=843 was sufficiently powered for the majority of the main analyses, detection of categorical stasis predictors was underpowered.

4.2.3 BA treatment delivery modes

Patients received BA treatment based on NICE guidelines for depression (NICE, 2016) delivered via three different modes; LI one-to-one BA, HI one-to-one BA and HI group BA. The fundamental rationale of all three treatments were based on the core principle of BA - to encourage increased participation in rewarding activities (Martell et al., 2001). Patients are encouraged to recognise the link between activity and mood and

then learn to schedule activities to increase their activation. To implement the BA strategies, all modes of treatment involved setting patients between-session activities (i.e. 'homework') that were fed back and reviewed at each treatment session.

The differences in the mode of intervention were centrered around the *'intensity'* and *format* of the treatment, summarised in Table 4.1.Following the CBT competency framework (Roth & Pilling, 2007; University College London, 2018), the LI mode was a guided self-help version of BA based on elements of Lejuez, Hopko and Hopko's brief behavioural activation for depression treatment manual (BATD; 2001). It comprised the basic activation principles of BA with treatment components grounded in information gathering to develop an ABC model (antecedents, behaviours, consequences), self-monitoring, development of activity hierarchies and BA diaries and use of rewards for achieving goals (University College London, 2018). LI BA was delivered one-to-one with a PWP (by a total of 58 PWPs). The PWPs' role involved supporting patients set goals for treatment and work through the self-help materials. According to the CBT competency frameworks for HI interventions (Roth & Pilling, 2007; University College London, 2018), the HI versions were full therapeutic courses of BA based on Martell et al.'s (2001) version of BA. HI BA comprised the basic activation principles of BA with additional components incorporated into the protocol designed to align activation with patient values, and therefore to have a more personally meaningful function (including functional analysis, values work, TRAP/TRAC formulation and dealing with rumination strategies). Martell et al.'s (2010) clinicians guide was used to inform the development of the HI BA protocol into an eight-session treatment manual suitable for group delivery. HI BA was delivered either one-to-one with a CBT therapist (total of 13 therapists) or in a group facilitated by two CBT therapists (total pool of 17 facilitators). HI group BA was delivered across 25 groups with an average group size of eight patients.

Table 4.1. *Summary of BA modes of delivery*

	Mode of BA		
	LI one-to-one BA	**HI one-to-one BA**	**HI group BA**
Stepped care level	Step 2	Step 3	Step 3
Format	One-to-one	One-to-one	Group
Depression severity	Mild-moderate	Moderate-severe / complex	Moderate-severe / complex
No. of sessions offered	6	10-14	8
Duration of sessions	35 mins	50-60 mins	120 mins
Frequency	Typically fortnightly	Weekly	Weekly
Delivered by	x1 PWP	x1 CBT therapist	x2 CBT therapists
Training qualification	Postgraduate certificate in LI interventions	Postgraduate diploma in HI interventions	Postgraduate diploma in HI interventions
Accredited	Not required	BABCP	BABCP
Supervision	1 hr case management (weekly) 1 hr clinical; individual or group (fortnightly)	1 hr clinical; individual (weekly)	Issues taken to individual supervision & 1 hr BAG working groups (quarterly)
Setting	GP surgery/ community venue / telephone	GP surgery / community venue	Central IAPT base / community venue
Protocol	BA self-help materials Based on Lejuez et al. (2001)[a]	BA treatment protocol Based on Martell et al. (2001)[a]	BAG treatment manual & patient workbook Based on Martell et al. (2010)
BA components	*Core principles* Activity-mood monitoring Activity identification and scheduling	*Core principles;* Activity-mood monitoring, activity identification and scheduling *Additional components;* Values, TRAP/TRAC, problem solving, developing responses to rumination	*Core principles;* Activity-mood monitoring, activity identification and scheduling *Additional components;* Values, TRAP/TRAC, problem solving, developing responses to rumination

[a] Sources as indicated by the UCL competency frameworks for LI and HI interventions (available from https://www.ucl.ac.uk/clinical-psychology/competency-maps/cbt-map.html). Abbreviations; BA: behavioural activation; LI: low intensity; HI: high intensity; PWP; psychological wellbeing practitioner; CBT; cognitive behavioural therapy; BABCP; British Association of Behavioural and Cognitive Psychotherapy; TRAP/TRAC; trigger, response, avoidance pattern/trigger, response, alternative coping.

4.2.4 Outcome measures

The outcome measures consisted of the IAPT minimum dataset (PHQ-9, GAD-7 and WSAS – see below). Copies of the outcome measures can be found in Appendix D. These self-report measures are completed at every contact as part of IAPT routine outcome monitoring.

4.2.4.1 Patient Health Questionnaire-9 (PHQ-9)

The PHQ-9 is a nine item self-report questionnaire scored from 0-27 (Kroenke, Spitzer, & Williams, 2001). It is designed to detect depression within primary care settings and is administered as part of routine outcome monitoring in IAPT services. The PHQ-9 was used as the primary outcome measure for evaluating change in depression symptoms. It has sensitivity and specificity scores of 92% and 80% respectively at the ≥10 clinical cut off point (Gilbody, Richards, Brealey, & Hewitt, 2007) and excellent internal consistency (Cronbach α= .89) and test-retest reliability (correlation = .84; Kroenke et al., 2001). Patients' scores are classified into depression severity according the following thresholds; 0-4 = no depression, 5-9 = mild, 10-14 = moderate, 15-19 = moderately severe and 20-27 = severe. A score of 10 and above is classed as indicating clinically significant symptoms of depression.

4.2.4.2 Generalised Anxiety Disorder-7 (GAD-7)

The GAD-7 is a seven item self-report questionnaire scored from 0-21 (Spitzer, Kroenke, Williams, & Löwe, 2006). It is designed to detect anxiety within primary care settings and is administered as part of routine outcome monitoring in IAPT services. The GAD-7 was used as a secondary outcome measure for evaluating anxiety symptoms. It has sensitivity and specificity scores of 92% and 76% respectively at the ≥8 clinical cut off point and excellent internal consistency (Cronbach α= .92) and test-retest reliability (correlation = .83; Spitzer et al., 2006). Patients' scores are classified into anxiety severity according the following thresholds; 0-4 = no anxiety, 5-10 = mild,

11-15 = moderate and 15-21 = severe. A score of 8 and above is classed as indicating clinically significant symptoms of anxiety.

4.2.4.3 Work and Social Adjustment Scale (WSAS)

The WSAS is a five item self-report measure of functional impairment as a result of mental health problems (Mundt, Marks, Shear, & Greist, 2002). It is designed to assess the impact the patient's symptoms are having on their work, home life, leisure activities and social relationships. It has excellent internal consistency in depression samples (Cronbach α= .81 to .92) and good test-retest reliability (correlation = .73; Mundt et al., 2002). Scores range from 0-40 and are classified according to the level of functional impairment; 0-9 = sub-clinical populations, 10-20 = significant and 20+ = moderately severe to severe.

Information on patient demographics, service usage, clinical and treatment characteristics were accessed from the dataset to provide pre-treatment markers and predictor variables. Patient demographics included age, gender, employment status and socioeconomic status (SES; as defined by the Index of Multiple Deprivation 2015 score derived from post codes; *The English Index of Multiple Deprivation (IMD)*, 2015). Service usage included contact dates, appointment type (assessment, treatment, review, follow-up) and number of sessions attended. Clinical characteristics included baseline depression, anxiety and pre-treatment functioning. Treatment characteristics included level of stepped care (step two or three), mode of therapy (1:1 or group) and therapist's code.

4.2.5 Data analysis

4.2.5.1 Data preparation

The dataset was cleaned and prepared for analysis. To aid interpretation of variables with multinomial categories, ecologically valid category levels were merged to create dichotomous classifications (Firth, Barkham, Kellett, & Saxon, 2015). The

employment variable became dichotomised as *employed/retired* (encompassing 'employed'; *n*=302, 'retired'; *n*=43, and 'student'; *n*=102) and *unemployed* (encompassing 'unemployed'; *n*=107, 'long-term sick/disabled'; *n*=124, and 'homemaker/full-time carer; *n*=38). The ethnicity variable became dichotomised as *white* (*n*=683) and *ethnic minorities* (encompassing 'Asian'; *n*=41, 'Black'; *n*=14, 'Chinese'; *n*=3, 'Somali'; *n*=2, 'Yemeni'; *n*=8, 'mixed'; n=13, and 'other'; *n*=22). IMD deciles were grouped into low (deciles 1-3; *n*=478), medium (deciles 4-6; *n*=128) and high (deciles 7-10; *n*=174) SES classifications.

The primary outcome was depression as measured by the PHQ-9; treatment effects on depression were evaluated through both summary-level outcomes (as a continuous variable: PHQ-9 scores) and patient-level outcomes (as a categorical variable; recovery classification based on change on the PHQ-9). Secondary outcomes were summary effects on anxiety (GAD-7 scores) and impaired functioning (WSAS scores). Recovery classifications were defined by applying reliable and clinically significant change criteria to depression outcomes (Jacobson & Truax, 1991). Reliable change is deemed to have occurred when change in patients' scores exceeds the measurement error of the measure. For the current study, the reliable change threshold calculated for use with IAPT outcome measures (IAPT, 2014) was applied to establish the five recovery categories defined below;

1) *Deterioration* was recorded when there was a reliable increase in PHQ-9 scores of ≥6.
2) *Improvement* was recorded when there was a reliable decrease in PHQ-9 scores of ≥6.
3) *Clinical change* (i.e., IAPT moving to recovery metric) was recorded when PHQ-9 scores moved from above to below the clinical cut-off (<10; i.e., *change in caseness*).

4) *Recovery* was recorded when there was an decrease in PHQ-9 scores of ≥6 (i.e., improvement), in addition to 'clinical change'. Accordingly, the 'recovery' category was not mutually exclusive with the 'clinical change' category.
5) A *stasis* outcome was recorded for cases where no reliable change occurred in either direction on the PHQ-9 (i.e., neither improvement or deterioration was present).

As patients below depression casesness at pre-treatment were excluded from the study, a *harm* outcome was not possible (i.e., reliable deterioration in addition to moving from below to above the clinical cut-off).

4.2.5.2 Handling missing data

Data were analysed using the intention-to-treat (ITT) principle, including all patients who entered treatment and had at least two sets of outcomes in the analysis. As outcomes were collected at every session, missing data were accounted for in pre-post treatment analyses using last observation carried forward (LOCF) imputation. The final available outcome was used as the post score. Although LOCF has documented statistical limitations (Lachin, 2016), it was deemed clinically applicable to IAPT criteria in that patients are classified as having 'received treatment' if they have attended at least two treatment sessions. As at least two PHQ-9 scores was a requirement for study inclusion, pre-post change could be calculated for all patients on the primary outcome. In terms of the secondary outcomes, there was no missing data for the GAD-7. Pre-post change could not be calculated for *n*=117 WSAS scores.

4.2.5.3 Evaluating clustering effects

The dataset had an inherent nested structure, with patients who received HI group BA nested within separate BAG groups (i.e., therapist dyads) and LI one-to-one BA and HI one-to-one BA patients nested within therapists. Nested data are more likely

to be similar to data within the same cluster than to data from other clusters (Baldwin et al., 2011). Consequently, highly clustered outcomes can impact assumptions of independence and affect analyses (Killip, Mahfoud, & Pearce, 2004). Intra-class correlation coefficients (ICC) were used to estimate the level of variance attributable to therapist/group level factors, in order to give an indication of the extent of clustering in the dataset (Equation 1).

$$ICC = \frac{MS_{\text{Therapist Cluster}} - MS_{\text{error}}}{MS_{\text{Therapist Cluster}} + (\text{average cluster size} - 1)MS_{\text{error}}} \quad \text{(Equation 1)}$$

Note: Mean square (MS) values were obtained from a one-way ANCOVA of post-treatment score using therapist/group code as the fixed-factor and pre-treatment score as a covariate

Using the ICC estimate, the design effect (DE) was calculated (Equation 2). A DE of greater than two was used as an indication of significant co-dependence that would be unsuitable for analysis on a single-level (i.e., would require use of a multi-level model) (Muthen & Satorra, 1995).

$$DE = 1 + (\text{average cluster size} - 1)*ICC \quad \text{(Equation 2)}$$

The mean cluster size was 11.39. ICCs calculated for PHQ-9 (0.07), GAD-7 (0.05) and WSAS outcomes (0.07) therefore produced design effects of 1.72, 1.52 and 1.72 respectively. As all the DEs were less than two, single level analyses were deemed appropriate.

4.2.5.4 Statistical analyses

Four sets of analyses were performed to 1) evaluate the summary and patient-level treatment effects across BA modes, 2) establish the dose of BA required to

experience symptom reliable improvement across BA modes, 3) investigate predictors of stasis outcomes, and 4) compare characteristics of patients who improve with those who experience stasis (see Figure 4.1). All analyses were conducted in SPSS version 24. Data were checked for violations of assumptions appropriate to each analysis (normality, homogeneity of variance/covariance, outliers, expected cell counts, multicollinearity checks).

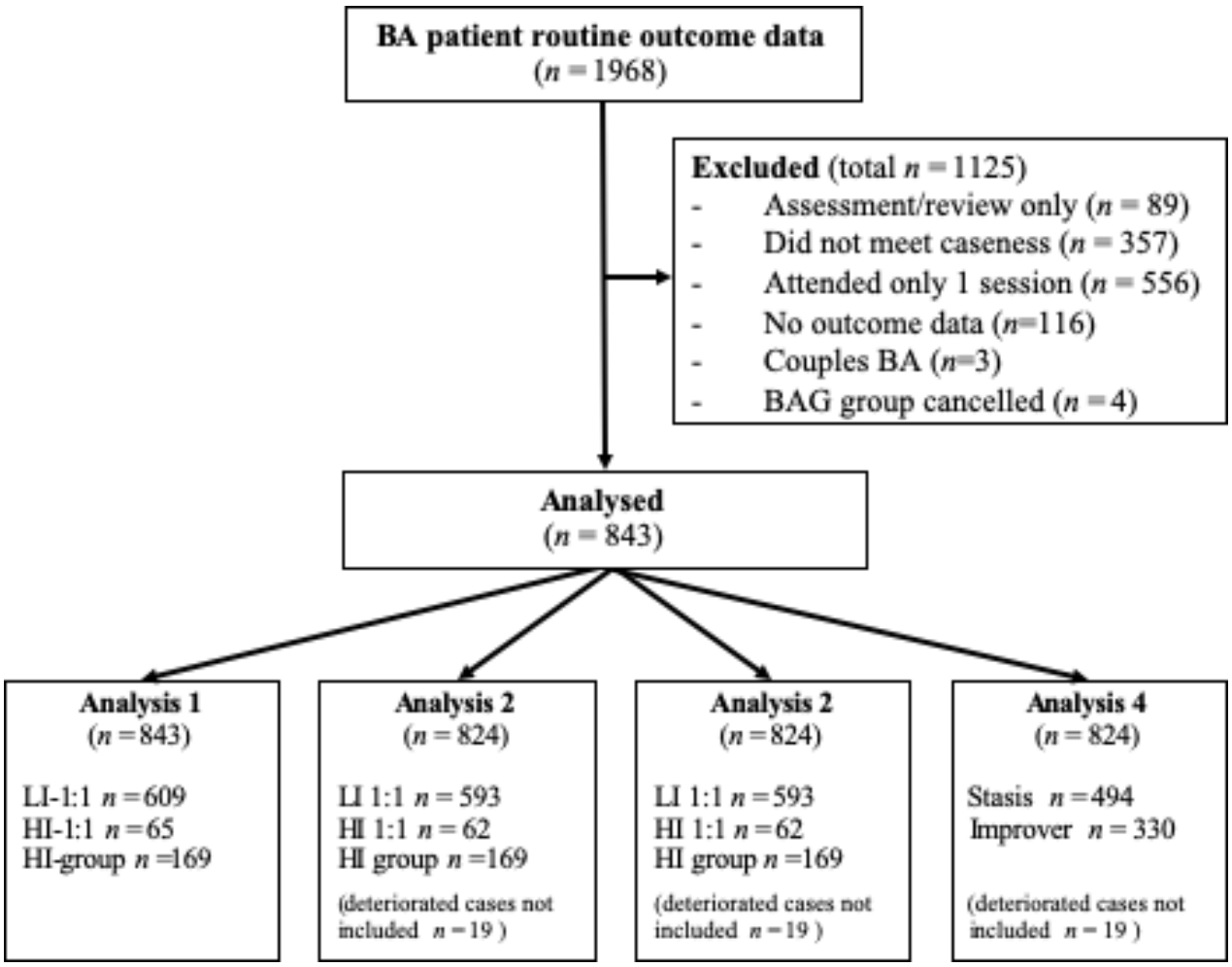


Figure 4.1. Flowchart of included sample across study analyses

4.2.5.4.1 Analysis 1: Summary-level and patient-level effect of BA modes

To evaluate summary-level outcomes, mean pre-and post-treatment scores were calculated for each BA mode. Within-group (pre-post changes) Cohen's *d* effect sizes and 95% confidence intervals (CI) were calculated to estimate the magnitude of the

treatment effects for all outcome variables. The repeated-measures effect size (d_{RM}) procedure was employed, accounting for the correlation (*r*) between pre- and post-scores and using the average standard deviation (SD) (Morris & DeShon, 2002). Cohen's *d* thresholds of 0.2, 0.5 and 0.8 were considered small, moderate, and large effect sizes respectively (Cohen, 1992). Repeated-measures ANOVAs were performed on the primary and secondary outcomes, with *pre-post scores* as the within-group factor, *BA mode* as the between-group factor, and a test of the *time by mode interaction* effect. Significant *time by mode interactions* were explored using planned contrasts. Pre-post-control design effect sizes (d_{ppc}) were calculated to compare the between-group effects of each BA mode contrast, while accounting for unequal sample sizes and potential differences in pre-test scores (Morris, 2008). The d_{ppc} effect sizes were converted into 'numbers needed to treat' (NNT; Kraemer & Kupfer, 2006). NNT provides an estimate of the number of patients who would need to be treated by the BA intervention to produce one additional beneficial outcome over the comparator BA condition. To evaluate the effect at the patient-level, the proportion of individual recovery outcomes for each BA mode was established. To compare the rate of each outcome, chi-squared tests were performed on the proportions of each recovery category across BA treatment modes. Post-hoc z-tests were performed to identify which of the three modes were significantly different from each other.

4.2.5.4.2 Analysis 2: Dose-response for symptom improvement across BA modes

Survival analyses were conducted to establish the number of sessions required for at least 50% of patients to experience improvement in depression symptoms during BA treatment. The presence or absence of reliable change on the PHQ-9 (≥6 points) was used to classify patients as either 'improved' (coded as 1 to represent the event) or 'stasis' (coded as 0 to represent 'censoring'). When data was coded as 'censored', it meant either the event was never experienced (i.e., there was not sufficient change to

indicate improvement) or data on survival time was missing (under LOCF principles, this was taken to indicate no change). Deterioration outcomes were excluded from analysis to focus on 'stasis', rather than all negative outcomes. Time was recorded as the first session at which patients had experienced a reliable improvement in their depression symptoms or the final session they attended if they experienced a stasis outcome (i.e., were censored). Thresholds for 50%, 75% and 95% probability of recovery for treatment duration of each BA mode were obtained using a Kaplan-Meier plot (to enable comparison with previous IAPT dose-response estimates; (to enable comparison with previous IAPT dose-response estimates; Robinson et al., 2019). The log-rank (Mantel-Cox) test followed by pairwise comparisons were used to determine significant differences between modes (using a Bonferroni correction of $p < .017$). Number needed to treat was estimated from each BA mode improvement rate (1/improvement rate*100). Separate plots were also applied to each mode, with curves for post-treatment improvers versus stasis to establish at which session improvers diverged from stasis patients.

4.2.5.4.3 Analysis 3: Predictors of stasis outcomes

Backward-elimination logistic regression models were employed to investigate predictors of stasis versus improver outcomes across the entire sample. Patients were classified as either *improvers* (experienced some reliable improvement or more; ≥6 point PHQ-9 reduction) or *stasis* (experienced no reliable change in either direction; ± <6 PHQ-9 points). Deterioration outcomes were excluded from analysis to focus on 'stasis', rather than all negative outcomes. Predictor variables comprised demographic variables (age, gender, ethnicity, employment, SES), clinical variables (pre-treatment PHQ-9, GAD-7 and WSAS severity) and treatment variables (BA mode, number of sessions attended). Due to the risk of unreliable results in large samples with multiple predictors (Altman, Gore, Gardner, & Pocock, 1983), the sample was randomly split

into an 'estimation' (*n*=697) and 'validation' (*n*=683) sample. All predictors were inputted into an initial estimation model, before implementing the backward-elimination procedure allowing a final estimation model to be produced that only retained significant predictors (at $p < .05$ level). The model was then applied to the validation sample. Any predictors that failed to be validated (significant at $p < .05$) were removed to leave a final predictive model for risk of a stasis outcome. The Hosmer-Lemeshow test was used to assess goodness of model fit ($p < .05$ indicates poor fit; Lemeshow & Hosmer, 1982). Separate backward-elimination logistic regression models were then produced for BA modes separately to investigate whether risk of stasis outcomes varied within BA modes.

4.2.5.4.4 Analysis 4: Comparison of 'improver' versus 'stasis' patients

Patients were grouped according to the same classification used in analysis four ('stasis'; *n*=494, 'improved'; *n*=330). T-tests and chi-squared tests were used to compare *stasis* and *improver* patients on variables relating to demographics (age, gender, ethnicity, employment and SES,) clinical features (pre-treatment PHQ-9, GAD-7 and WSAS severity) and treatment factors (number of sessions attended). The analyses were performed on the entire sample and separately for each BA mode.

4.3 Results

4.3.1 Sample characteristics

Patients' demographic and clinical characteristics across steps of service and BA models are presented in Table 4.2. Patients receiving different modes of BA did not differ significantly in terms of their age, gender, ethnicity, employment status or on their mean baseline PHQ-9 score. There were significant differences across patients receiving different BA modes on baseline GAD-7, WSAS scores, percentage meeting caseness for anxiety, and percentage with low SES. More patients who also met clinical

caseness for anxiety received HI one-to-one BA over either LI one-to-one BA or HI group BA. HI group BA treated patients with more impaired functioning (compared to LI BA), but less severe co-morbid anxiety. Overall, a greater proportion of patients with low SES were treated with LI one-to-one BA than either versions of HI BA.

Table 4.2. *Demographic and clinical characteristics for BA modes of delivery*

	BA Mode				
	LI-1:1 (N=609)	**HI-1:1 (N=65)**	**HI-group (N=169)**	**Chi-square / ANOVA[a]**	***p* value**
Demographic					
Mean age in years (SD)	39.21 (15.20)	38.75 (14.20)	38.82 (16.16)	$F = 0.06$	.941
Gender, % female (*n*)	62% (378)	51% (33)	52% (81)	$X^2 = 7.54$	.023
Ethnicity, % White British (*n*)	77% (467)	85% (55)	86% (145)	$X^2 = 7.94$	.019
Employment, % employed (*n*)	58% (291)	59% (38)	70% (118)	$X^2 = 7.38$	.025
SES, % low	66% (402)	48% (31)	42% (45)	$X^2 = 32.82$	**<.001 LI>HI & Grp**
Clinical					
Mean PHQ-9 score (SD)	18.11 (4.62)	18.17 (4.25)	17.45 (4.33)	$F = 1.56$	.214
Mean GAD-7 score (SD)	14.36 (4.75)	14.89 (3.89)	12.97 (4.75)	$F = 6.97$	**.001 Grp<LI & HI**
Mean WSAS score (SD)	21.01 (8.28)	22.40 (8.06)	23.20 (7.38)	$F = 5.40$	**.005 Grp>LI**
Anxiety, % caseness (*n*)	90% (550)	100% (65)	86% (145)	$X^2 = 10.73$	**.005 HI>LI & Grp**
Mean number of sessions (SD)	2.86 (1.32)	5.43 (3.39)	5.80 (1.81)	$F = 207.48$	**<.001 Grp & HI>LI**

[a] Differences in pre-treatment characteristics between the treatment modes were assessed with ANOVA (Welch test for large sample sizes) for continuous variables and chi-square for categorical variables; Significant p-values highlighted in **bold** (categorical variables were assessed against a Bonferroni adjustment ($p < .016$) to account for multiple post-hoc z-tests of pairwise proportions across BA modes). Abbreviations; LI-1:1: low intensity one-to-one BA; HI-1:1: high intensity one-to-one BA; HI-group: high intensity group BA.

Figure 4.2 displays the number of sessions attended in the context of service session limits across BA modes. For LI and HI one-to-one BA, session attendance dropped off as service limits were approached, whereas HI group BA attendance remained more constant. There were a few cases of patients receiving more LI and HI one-to-one BA sessions than the service limits. Both HI one-to-one BA (p<.001) and HI group BA (p<.001) patients attended significantly more sessions than LI one-to-one BA, but did not differ from each other (p=.676).

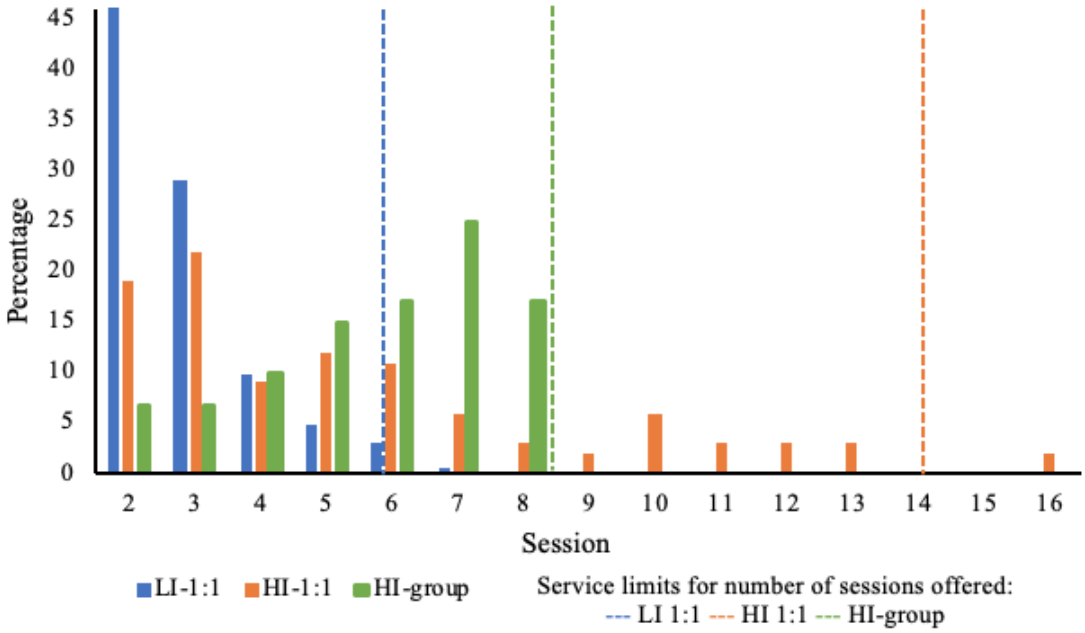


Figure 4.2. Number of sessions attended across modes of BA (N = 843)

4.3.2 Analysis 1: Treatment effect of BA modes

The first aim was to compare the summary and patient level effects of different treatment modes of BA.

4.3.2.1 Summary-level treatment outcomes across BA modes

Table 4.3 presents the means, SDs and pre-post effect sizes for all outcome measures within each mode of BA delivery. Analysis of pre-post effect sizes demonstrated moderate post-treatment reductions in depression symptoms after LI one-to-one BA, in contrast to large reductions after both HI one-to-one BA and HI group

BA. Small-moderate effects of anxiety reductions were observed after LI one-to-one BA, whereas moderate effects were observed for both HI versions of BA. Improvements in functioning represented small, small to moderate and moderate effects after LI one-to-one, HI one-to-one BA and HI group BA respectively.

Table 4.3. *Means (SD), pre-post effect sizes (dRM) and NNT for modes of BA delivery on primary and secondary outcomes (N =843)*

	BA Intervention		
	LI-1:1 (N=609)	**HI-1:1 (N=65)**	**HI-group (N=169)**
PHQ-9			
Pre-treatment	18.11 (4.62)	18.17 (4.25)	17.45 (4.33)
Post-treatment	14.40 (6.57)	13.03 (6.78)	12.73 (5.80)
Change score	-3.71 (5.74)	-5.14 (6.37)	-4.72 (5.17)
Pre-post *r*	.52	.41	.51
d_{RM} (95% CI)	0.68	0.86	0.94
	(0.59 to 0.76)*	(0.55 to 1.17)*	(0.76 to 1.12)*
GAD-7			
Pre-treatment	14.36 (4.75)	14.89 (3.89)	12.97 (4.75)
Post-treatment	11.92 (5.74)	11.28 (5.99)	9.96 (5.19)
Change score	-2.44 (4.85)	-3.62 (5.49)	-3.01 (4.25)
Pre-post *r*	.59	.45	.64
d_{RM} (95% CI)	0.51	0.70	0.71
	(0.44 to 0.59)*	(0.41 to 0.98)*	(0.57 to 0.86)*
WSAS			
Pre-treatment	21.01 (8.28)	22.40 (8.06)	23.20 (7.38)
Post-treatment	18.38 (9.77)	17.43 (9.74)	18.36 (8.75)
Change score	-2.63 (7.30)	-4.97 (9.80)	-4.84 (7.64)
Pre-post *r*	.68	.41	.56
d_{RM} (95% CI)	0.37	0.51	0.64
	(0.29 to 0.44)*	(0.23 to 0.79)*	(0.49 to 0.80)*

Note: Pre-post effect sizes (d_{RM}) have been calculated taking the correlation (*r*) between pre-post scores into account and using average SD, as suggested by Morris and DeShon (2002); *indicates significant effect as CI do not cross zero. Abbreviations; LI-1:1: low intensity one-to-one BA; HI-1:1: high intensity one-to-one BA; HI-group: high intensity group BA.

A two-way, mixed ANOVA showed depression symptoms significantly decreased following BA treatment (pre-post main effect; $F(1, 840) = 248.45$, $p < .001$,

η^2 = .23) and significantly differed across BA treatment modes (BA treatment effect; $F(2, 840) = 4.12$, $p = .017$, $\eta^2 = .01$). The interaction between time and mode of BA on depression scores (Figure 4.3) was significant ($F(2, 840) = 3.47$, $p = .032$, partial η^2 = .008) indicating amount of pre-post symptom reduction differed between modes of BA. Planned contrasts on the time by BA mode interaction indicated significantly greater pre-post reductions after HI group BA ($p = .041$) compared to LI one-to-one BA. The greater reductions after one-to-one HI BA compared to LI BA fell short of the significance threshold ($p = .053$). The difference between HI one-to-one BA and HI group BA reductions in depression was not significant ($p = .610$). The reductions in symptoms after HI one-to-one BA and HI group BA compared to LI one-to-one BA represented small effects (see Table 4.4), meaning for every 6-8 patients treated with HI-versions of BA, one additional patient will experience a benefit compared to LI BA.

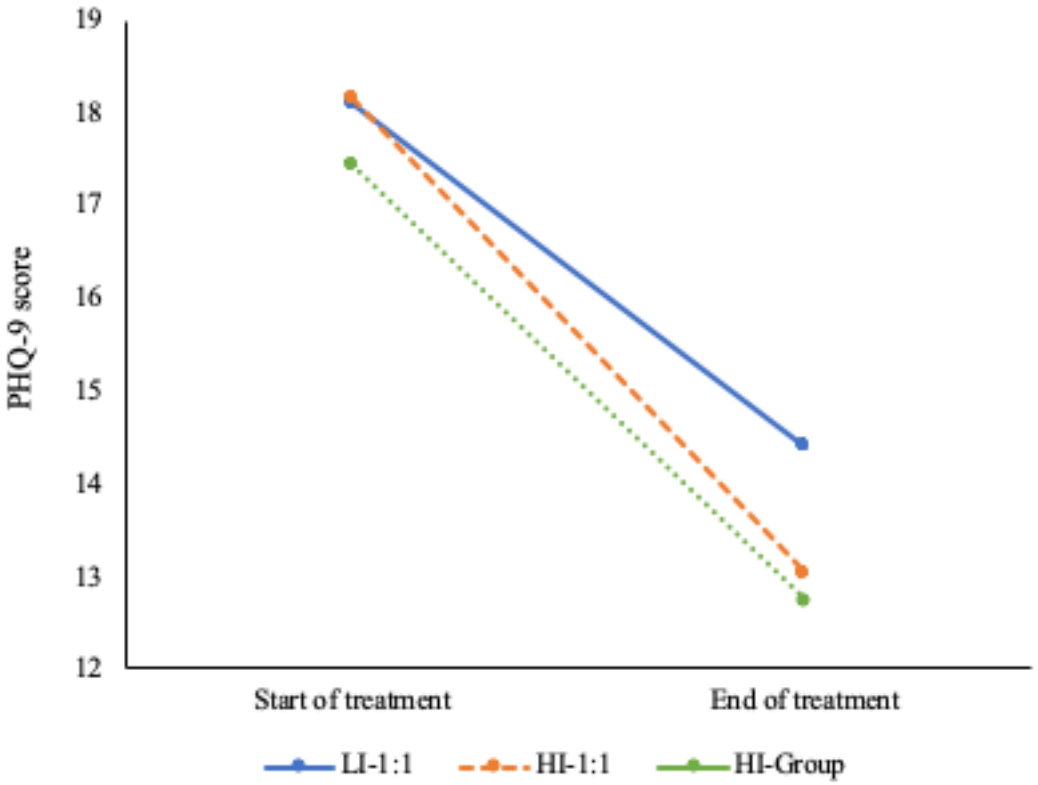


Figure 4.2. Pre-post treatment depression scores according to mode of BA delivery

Similar patterns were found for two-way, mixed ANOVA conducted on the impaired functioning WSAS scores (Figure 4.3). There was a significant interaction between time and mode of BA on functioning (interaction; $F(2, 723) = 6.84$, $p = .001$, partial $\eta^2 = .02$; pre-post main effect; $F(1, 723) = 113.84$, $p < .001$, $\eta^2 = .14$; BA treatment effect; $F(2, 723) = 1.16$, $p = .313$, $\eta^2 = .003$). Planned contrasts showed the HI versions of BA produced significantly greater improvements in functioning than LI one-to-one BA (HI one-to-one BA; $p = .021$; HI group BA; $p = .001$), but did not differ from each other ($p = .908$) (see Table 4.4).The interaction between time and mode of BA on anxiety symptoms (GAD-7) was not significant (interaction; $F(2, 840) = 2.41$, $p = .091$, partial $\eta^2 = .01$; pre-post main effect; $F(1, 840) = 156.18$, $p < .001$, $\eta^2 = .16$; BA treatment effect; $F(2, 840) = 8.91$, $p < .001$, $\eta^2 = .02$).

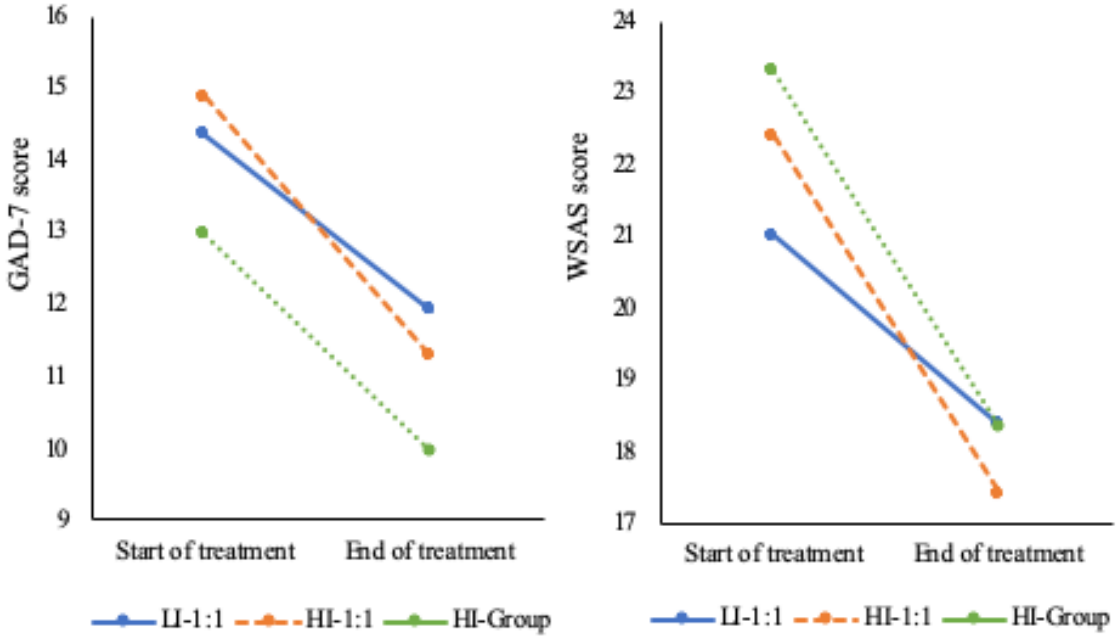


Figure 4.3. Pre-post treatment anxiety (GAD-7) and functioning (WSAS) scores according to mode of BA delivery.

Table 4.4. *Pre-post-control design effect sizes (d_{ppc}) and NNT for BA mode contrasts on primary and secondary outcomes*

	BA Mode Between-Group Contrasts		
	LI-1:1 vs. HI-1:1	**LI-1:1 vs. HI-group**	**HI-1:1 vs. HI-group**
PHQ-9			
d_{ppc}	0.31	0.22*	-0.09
Interpretation	Small	Small	Minimal
NNT	5.76	8.09	-19.71
GAD-7			
d_{ppc}	0.25	0.12	-0.13
Interpretation	Small	Minimal	Minimal
NNT	7.13	14.79	-13.65
WSAS			
d_{ppc}	0.28*	0.27*	-0.02
Interpretation	Small	Small	Minimal
NNT	6.37	6.61	-88.62

Note: Interpretation of d_{ppc} is based on Cohen's thresholds of 0.2, 0.5 and 0.8 representing small, moderate, and large effect sizes respectively (Cohen, 1992); NNT = number needed to treat; *indicates significant planned contrast. Abbreviations; LI-1:1: low intensity one-to-one BA; HI-1:1: high intensity one-to-one BA; HI-group: high intensity group BA.

4.3.2.2 Patient-level treatment outcomes across BA modes

Table 4.5 presents the proportion of patient-level treatment responses at post-treatment for each mode of BA. Prior to receiving BA treatment, all patients scored above the clinical cut-off for depression on the PHQ-9. At post-treatment, more than 50% of patients experienced a stasis outcome in all three modes of BA. Although HI versions of BA had higher rates of recovery and improvement compared to LI BA, only stasis rates and clinical change were significantly different across the BA modes. Pairwise comparisons showed significantly higher rates of stasis were present after LI one-to-one BA compared to the HI version of one-to-one BA. The difference between LI one-to-one BA and HI group BA stasis and improvement rates was not statistically

significant. Deterioration rates were low across the entire sample with no significant differences between BA modes.

Table 4.5. *Individual treatment response for BA modes of delivery at treatment completion*

Post-treatment PHQ-9 recovery status	**LI-1:1 (n=609)**	**HI-1:1 (n=65)**	**HI-group (n=169)**	**Chi-square X^2 (*p* value)**	**Post-hoc z-tests**
Recovered	20% (119)	29% (19)	25% (42)	4.83 (.090)	-
Improved	12% (70)	14% (9)	17% (28)	3.16 (.206)	-
Clinical change	24% (147)	35% (23)	31% (53)	**6.44 (.040)**	HI-1:1, HI-grp
Stasis	66% (376)	52% (30)	59% (88)	**7.40 (.025)**	LI-1:1 > HI-1:1
Deteriorated	3% (16)	5% (3)	0% (0)	5.93 (.052)	-

Note. Significant values highlighted in **bold.** Abbreviations; LI-1:1: low intensity one-to-one BA; HI-1:1: high intensity one-to-one BA; HI-group: high intensity group BA.

4.3.3 Analysis 2: Dose-response effect for BA modes

The second aim was to investigate the number of sessions required for patients to experience improvement in their depression symptoms across the different modes of BA. Figure 4.4 plots the treatment duration survival curves for LI one-to-one BA, HI one-to-one BA, and HI group BA against 50%, 75% and 95% probability of improvement thresholds.

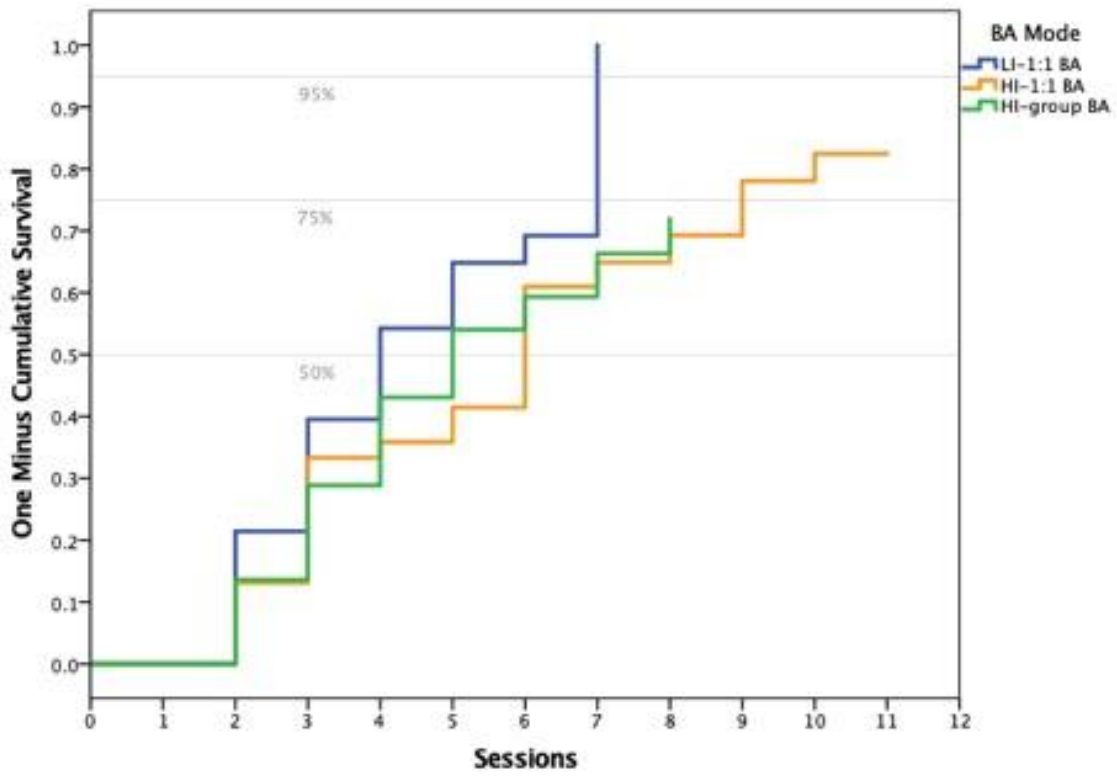


Figure 4.4. Survival curves for time to improvement for BA modes of delivery.

Table 4.6 summarises the improvement thresholds, mean and median time to improvement and corresponding NNT estimates for BA modes. For LI one-to-one BA, four sessions were required to identify 50% of improvers and by session seven there was a 95% probability of improvement. For HI one-to-one BA, six sessions were required to identify 50% of improvers, while 9 sessions identified 75% of improvers (95% threshold could not be calculated). For HI group BA, five sessions were required to identify 50% of improvers, but thresholds for 75% and 95% improvement could not be calculated. Log-rank (Mantel-Cox) test showed there were significant differences in survival times between BA modes (X^2 = 10.59, p = .005). Pairwise comparisons identified significant differences between LI one-to-one BA and HI group BA (X^2 = 8.31, p = .004) and LI one-to-one BA and HI one-to-one BA (X^2 = 5.66, p = .018), but not HI group BA and HI one-to-one BA (X^2 = .111, p = .739). The NNT corresponding to improvement rates showed that although LI one-to-one BA appeared to produce

faster improvement rates overall, more patient needs to be treated to experience an additional beneficial outcome compared to HI one-to-one BA and HI group BA.

Table 4.6. *NNT, mean and median survival times to improvement for modes of BA*

BA Mode	***N*=824**	**NNT**	**50% threshold**	**75% threshold**	**95% threshold**	**Mean survival time (95% CI)**	**Median survival time (95% CI)**
LI-1:1	592	3.14	4	7	7	4.51 (4.24 - 4.78)	4.00 (3.45 - 4.55)
HI-1:1	61	2.33	6	9	-	6.21 (5.25 - 7.17)	6.00 (5.19 - 6.81)
HI-group	169	2.38	5	-	-	5.35 (4.98 - 5.71)	5.00 (4.01- 5.99)

Abbreviations; BA: behavioural activation; NNT: number needed to treat; LI-1:1: low intensity one-to-one BA; HI-1:1: high intensity one-to-one BA; HI-group: high intensity group BA.

Figure 4.5 plots the separate survival curves for BA modes categorised by end of treatment improver versus stasis outcomes. Across all modes of BA, survival curves for improver patients were differentiated from stasis patients after two sessions. False improvement (reliable change that did not remain at the end of treatment) was found in 5.0% of LI one-to-one BA patients, 14.7% of HI one-to-one BA patients and 27.3% of HI group BA patients. Applying 50-95% improvement thresholds to all improver patient outcomes suggested the optimal number of sessions ranges between 2-4 for LI one-to-one BA, 3-9 for HI one-to-one BA and 4-7 for HI group BA.

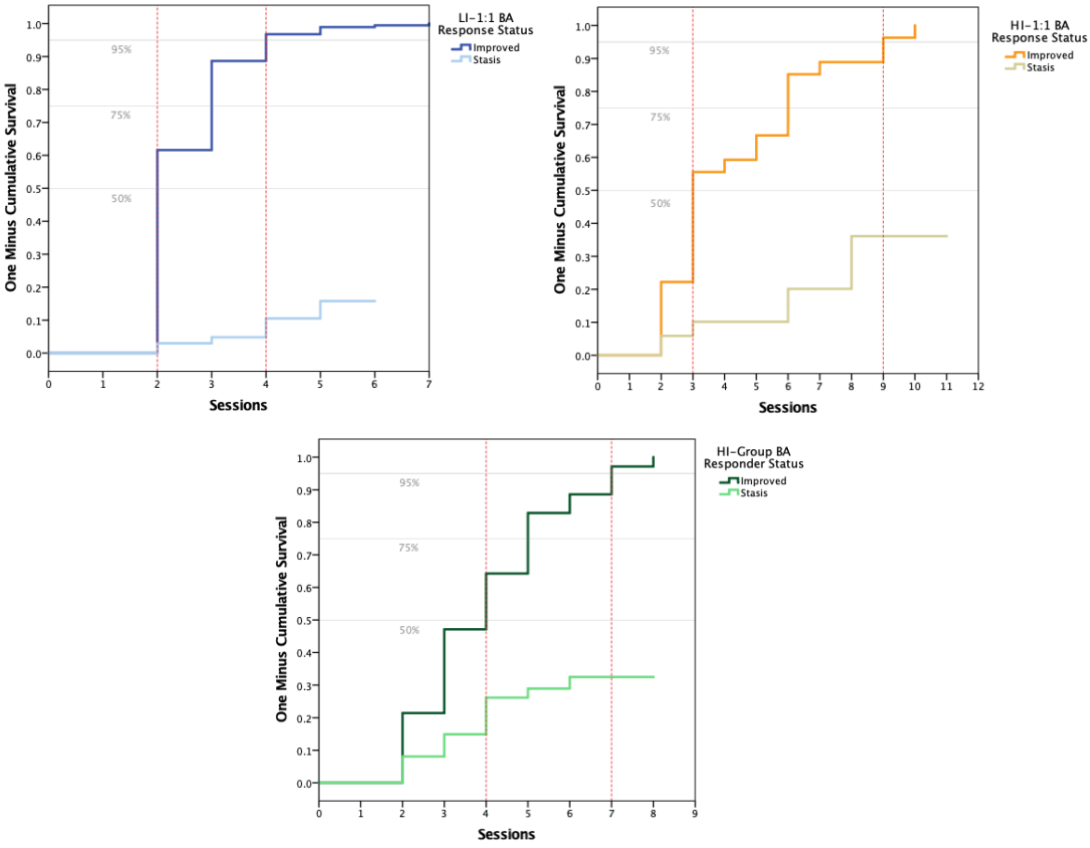


Figure 4.5. Survival curves for improver versus stasis outcomes across separate BA modes.

4.3.4 Analysis 3: Predicting risk of a stasis outcome after BA treatment

The third aim was to explore predictors of stasis outcomes after BA treatment for depression. Estimation and validation logistic regression models for all modes of BA are presented in Table 4.7. When applied to the initial estimation sample, a logistic regression model using backward elimination produced a model that explained 20% of the variance in stasis outcomes, with correct classification of 67.3%% of cases. Variables that remained in the model suggested that risk of a stasis outcome was predicted by the following: not being in employment; attending fewer sessions; starting treatment with higher levels of anxiety and functional impairment and lower levels of depression.

When the final estimation model was applied to the validation sample, pre-treatment anxiety levels and unemployment were not validated as significant predictors of stasis. Therefore, they were removed. The remaining variables (attending fewer sessions; greater pre-treatment functional impairment; lower pre-treatment depression levels) were validated as significant predictors of stasis outcomes. These three variables comprised the final model, classifying 65.1% correctly.

Table 4.7. *Predictors of stasis for BA treatment validated across estimation and validation samples*

	Full estimation sample model (n = 415)				Final estimation sample model (n = 415)				Final validation sample model (n = 409)			
	Nagelkerke R^2 = .233 Hosmer-Lemeshow X^2 = 5.87, p = .662				Nagelkerke R^2 = .199 Hosmer-Lemeshow X^2 = 11.05, p = .199				Nagelkerke R^2 = .101 Hosmer-Lemeshow X^2 = 2.47, p = .963			
Variable	B	SE	β	*P*	B	SE	β	*p*	B	SE	β	*p*
Constant	*3.883*	*.807*	*48.593*	***<.001***	*2.606*	*.604*	*13.551*	***<.001***	*1.801*	*.513*	*6.054*	***<.001***
Number of sessions	-.289	.081	.749	**<.001**	-.217	.057	.805	**<.001**	-.214	.054	.807	**<.001**
Pre-treatment WSAS	.065	.018	1.068	**<.001**	.068	.018	1.070	**<.001**	.050	.017	1.051	**.003**
Pre-treatment PHQ-9	-.223	.044	.800	**<.001**	-.223	.043	.800	**<.001**	-.083	0.30	.920	**.006**
Pre-treatment GAD-7	.083	.036	1.087	**.020**	.075	.035	1.078	**.033**				
Unemployed (ref = employed)	.365	.281	1.440	.195	.571	.260	1.617	**.028**				
BA mode: LI-1:1 (ref)				.238								
BA mode: HI-1:1	.373	.471	1.453	.427								
BA mode: HI-group	.764	.453	2.148	.091								
SES: low (ref)				.149								
SES: medium	-.406	.347	.666	.242								
SES: high	-.628	.339	.534	.064								
Age	-.013	.009	.987	.147								
Gender: female (ref = male)	-.270	.268	.763	.313								
Ethnicity: white (ref = minority)	-.361	.351	0.697	.304								

Abbreviations; ref: reference category; BA: behavioural activation; LI-1:1: low intensity one-to-one BA; HI-1:1: high intensity one-to-one BA; HI-group: high intensity group BA; SES: socioeconomic status.

To explore whether predictors of stasis differed within modes of BA treatment, separate backward elimination logistic regression models were also applied. Table 4.8 reports the final models for each BA mode. Stasis outcomes after LI one-to-one BA were predicted by attending fewer sessions, being unemployed, low SES (compared to high SES), greater impaired functioning and lower depression severity prior to treatment. HI one-to-one BA stasis outcomes were predicted by attending fewer sessions and being male. HI group BA stasis outcomes were predicted by having lower SES (compared to medium SES), greater impaired functioning and lower baseline depression severity. Fit statistics demonstrated acceptable fit for all three models.

Table 4.8. *Significant predictors of stasis across modes of BA therapy*

	LI-1:1 BA (n = 593)		HI-1:1 BA (n = 62)		HI-group BA (n = 169)	
	Nagelkerke R^2 = .177 Hosmer-Lemeshow X^2 = 2.86, p = .943		Nagelkerke R^2 = .220 Hosmer-Lemeshow X^2 = 8.15, p = .419		Nagelkerke R^2 = .232 Hosmer-Lemeshow X^2 = 5.28, p = .727	
Variable	β	*p*	β	*P*	β	*p*
Constant	*11.885*	***<.001***	*8.318*	***.003***	*25.647*	***.004***
Number of sessions	.675	**<.001**	.799	**.013**		
Unemployed (ref = employed)	1.704	**.019**				
SES: Low (ref)		**.038**				**.012**
SES: Medium	1.110	.724			.163	**.003**
SES: High	.535	**.020**			.426	.088
Baseline WSAS	1.066	**<.001**			1.080	**.024**
Baseline PHQ-9	.889	**<.001**			.792	**.001**
Baseline GAD-7						
Gender: female (ref = male)			.295	**.036**		
Ethnicity: white (ref = minority)						
Age						
Correctly classified %	70.9%		70.5%		67.6%	

Abbreviations; ref: reference category; BA: behavioural activation; LI-1:1: low intensity one-to-one BA; HI-1:1: high intensity one-to-one BA; HI-group: high intensity group BA; SES: socioeconomic status.

4.3.5 Analysis 4: Comparing 'improver' versus 'stasis' patients

The final aim was to compare characteristics of patients who experienced improvement after BA with those who experienced a stasis outcome. Table 4.9 presents the characteristics of improver versus stasis patients after BA treatment overall and across the three BA modes. Across the entire sample, patients who experienced a stasis outcome had attended fewer sessions, had significantly higher impaired functioning and anxiety scores and included a greater proportion of people with lower SES and current unemployment. LI one-to-one BA stasis patients were younger, had more impaired functioning, attended fewer sessions and were more likely to be currently unemployed and have lower SES. HI group BA stasis patients had higher baseline co-morbid anxiety symptoms. Meanwhile, HI one-to-one BA stasis patients were not discerned by any factor.

In summary, demographic factors including age, gender, ethnicity did not affect risk of a stasis outcome. BA stasis outcomes were predicted by fewer sessions attended, higher functional impairment and lower depression scores, with fewer sessions emerging as the strongest predictor of stasis likelihood across analyses (although not present in all models). In the entire sample, stasis patients were significantly differentiated from improver patients based on attending fewer sessions, having greater functional impairment and levels of anxiety, in addition to unemployment and lower SES. Within the LI BA sample, stasis patients were also found to be younger than responders, whereas there was limited distinction between stasis and responder patients in both HI versions of BA (only higher anxiety in HI 1:1 BA).

Table 4.9. *Characteristics of improver versus stasis patients after BA treatment for depression*

Variable	Entire BA sample		LI-1:1 BA		HI-1:1 BA		HI-group BA	
	Improved (*n* =330)	Stasis (*n* =494)	Improved (*n* =217)	Stasis (*n* =376)	Improved (*n* =32)	Stasis (*n* =30)	Improved (*n* =81)	Stasis (*n* =88)
Mean (SD)								
Age	40.20 (15.62)	38.33 (15.18)	**41.14 (15.58)**	**38.00 (15.00)**	37.47 (15.04)	40.93 (13.56)	38.75 (15.91)	38.88 (16.48)
Pre-treatment PHQ-9	17.83 (4.66)	18.24 (4.43)	17.92 (4.88)	18.39 (4.43)	17.94 (3.68)	18.90 (4.70)	17.56 (4.40)	17.35 (4.29)
Pre-treatment GAD-7	**13.72 (4.62)**	**14.44 (4.77)**	13.99 (4.63)	14.64 (4.80)	**13.91 (3.95)**	**16.03 (3.57)**	12.91 (4.82)	13.02 (4.72)
Pre-treatment WSAS	**20.48 (7.86)**	**22.71 (8.15)**	**19.40 (7.80)**	**22.28 (8.44)**	22.19 (4.04)	23.20 (8.05)	22.37 (7.64)	23.95 (7.09)
No. of sessions attended	**4.19 (2.38)**	**3.31 (1.83)**	**3.20 (1.69)**	**2.65 (1.00)**	6.31 (3.60)	4.77 (3.06)	6.01 (1.67)	5.61 (1.91)
%								
Gender (% female)	60%	58%	64%	61%	59%	40%	52%	52%
Ethnicity (% minority)	18%	23%	20%	25%	13%	20%	14%	15%
Employment (% employed)	**68%**	**56%**	**68%**	**52%**	68%	47%	67%	73%
SES (% low) IMD decile	**52%**	**66%**	**58%**	**70%**	50%	43%	32%	51%

[a] Improver versus stasis differences were assessed with ANOVA for continuous variables and chi-square for categorical variables; Significant differences highlighted in **bold**. Abbreviations; LI-1:1: low intensity one-to-one BA; HI-1:1: high intensity one-to-one BA; HI-group: high intensity group BA; SES: socioeconomic status; IMD; Index of Multiple Deprivation.

4.4 Discussion

4.4.1 Summary of results

The purpose of this study was to investigate depression treatment response after stepped-care delivery of BA, varying by intervention intensity, format and dose. In addition, the study explored factors that might predict whether a patient will fail to benefit from BA treatment (termed stasis) in a routine psychological care context.

BA was shown to be effective at reducing depression regardless of format (1:1 compared to group). However, larger treatment effects (both versions of HI BA) and fewer stasis rates (HI one-to-on BA only) were seen for high intensity compared to low intensity versions of BA. Between 4-6 sessions were needed to identify 50% of treatment improvers, with the majority of improver patients identified after between seven LI and nine HI (1:1 only) sessions respectively. LI BA produced faster rates of improvement compared to HI versions, but needed to treat more patients to get a beneficial outcome. End of treatment stasis rates ranged between 52% (for HI) and 66% (LI), and the probability of benefitting from treatment was distinguishable from risk of stasis after only two sessions of BA (irrespective of intensity or format). Risk of a BA stasis outcome was predicted by attending fewer sessions, greater impaired functioning prior to treatment and lower pre-treatment depression levels. Fewer sessions was the strongest predictor of stasis overall (apart from for the HI group BA model). Stasis patients were also distinguished from improvers across the sample as a result of having fewer treatment sessions and greater impaired functioning, in addition to higher anxiety, unemployment and lower SES. However, consistent differences between stasis and improver patients were unable to be identified across separate BA modes.

4.4.2 Variability in BA treatment response

The reductions in depression after all modes of BA observed at the summary level reflect clinical trial findings in support of BA as an effective, flexible treatment for

depression (Dimidjian et al., 2011; Ekers et al., 2014; Richards et al., 2016). However, the patient-level outcomes demonstrated substantial variability in individual BA treatment response. The number of patients improving or recovering after BA in routine services was closer to the lower psychotherapy estimates (~35%) reported by Hansen et al. (2002), whereas the rates of deterioration were considerably lower (1-3% compared to estimates of 10%). However, this study has focused directly on patients who do not experience any change (in either direction) in their depression symptoms after treatment with BA. Such stasis outcomes are extremely prevalent in routine services, with 52-66% of BA patients failing to benefit to any meaningful degree.

4.4.2.1 Predicting depression stasis outcomes

Demographic factors (age, gender, ethnicity) failed to predict risk of a stasis outcome, consistent with findings from other non-response investigations (Van et al., 2008; Vittengl et al., 2016) and general negative outcome studies (Bohart & Wade, 2013; Hamilton & Dobson, 2002). The present findings demonstrated an inconsistent association between patient factors (employment, SES), baseline clinical markers (symptom severity, functioning) and treatment factors (format, intensity, sessions attended) and stasis risk, with number of sessions attended, impaired functioning, less severe depression (and to some extent low SES and unemployment) emerging as viable predictors. Depression severity has previously been identified as a predictor of treatment response, with some studies finding that higher levels of baseline depression are associated with poorer outcomes (Delgadillo, Moreea, et al., 2016; Thibodeau et al., 2015) and other non-response studies finding lower initial depression severity predicts treatment non-response (Van et al., 2008). The present results perhaps reflect that more severely depressed patients appear to respond particularly well to BA (compared to cognitive therapy) (Coffman et al., 2007). There also may be some influence from a floor effect. Starting treatment with lower depression scores is also likely to restrict the

amount of change that can occur, meaning meeting reliable improving criteria is more difficult. Greater baseline impaired functioning also predicted risk of a stasis outcome (in the entire sample and for both LI one-to-one BA and HI group BA), reflecting outcomes seen in other IAPT dataset studies (Delgadillo, Moreea, et al., 2016).

Employment status has also been linked to therapy outcome in the literature (Delgadillo, Moreea, et al., 2016; Thase et al., 1993). However, there were somewhat inconsistent findings across BA modes. Unemployment failed to predict stasis in the entire sample or for HI BA, but did predict stasis after LI one-to-one BA and distinguish stasis patients from improvers overall. These findings may suggest that the impact of unemployment on poor treatment response is more prominent for LI versions of BA. However, it is worth noting that the reduced power to detect categorical predictors may have contributed to the failure to identify employment in the smaller HI BA samples.

The factor that most strongly predicted stasis and also distinguished improvers from stasis patients across the entire sample, was the treatment-related factor – number of sessions attended. Attending fewer sessions predicted having a stasis outcome, with improver patients attending significantly more treatment sessions than stasis patients. Interestingly, number of sessions is arguably the only predictor investigated that has the potential to be changed (i.e., you cannot change a patient's gender), and therefore represents a useful area for intervention in tackling stasis. However, number of sessions was not a significant stasis predictor in the HI group BA model showing patients engaged with group BA (i.e., attended sessions) without necessarily being more likely to experience meaningful change in their symptoms. It is not clear why this is the case or whether this finding is related to group delivery, however a similar finding was also seen in a routine practice study of group CBT. Responders and non-responders were not distinguished by the number of sessions they attended (Thimm & Antonsen, 2014). It may be that the support from, and sense of commitment to other group members

encourages attendance, even if clinical outcome benefits are not being experienced. Nevertheless, attending sessions without seeing a benefit reflected in their outcome highlights that HI group BA may need to find ways to make the treatment effects more potent for a wider subset of people who attend sessions.

4.4.3 Impact of BA treatment factors

4.4.3.1 Stepped-care BA dose-response effect

The link between dose (i.e. sessions) and positive outcome is well established in the literature (Hansen et al., 2002; Howard et al., 1986; Robinson et al., 2018). In light of the number of sessions consistently predicting stasis, the dose-response analyses provide further information on the impact of dose for treatment improvement versus stasis. The optimal number of sessions for LI one-to-one BA supports previous estimates of LI interventions (4-7 sessions; Robinson et al., 2019), with outcomes atrophying after 7-8 sessions (Firth, Barkham, Kellett, et al., 2015). Unfortunately, the upper end of dose-thresholds could not be calculated for HI versions of BA to enable accurate comparison with estimates from the literature. However, when using thresholds for number of sessions required to identify 50%-95% of treatment improvers only, both LI (2-4 sessions) and HI (4-7/3-9) BA interventions (regardless of format) were found to have 50-95% boundaries that tend to be less than those of other CBT-based interventions for affective disorders (LI: 4-7 sessions, HI: 5-14 sessions; Robinson et al., 2019). This faster pattern of response perhaps reflects the more simple, parsimonious nature of BA as a treatment for depression (Richards et al., 2016). This appears particularly true in the low intensity version of the BA model which focusses on early activation in treatment contracts. On the other hand, one-to-one HI BA is formulation driven. The time spent on formulations may limit the amount of change work that goes on in early sessions, hence the comparative slower pattern of response with LI BA.

Although LI one-to-one BA produced improvements from fewer sessions than HI BA, inspection of corresponding NNT estimates showed that more patients needed to be treated to produce those treatment responses. Comparing the mean number of sessions attended in the context of the dose-response estimates also helps explain the greater improvement rates seen for HI versions of BA. Given that LI one-to-one BA had a 50% probability of improvement by four sessions, but the mean number of sessions attended was less than three, the high rates of stasis are not surprising. In contrast, HI versions of BA had a 50% probability of improvement by five/six sessions and the mean number of sessions attended was 5.4 (HI one-to-one BA) and 5.8 (HI group BA). Taken together, this pattern further reinforces the need to ensure that patients attend for an adequate dose of therapy in order to reduce the risk of a stasis outcome.

4.4.3.2 Effect of BA format

Looking further at how treatment delivery factors impacted on outcomes, the comparability of treatment response and dose-response for HI BA delivered 1:1 or in a group suggests group delivery can be as effective as individual therapy. In contrast to some previous reports of potential larger effects for individual therapies (Cuijpers & Straten, 2008), the opposite was found here. Although not significantly different, slightly larger effect sizes were seen after HI group BA compared to HI one-to-one BA. Group and individual HI BA also did not significantly differ on stasis, improvement or recovery rates. Due to the different nature of the therapeutic context, it is possible the benefits of 1:1 and group BA are in part produced through differential therapeutic processes, irrespective of the protocol. Patients receiving 1:1 therapy are likely to benefit via a stronger, more direct therapeutic bond with the clinician (Ardito & Rabellino, 2011), whereas in groups there will be a less direct therapeutic interaction with the clinician, but a greater focus on group dynamics with other patients (Yalom & Leszcz, 2005). The comparable treatment effects (using the same HI protocol) observed

here imply that the benefits of group dynamics can be as therapeutically effective as 1:1 therapeutic work.

It is also possible that the comparability (or even slight advantage) of group BA was a function of the type of patients who are allocated to group treatment in routine services. Fewer patients with anxiety meeting clinical caseness were referred to group BA. BA is an intervention targeted at depression symptoms, therefore high levels of co-morbid anxiety/greater complexity may inhibit the beneficial effects and this difference in sample characteristics could partially explain the group BA effects relative to the 1:1 BA outcomes.

4.4.3.3 Effect of BA intensity

Larger effects observed for HI BA imply that the added complexity incorporated into HI protocols (such as values work and functional analysis) might be enabling greater treatment benefits albeit at a delayed pace. On the other hand, fewer patients benefit from LI BA (compared to HI one-to-one BA), but those who do, experience quicker improvements. Despite the stepped-care process, LI and HI patients did not significantly differ in their baseline depression severity, meaning that smaller effects for LI BA are less likely to be attributable solely to a floor effect. Instead, these effects might be explained by the differential timing in the introduction of the change mechanisms in LI and HI BA protocols. The activation component of BA is thought to be the key process of change (Curran et al., 2012), therefore as the 'doing' is prescribed immediately by the LI protocol, those who are able to immediately engage (for whatever reason) experience the benefits quicker (as seen in the dose-response findings). However, more patients need to be treated with LI BA to get beneficial results. This suggests some patients may find the LI activation method too simplistic or struggle to increase their approach behaviours in spite of their mood. Hence, the greater rates of stasis and poorer retention seen for LI BA. Additional features in HI protocols

that contextualise and ground activation work in what the patient values before prescribing the 'doing' might help these types of patients be better able to implement sustainable behaviour change later on. Change is produced at a slower rate as there is a delay in the positive reinforcement being translated into mood change. However, the non-depressed behaviours that are elicited are more effective at alleviating depression, thus producing the better treatment outcomes for HI BA. It should be noted that patient level outcomes did not significantly differ between LI BA and HI group BA. The comparable individual rates of outcome (e.g., stasis, improvement) in the context of larger overall reductions in depression after HI group BA compared to LI BA suggest the more potent HI treatment effect does not translate to everyone in a group context. This may reflect that some patients would be better suited to one-to-one versions of HI protocols in order to gain benefit, as seen by the larger treatment effects and lower rates of stasis after individual HI BA.

4.4.4 Clinical and research implications

In light of the differential summary and patient level (for HI one-to-one BA) treatment outcomes and dose-responses for LI versus HI BA, a method that can predict the types of patients who are unlikely to benefit from the immediate activation in LI BA could improve BA outcomes. Furthermore, it may be that people who responded well to HI BA, would have also responded well to LI BA. Or people who responded well to one-to-one BA, would also have responded well to group BA. The ability to predict who will benefit from what would also enable resources to be allocated more efficiently. Those who are ready, willing and able to engage in rapid activation need to be treated at step 2 for LI BA, whereas those who are likely to need more time to contextualise their activation and define their personal values before activating can be stepped straight to HI BA. Similarly, those who would benefit from BA regardless of format could be treated in an organisationally efficient group, whereas those who need more one-to-one

therapeutic support to benefit can be referred to individual treatment. The fact that lower levels of baseline depression, the current metric most commonly used to determine initial stepped care treatment allocation, predicted stasis outcomes implies predicting who will benefit from LI and HI BA requires a more complex selection process (Delgadillo, Moreea, et al., 2016; Webb et al., 2018). Furthermore, HI and LI patients could not be differentiated in terms of baseline symptom severity in the present study, perhaps suggests that the stepped-care system is not being applied as intended. Innovative machine learning paradigms offer a more powerful method for predicting outcomes that are beginning to be applied to psychotherapy (Cohen & Derubeis, 2018). Such techniques could be useful in establishing more sophisticated stepped care processes for allocating patients to LI or HI BA/individual or group treatments and help to reduce first time stasis outcomes.

Given how widespread stasis outcomes are in routine practice, it is important that they are tackled directly, as well as indirectly (using methods suggested above). Outcomes could be improved by utilising factors associated with stasis to try and tackle the problem more directly within treatments. A lower number of sessions attended emerged as the strongest predictor of stasis in general. Finding ways to increase treatment retention (and therefore reduce risk of a stasis outcome) could enhance treatment protocols. The unexplained variance in the predictive models reported here suggests there may be additional factors that are able to inform how to predict stasis outcomes. Patient characteristics tend to be poor and inconsistent predictors of stasis. Factors related to the treatment process, such as time spent on a waiting list or engagement with treatment may explain more of the variation (as outlined in Chapter 3, section 3.3.1.3). Research is needed to continue identifying factors that are associated with stasis and that could be targeted in treatment protocols.

HI group BA produced larger treatment effects compared to LI BA, but this was not reflected in the individual patient outcomes. Session attendance also did not affect likelihood of a stasis outcome for HI group BA. These findings imply that although patients engage with the treatment (through attending) and group BA produces significant reductions in depression overall, the amount of patients who experience those reductions could be improved. HI group BA could be enhanced to extend the benefits that are clearly possible, to more patients, hence reducing stasis outcomes. Likewise, the present results identified attending fewer sessions as the strongest predictor of stasis overall, yet a large proportion of patients (*n*=556) only attended one session (and therefore were not eligible to be included in the analysis: see Figure 4.1). Clearly there is something going on with these patients which is important to understand in the context of stasis. Understanding why, and finding ways to encourage these patients to persevere with treatment is a useful avenue to explore within nonresponse research.

4.4.5 Limitations and future research

The study had several limitations. It was a study of routine practice treatment delivery so treatment conditions could not be randomised. Comparisons between HI and LI/one-to-one and group versions of BA may be affected by latent factors that resulted in patients being treated by those interventions. A disproportionate number of patients were treated with LI compared to HI BA. Although a greater number of LI BA cases would be expected in a stepped care model, the amount of cases treated with individual HI BA was far lower than was expected for the timeframe. It is likely more people were treated with HI BA, but due to the system used in IAPT services were not labelled correctly (HI interventions often get coded under the generic CBT bracket, rather than distinguished as BA). As a result, the conclusions regarding HI one-to-one BA are less reliable. Similarly, the predictors of stasis identified may have been a function of the

sample sizes. The larger LI sample produced more independent predictors of stasis, whereas the smallest HI sample produced the fewest. Although the use of a validation sample tried to control for this effect in the complete sample, the individual BA mode models may have been affected. Likewise, the sample size was underpowered to adequately detect categorical stasis predictors, which means continuous predictors may have been emphasised in the models. Furthermore, variables included in the models were limited to data that was collected in routine practice, therefore only certain patient and treatment factors could be tested. Factors relating to treatment process were not available and restricted the scope of the prediction models. Although the resulting models are applicable to IAPT data systems, there is still a need for wider investigation into other factors that affect stasis. Controlled and randomised direct comparisons of group versus one-to-one BA and HI versus LI BA interventions with (suitably powered) specifically measured treatment process predictors are indicated to see if the present results can be replicated and extended.

The variation in number of BA sessions attended was suboptimal for calculating the upper ends of the dose-response limits. There were limited cases for later sessions of BA, especially for the one-to-one versions, making the upper thresholds more unreliable. Upper thresholds may also have been arbitrarily influenced by the maximum number of sessions offered in the service for each mode of BA. While the thresholds are confined, understanding the dose-response effect has to balance optimal doses of therapy with what is feasible in clinical practice. With this is mind, the term 'optimum dose' should have a caveat that it refers to the optimal dose of therapy for the context it is currently delivered in. The lack of a follow-up means the durability of effects for the different modes of BA could not be compared. Additional longitudinal research should establish whether the additional components of HI BA protocols enable more sustained behaviour change after treatment has ended. There were no treatment fidelity or

competency checks to confirm whether the LI and HI BA interventions were genuinely protocol-adherent. The results have to be interpreted under the assumption BA was delivered as intended across BA modes.

Session-by-session scores from the last available session were used to provide post-treatment scores using the last observation carried forward (LOCF) method. The LOCF method does have acknowledged faults and reduces accuracy of the results (Lachin, 2016). Multiple imputation methods would provide a more robust method for dealing with missing data. Cluster structures in the data (i.e., patients clustered in groups or within therapists) increased the risk of data dependence violations in the analyses. Checks were conducted to verify the use of single-level analyses, but it has been argued that ICCs as low as 0.01 can still impact analyses (Baldwin et al., 2011). The use of self-report outcomes are known to be at risk of validity issues, such as social desirability when providing sensitive information (Tourangeau & Yan, 2007). The self-report measures also make diagnostic certainty difficult, as there was no specific diagnostic assessment. Use of combinations of self-report and clinician-rated measures in future investigation would help address these issues.

4.4.6 Conclusion

In conclusion, format of BA delivery in routine practice does not affect treatment response, but intensity of the BA model does to some extent. Optimal doses of therapy are 4-6 sessions for LI BA and 6-9 sessions for HI BA, implying patients can experience faster rates of improvement after LI BA. However, more patients need to be treated to get an additional beneficial outcome. Methods for distinguishing who can benefit from the less intensive, simpler BA model and who will get an additional benefit from the more intensive, complex BA mode would enable improved outcomes and more efficient allocation of resources. Over half of patients risk experiencing a stasis outcome after BA treatment. Stasis outcomes can be distinguished after only two BA sessions

(irrespective of intensity or format) and are predicted by attending fewer sessions, greater impaired functioning and starting treatment with lower levels of depression. BA is evidently an effective treatment for depression, but prevalence of stasis highlights treatment effects could be enhanced. HI group BA in particular, shows the large beneficial reductions that are significantly larger than LI BA at the summary level, do not translate to individual patient outcomes (rate of stasis between LI BA and HI group BA not significantly different). Therefore there is potential for HI group BA to be made more beneficial for a greater proportion of patients (i.e. reduce stasis). Treatment process factors that predict when patients may fail to benefit provide opportunities to intervene to reduce stasis outcomes. A large proportion of patients only attended one treatment session (and could not be included in the present analysis), which clearly has implications for reducing stasis outcomes going forward. Poor session attendance was the strongest predictor of stasis outcomes overall, therefore future research should aim to develop strategies that can increase treatment retention as a method for improving outcomes.

CHAPTER 5

Testing an Intervention to Improve Outcomes at Group Behavioural Activation for Depression: A Cohort Comparison Study

The previous chapter demonstrated the effect of the intensity, format and dose of BA interventions on depression stasis outcomes, and associated stasis predictors. The findings supported the use of group BA in routine practice, but highlighted that there is a need to improve treatment to enable benefits to be experienced by more patients and hence drive down stasis rates. In addition, a consistent theme in the findings highlighted the importance of session attendance for reducing stasis risk after BA. The notion of 'quality improvement' in healthcare has been applied to psychotherapeutic interventions for depression in an attempt to improve outcomes. Yet evidence-driven quality improvement interventions integrated into existing treatments are few and far between. The relative simplicity of behavioural interventions means that behavioural activation (BA) is particularly well suited to the modification and addition to treatment protocols to potentially improve treatment outcomes. The third empirical study is therefore focused on a quality improvement approach to an existing group BA treatment within routine practice. The objective was to integrate, and test two treatment augmentations embedded in group BA - a 'bottom-up' data-driven enhancement, and a 'top-down' theory-based enhancement aimed at reducing drop-out rates and stasis outcomes respectively.

5.1 Introduction

5.1.1 Depression stasis outcomes

Despite the array of evidence-based therapies that can effectively treat depression, clearly the available treatments are no panacea. Marked variability in real

world outcomes occurs, with improvement rates ranging from 35-64% (Gyani, Shafran, Layard, & Clark, 2013; Hansen, Lambert, & Forman, 2002; Richards & Borglin, 2011). Even for the higher end of the improvement range, there are still a significant proportion not benefitting. Failure to benefit from therapy could be due to a worsening of symptoms or merely a lack of any improvement. Considering that it is thought up to 10% of patients deteriorate when receiving psychological treatment (Hardy et al., 2017; Lambert, 2013), when using the most extreme estimates it can be surmised that over 50% may fail to experience any change in their depression symptoms (Cahill, Barkham, & Stiles, 2010). These 'stasis' outcomes not only leave patients in a state of distress, but failure to experience any meaningful benefit might yield negative attitudes about seeking future treatment (Meltzer et al., 2003; Ten Have et al., 2010).

The considerable rate of stasis outcomes highlights the potential to improve the quality of extant evidence-based depression interventions and stresses that much more could be done to improve treatments. Given the high demand for effective and efficient treatments, associated efforts have been directed towards quality improvement in order to enhance outcomes in real world services (Lambert, 2007). Quality improvement can be defined as "the combined and unceasing efforts of everyone – healthcare professionals, patients and their families, researchers, payers, planners and educators – to make the changes that will lead to better patient outcomes (health), better system performance (care) and better professional development (learning)" (Batalden & Davidoff, 2007, p. 2). Quality improvement evidence has shown improving the quality of depression care through service and therapy quality improvement interventions can produce better outcomes for patients and this evidence will now be reviewed.

5.1.2 Quality improvement techniques in a psychotherapy context

5.1.2.1 Service quality improvement interventions

5.1.2.1.1 'Quality assurance' approaches

Quality assurance methods are indirect service delivery level quality improvement strategies that have focused on establishing and ensuring the implementation of best practice in routine care. Outcomes following treatment can be up to three times lower in routine services than in clinical trials (Barkham et al., 2008; Gibbons, Wiltsey Stirman, Derubeis, Newman, & Beck, 2013; Hansen et al., 2002). The difference stems, at least in part, from the different conditions that treatment is delivered in (Barkham et al., 2008), with research trials characterised by stringent internal validity and high levels of therapist supervision and treatment fidelity checks (Rothwell, 2005). However, in naturalistic settings, typical treatment delivery often falls short of required standards of care and the monitoring of treatment integrity is much more haphazard (Hansen et al., 2002). Patients can have difficulty accessing suitable evidence-based therapy (Wang et al., 2005), with therapists who have received adequate training in the most effective treatment methods (Beidas & Kendall, 2010). Even when they do, it is common for the 'dose' of treatment to be insufficient (Wang et al., 2005), or for unsupported modifications to be made to the therapy, so that it is no longer delivered as intended by the evidence base (Bruijniks, Franx, & Huibers, 2018; Waller, 2009). When there is 'drift' away from the treatment protocols, outcomes have been shown to suffer (Delgadillo, Kellett, et al., 2016).

Implementation of methods to ensure better translation of evidence-based practice into routine services have helped to indirectly improve the quality of real world depression treatment. Collaborative care systems have been successful in increasing access via primary care and ensuring ongoing support for suitable depression treatment (Jaycox et al., 2003; Wells et al., 2000). Evidence-based research has informed guidelines for clinical practice, to ensure that only empirically supported treatments and the associated best practice (e.g. required number of sessions) are recommended (Moriana, Gálvez-Lara, & Corpas, 2017; National Institute for Health and Clinical

Excellence [NICE], 2018). Regular supervision and the manualising of treatments have been used as methods to facilitate adherence to protocols and ensure continued delivery of evidence-based practice (Bambling, King, Raue, Schweitzer, & Lambert, 2006; Wilson, 1996). Set up in 2008, the Improving Access to Psychological Therapies (IAPT) model in the UK is one of the best examples of a systematic implementation of quality assurance criteria to improve mental health care. The model is widely regarded as a benchmark for services, having trained over 7,000 new therapists and now delivering NICE-recommended psychological treatment for depression and anxiety to 560,000 people a year while reporting consistently high recovery rates (Clark, 2011, 2018).

5.1.2.1.2 'Outcome and progress monitoring' approaches

Outcome and progress monitoring quality improvement approaches have utilised the formal measurement and monitoring of psychotherapy outcomes to enhance outcomes. More sophisticated methods of assessing the extent to which treatments are helping patients get better have been developed to monitor outcome. Response has been operationalised to enable classification of how much improvement is present at the end of treatment on a case by case basis (e.g., reliable change index [RCI]; Jacobson & Truax, 1991). *Outcome monitoring* of pre- and post-treatment symptom scores is rare in clinical services globally, but advocates of routine outcome monitoring argue that recording outcomes and making them publicly available is a useful method for driving up recovery rates, which should be adopted worldwide (IAPT approach; Clark et al., 2017). However, monitoring outcome at the end of treatment is limited, in that the chance to intervene for patients who have completed treatment has passed. What has shown to be even more effective in enhancing outcomes is routinely *monitoring progress* on a session-by-session basis throughout the course of treatment (Wampold, 2015). Increasing awareness of how a patient is progressing enables therapists to better

recognise when someone is at risk of a poor outcome, and offers the opportunity to adjust treatment accordingly (Lambert, 2013). Several systems have utilised routine outcome monitoring as an effective process for providing patient-specific feedback to therapists (and/or patients), with discernible improvements observed in patient outcomes (Delgadillo et al., 2018; Goldberg et al., 2016; Lambert, 2017).

5.1.2.2 Therapy quality improvement interventions

5.1.2.2.1 'Treatment enhancement' approaches

Treatment enhancements are more targeted direct quality improvement approaches that aim to enhance outcomes by improving the actual treatments for depression. Even in the optimum conditions of clinical trials, outcomes show that not all patients experience a benefit from treatments (Hansen et al., 2002). Evidently, as recovery rates do not come near to 100%, there is room for improvement. Treatment enhancements have been approached in a variety of ways. There has been a tendency to develop new treatment approaches that combine or target different processes of change, such as the 'third-wave therapies' (Dimidjian et al., 2016). Alternatively, there has been a focus on quality improvement through using pharmacological and technological adjuncts to treatment. Pharmacological adjuncts, such as the delivery of psychological therapies in combination with standard antidepressant medication, have received considerable attention, but with varied results (Cuijpers, Ven Straten, Warmerdam, & Andersson, 2009; Thase et al., 1997). Recently, sophisticated methods in the treatment of anxiety have used neuroscience evidence of brain processes to successfully enhance specific elements of psychological therapies with pharmacological intervention (Craske, Hermans, & Vervliet, 2018).

Technological advances have resulted in a growing interest in using technology as an adjunct to psychotherapy, again with varied but generally positive effects on outcomes. For example, online platforms (Ahern, Kinsella, & Semkovska, 2018), smart

phone applications (Ly et al., 2015), and automatic text messaging (Aguilera, Bruehlman-Senecal, Demasi, & Avila, 2017) have all been used to support and enhance face-to-face treatment techniques. Less common are examples of simple, low-cost therapy enhancements or adjuncts. Predictors of outcome from treatment data have been used to produce patients suitability guidelines for a group cognitive behavioural treatment (CBT), which improved subsequent group outcomes (Kellett, Newman, Matthews, & Swift, 2004). Meanwhile, integrating a theory-informed action plan adjunct prior to treatment has been effective at increasing psychotherapy attendance (Avishai, Oldham, Kellett, & Sheeran, 2018). Overall, the data- or theory-informed strategies that have been successfully implemented have generally taken the form of an adjunct to treatment, rather than the specific enhancement of treatment content.

5.1.3 Where next for quality improvement for psychological therapy for depression?

The concept of quality improvement is clearly important for both patients and services, but could more be done? Ensuring the translation of evidence-based research into clinical practice through quality assurance is essential. There is a body of research focusing on the issues of quality assurance implementation in routine services. Similarly, the idea of outcome and progress monitoring is well-established and significant work has been dedicated to implementing change. Yet these types of organisational level approaches often require whole systematic service and culture changes to see meaningful sustained impact (Lambert, 2007; Von Korff & Goldberg, 2001). This takes time, resources and the backing of funders and organisations (such as with the IAPT model; Layard & Clark, 2014). In the meantime, more efforts should be directed towards trying to find additional, easily implemented and low-cost therapy strategies to enhance depression outcomes.

Treatment enhancements provide an opportunity for simpler quality improvement implementation. However, research into improving psychological treatments through therapy quality improvement interventions has been vastly underfunded, especially in comparison to pharmacological treatments ("Therapy deficit," 2012). In their review of methods for studying quality improvement, Portela et al. (2015) noted that the healthcare improvement literature is largely made up of commentaries or narrative reviews describing what quality improvement interventions could do. This gap is reflected in the limited available empirical studies on the implementation and testing of quality interventions specific to enhancing psychological therapies for depression. Generally, enhancement attempts have looked to develop new treatments or adjunct therapies with technology. Not enough attention has been paid to improving the quality of the existing interventions. Whilst a range of clinically effective therapies have been developed, less effort has been applied to then shaping and enhancing the interventions (and testing whether outcomes improve; i.e. making a therapy that works, work better). The preference for developing new treatments comes with high 'financial' and 'time' costs that at best, produce small improvements. Few quality improvement enhancements have utilised theory-based augmentations aimed at increasing the salience of what we already know works and integrated them directly into existing therapy. By using strategies that are informed by research and are theoretically coherent with an extant intervention, therapy augmentations could be easily integrated into routine practice. Time spent integrating simple, cost-effective therapy enhancements could therefore produce effective outcome improvement at a significantly reduced cost (Oldham, Kellett, Miles, & Sheeran, 2012).

5.1.4 Quality improvement integrated into treatment of depression: The case of BAG

A growing empirical evidence base supports behavioural activation (BA) as one of the most effective treatments for depression (Ekers et al., 2014; Richards et al., 2016). BA is based on simple and parsimonious principles that can be effectively disseminated via an organisationally efficient group format (Kellett et al., 2017; Sturmey, 2009). The demand for depression treatment has never been higher and is only expected to increase in the coming years (World Health Organisation, 2017). BA delivered in groups (BAG) therefore represents a valuable treatment option for depression. So, can the effectiveness of group BA be improved through a quality improvement treatment enhancement?

The relatively simple and pragmatic nature of behavioural interventions means that BA appears well suited to modification and additions to treatment without affecting the integrity of the intervention. Moreover, BAG has a treatment protocol therefore augmentation with a treatment enhancement approach is easier than for concept driven interventions. Integrated treatment enhancements need to be targeted at the key barriers and facilitators of change (van Bokhoven, Kok, & van der Weijden, 2003). Applicable theory-driven strategies can then be chosen that are likely to influence those factors that either obstruct, or facilitate the desired change (Avishai et al., 2018). Applying this concept to behavioural interventions, *treatment dose* and *treatment compliance* are two crucial ingredients for a positive outcome, with the potential to be receptive to theory-informed BAG treatment enhancement. For both of these therapy processes, outlined below are the a) link to outcome, b) theoretical perspective of influencing factors and c) a theory-driven quality improvement strategy to bring about change in the form of a treatment augmentation.

5.1.4.1 Treatment dose focused BAG augmentation targeted at attendance

5.1.4.1.1 Relationship between attendance and BA outcome

The dose-response evidence base demonstrates that attendance at a minimum number of sessions is crucial to psychotherapy outcome in routine practice (Cahill et al., 2003; Delgadillo et al., 2014; Hansen et al., 2002). Much like for medication, patients need to receive an adequate 'dose' of therapy to facilitate change. Termination of therapy before receiving an appropriate dose places patients at risk of experiencing a stasis outcome (Cahill et al., 2003). However, the amount of psychotherapy that is delivered in naturalistic settings often falls short of the recommended amount (Hansen et al., 2002; Kessler et al., 2003; Wang et al., 2005). One reason for insufficient treatment doses is due to patients dropping out of therapy and failing to finish a course of treatment (Barrett et al., 2008). Treatment non-completers are known to have worse outcomes than treatment completers (Cahill et al., 2003). Therefore, retaining patients in therapy for longer could improve outcomes, thus indirectly reducing the rate of stasis (Page & Hooke, 2009).

5.1.4.1.2 Patient expectations as a predictor of attendance

Expectations about rate of improvement and required number of sessions is one reason that has been proposed to explain why some patients drop-out of treatment prematurely (Barrett et al., 2008; Swift et al., 2012). Attendance is predicted by patient therapy expectations, with patients rarely attending more sessions than they initially expected to (Callahan, Aubuchon-Endsley, Borja, & Swift, 2009; Scamardo, Bobele, & Biever, 2004). It follows that discrepant expectations are associated with drop-out (Hansen, Hoogduin, Schaap, & de Haan, 1992). The reason for discrepancies between expectations and reality appears to be that many patients underestimate the necessary dose of therapy required to experience a significant improvement (Swift & Callahan, 2008). Consequently, when patients' expectations for therapy duration are unrealistic and fall short of the adequate dose, they are more likely to drop out of treatment (Mueller & Pekarik, 2000; Tryon, 1999). This pattern would suggest that aligning

patient expectations with the realities of therapy using dose-response evidence could improve treatment completion.

5.1.4.1.3 Expectation management techniques

Studies have demonstrated that patient expectations can be manipulated using low-cost expectation management techniques. Role induction strategies to prime patients' expectations about the treatment rationale, what the therapy will involve, and what their role as the patient will entail have improved patient expectations (Constantino, Ametrano, & Greenberg, 2012) and treatment completion rates (Delgadillo & Groom, 2017). However, the impact of strategies to foster realistic dose-response expectations about rates of treatment response on drop-out rates is relatively under-investigated and the evidence inconclusive. When patients' estimates and expectations for treatment duration were discussed with their therapist as a component of therapy, drop-out rates were no different to those who had not engaged in such discussions (Reis & Brown, 2006). However, the results are potentially undermined by a realistic expected therapy duration set at three sessions or more, which is generally less than would be expected for recovery.

Other efforts have used psychoeducation materials informed by the dose-response evidence base to prepare patient expectations about the required number of sessions needed for improvement. Dose-response informed education brought patient pre-therapy expectations about duration more in line with a dose that would invoke meaningful symptom improvement (Swift & Callahan, 2008). The subsequent impact on actual attendance has seen inconsistent results. Delgadillo, Moreea, Murphy, Ali and Swift (2015) produced orientation leaflets based on theory and dose-response evidence specific to the intervention being delivered, to influence a variety of treatment expectations (including duration for improvement). The leaflets were posted to patients as part of a treatment information pack prior to attending low intensity guided self-help

interventions, but had no eventual effect on attendance. Swift and Callahan (2011), on the other hand, demonstrated that patients who received dose-response information, based on Hansen et al.'s (2002) estimate of 13-18 sessions for 50% patients to improve, remained in therapy significantly longer. Both these studies employed psychoeducation prior to patients attending for treatment, rather than as a treatment component. Whether patients properly read and processed the information and then remembered it during therapy is difficult to establish. Patients receiving education in the latter study were also asked to indicate the number of sessions they expected to attend as part of the process, which may have ensured increased dose-response comprehension and been a contributing factor to the more positive results.

5.1.4.1.4 Psychoeducation BAG augmentation for increasing attendance

The existing literature has utilised attendance expectation management as a method for providing realistic expectations, with mixed success in terms of influencing drop-out rates. When the expectation management strategy was integrated into the therapy, the validity of the dose-response information to the therapy being received was questionable. On the other hand, the strategies that have used evidence-based psychoeducation have all been administered prior to therapy where sufficient patient comprehension is hard to guarantee. Combining these two strategies through 1) adopting psychoeducation as a treatment component incorporated directly into the BAG protocol, and 2) using 'bottom-up' dose-response information informed by practice-based evidence from BAG (i.e., the actual intervention being delivered), might increase the salience of the message. The implication would be that if patients are better aware of the required dose-response, those who do not experience immediate improvement might feel less disillusioned and continue to persevere attending sessions. As increased attendance is related to better outcomes, BAG stasis outcomes could be indirectly targeted by an enhancement that reduces drop-out.

5.1.4.2 Treatment compliance focused BAG augmentation targeted at stasis

5.1.4.2.1 Relationship between treatment compliance and BA outcome

Treatment compliance also plays an important role in terms of patient response to BA. Behavioural interventions are distinctive in that the focus is on 'doing', rather than the traditional 'talking therapies' approach (Kanter et al., 2010). Patients are tasked with between-session work to put the activation concepts into practice (i.e., change their behaviour). As a result, the majority of the fundamental work to produce change in BA occurs within this between-session activity scheduling homework (Hopko et al., 2011). Clinical outcomes are predicted by the amount of completed between-session tasks (Rees, McEvoy, & Nathan, 2005). Patients' engagement and implementation of the activities is therefore crucial to achieving their goals and experiencing a positive outcome (Addis & Jacobson, 2000; Burns & Spangler, 2000; Kazantzis, Whittington, & Dattilio, 2010). In turn, failure to implement the activation strategies has been proposed as a contributing factor to non-response in BA (Hopko et al., 2011).

5.1.4.2.2 Intention-behaviour gap in depression

It is a common phenomenon for individuals to set goals, but to have difficulty implementing them. The procrastination and avoidance of depression only adds to this difficulty (Krämer, Helmes, Seelig, Fuchs, & Bengel, 2014). In the case of an individual with depression receiving BA treatment, they may set a between-session goal to increase their activation. Despite having definite intentions to strive for their goal, they might struggle to act on their intention in-between sessions. In social psychology, this is termed the 'intention-behaviour gap' (Sheeran & Webb, 2016). Intentions are known to predict behaviour (Sheeran, 2002). Merely setting a goal (e.g., increase activation) is rarely sufficient without also forming intentions for how to strive for the goal (e.g., 'I intend to go for a walk'). However, even high-intenders can experience trouble engaging in actions consistent with their intentions, demonstrating behavioural

intentions in isolation do not guarantee successful goal pursuit (Fife-Schaw, Sheeran, & Norman, 2007; Webb & Sheeran, 2006).

There are several reasons why difficulties putting intentions into action may occur (Sheeran & Webb, 2016). First, good opportunities for action might be missed or forgotten about, meaning that people struggle to get started in the first place. Second, barriers might derail the initiation of behaviour meaning people fail to stay on track and attain their goal. In depression, these difficulties are compounded by insufficient planning and maintenance self-efficacy skills. People with depression have been shown to produce fewer defined action plans and be more distracted by obstacles to their intentions than non-depressed individuals, yet fail to employ more coping planning strategies to compensate (Krämer et al., 2014). Given these amplified difficulties in depression, specific pre-planning and self-regulation techniques might be useful for increasing treatment compliance. Potential strategies can be drawn from behaviour change research into ways to help close the intention-behaviour gap. One such technique that has been developed, called 'implementation intentions', has proved effective for a variety of health behaviours.

5.1.4.2.3 'Implementation intentions' for facilitating goal attainment

Implementation intentions are specific plans about how, when and where goals will be acted upon, formed using an 'if-then' format (Gollwitzer, 1999). By forming an 'if-then' plan in advance, the 'if' defines a cue for behaviour initiation or an anticipated barrier that might inhibit action and the 'then' plans an appropriate response that aligns with the intended goal. For example, an 'if-then' plan for someone who is struggling to get out of bed in the morning might be: 'If my alarm goes off, then I will immediately get up and make a cup of tea'. Thus, by attaching intentions to the goal, progression from goal setting to goal striving is clearly defined.

Implementation intentions help to close the intention-behaviour gap by making processing mechanisms more efficient. The pre-planning builds a link between the environmental cue and the goal oriented response providing protection from internal or external interferences (Parks–Stamm, Gollwitzer, & Oettingen, 2007). Thereby, removing distracting decisions, the need for deliberation and enabling immediate action. Opportunities to act, or barriers which may interfere are more easily identified and can be used as triggers for the implementation of behaviour (Webb & Sheeran, 2008). Consequently, control over what action is selected is given over to contextual cues making the process of goal striving happen more automatically (Bayer, Achtziger, Gollwitzer, & Moskowitz, 2009). Spur of the moment decisions that are often detrimental to goal striving behaviours therefore become less likely.

The evidence for implementation intentions facilitating goal attainment is widespread in the field of health behaviours (Gollwitzer & Sheeran, 2006). Effective implementation of 'if-then' plans have been seen across numerous behavioural domains, including increased attendance at cervical screenings (Sheeran & Orbell, 2000), better medication adherence (I. Brown, Sheeran, & Reuber, 2009), and instigating healthier lifestyles, such as becoming more active (Arbour & Martin Ginis, 2009) or eating more fruit and vegetables (Kellar & Abraham, 2005). Application to clinical psychology is still relatively recent, but there is empirical support that implementation intentions can also be effective for people with mental health disorders. Toli, Webb & Hardy's (2016) meta-analysis demonstrated the beneficial effect of 'if-then' planning in facilitating goal achievement across various mental health problems where difficulties in putting plans into action are likely to be even more pronounced.

'If-then' plans have also been effective in increasing attendance at psychotherapy sessions (Sheeran, Aubrey, & Kellett, 2007), implementing self-help relaxation exercises (Varley, Webb, & Sheeran, 2011), and helping with attentional

control in social anxiety (Webb, Ononaiye, Sheeran, Reidy, & Lavda, 2010). Specific to depression, an intervention focused on forming 'if-then' plans doubled the rate of personal activity-related goal attainment and appeared to reduce depression symptoms (Fritzsche, Schlier, Oettingen, & Lincoln, 2016). More recently, the development of a new depression relapse prevention self-help intervention based on implementation intentions was feasible and acceptable to patients (Lucock et al., 2018). These studies have shown the promise for the implementation intentions technique as an adjunct to activation-based interventions, but they have yet to incorporated into existing depression treatment protocols.

5.1.4.2.4 The potential role of implementation intentions in reducing stasis

As a key aspect of psychotherapy is working towards goals in order to achieve outcomes, systematic integration of implementation intentions fits well into the process of setting between-session work (Toli et al., 2016). 'If-then' planning as a facilitator of behaviour change has a particular relevance to the BA approach, as the intervention requires patients to change their behaviour in order to experience change in their depression symptoms. Drawing on the empirical and theoretical evidence, integrating repeated, systematic 'if-then' plans into the process of setting between-session work at the end of each BAG session is a 'top-down' strategy that could facilitate behaviour change between-sessions. The evidence for 'if-then' plans suggests patients will be more likely to complete their between-session activation strategies if they forge a cue for action from their contextual surroundings. Given the association between homework compliance and outcomes, the effects of BAG could be enhanced as a result. Knock-on effects would be likely to be that fewer patients experience a depression stasis outcome.

5.1.5 Objective of the current study

To summarise, there is clearly a need to continue to develop and enhance psychological therapies for depression, evidenced by the rates of stasis outcomes for

routine clinical practice. While quality improvement approaches have been employed to increase outcomes after depression treatment, there is a paucity of theory-driven augmentations of the existing treatments. Low-cost, targeted augmentations, driven by underlying theory could be used to enhance extant evidence-based approaches. BAG is an effective and parsimonious treatment for depression. The ease and efficiency of treatment dissemination makes it an attractive option for clinical services and therefore BAG an ideal candidate for enhancement through treatment augmentation. Treatment dose and treatment compliance are two processes that are associated with BAG therapy response. When these two processes fall short of the required standards, they are associated with drop-out and stasis outcomes respectively. Theory-informed mechanisms therefore need to be targeted at preventing these unwanted outcomes. Managing patient expectations is a mechanism associated with treatment attendance. Using practice-based BAG data to produce dose-response psychoeducation integrated into the therapy could be used as a 'bottom-up' quality improvement augmentation to target BAG attendance. Meanwhile, increasing the amount of completed between-session work is a mechanism associated with depression symptom reduction. Implementation intention exercises incorporated into the process of setting between-session activities could be used as a 'top-down' quality improvement augmentation to target stasis.

5.1.5.1 Aims and hypotheses

The aim of the study was to test the effect of an enhanced version of BAG, augmented with two theory-informed quality improvement strategies. The study aimed to compare the augmented intervention with the existing BAG treatment on attendance, overall symptomology and depression stasis outcomes. It was hypothesised that in comparison to the existing BAG treatment, (1) a psychoeducation treatment augmentation would result in greater attendance rates at BAG, and (2) an

implementation intentions treatment augmentation would produce greater reductions in depression, anxiety and impaired functioning, and (3) fewer depression stasis outcomes after BAG.

5.2 Method

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*). Information and evidence about ethical approval can be found in Appendix C (research protocol, ethical approval confirmation).

5.2.1 Design

A cohort comparison design was employed to compare routine delivery of the existing BAG intervention (BAG-) with an augmented BAG intervention (BAG+). Patients recruited into the study received the BAG+ intervention. Retrospective anonymised routine outcome data from patients who had previously received BAG treatment (N=178; collected in study 1) were utilised as a historical control to represent BAG-. The BAG- sample was selected by matching patients in the historical dataset to the BAG+ sample on variables known to predict treatment outcome, using propensity score matching (PSM). PSM enables observational data to mimic the features of a RCT by balancing pre-treatment covariates in the experimental and control groups (Austin, 2011a).

5.2.2 Participants

5.2.2.1 Sample size

A sample size analysis using G*Power (Faul et al., 2007) indicated 27 patients would be needed in each group (total N=54) to detect a small to medium effect size

(*d*=0.3) with .80 power using a repeated measures, between-subjects ANOVA at p=0.10. The alpha value was set to 0.10 to represent the use of one-tailed tests due to the direction of the hypotheses being clearly specified in favour of BAG+.

5.2.2.2 Eligibility criteria

A STROBE diagram of patient flow through the dataset and sample selection method is presented in Figure 5.1. All patients who attended BAG+ in a UK Improving Access to Psychological Therapies (IAPT) service between January and December 2017 were invited to have their outcomes included in the study. Inclusion criteria were: (a) seeking treatment for a primary presenting problem of depression; b) referred following assessment by Psychological Wellbeing Practitioners (PWP); (c) attended at least one treatment session of a course of BAG; (d) at least 18 years old; and (e) gave informed consent for their data to be used in the study. As it was a study of routine practice, exclusion criteria were minimal to ensure a representative sample. The only exclusion criterion applied was not meeting criteria for depression caseness prior to commencing BAG (PHQ-9 score ≥10). Out of 34 patients who attended BAG+ across three groups, 31 met the criteria and had their outcomes included in the study (*N*=3 excluded due to not meeting depression caseness at the start of treatment).

The same inclusion and exclusion criteria were applied to the archived data of existing BAG delivered between November 2014 and December 2016, with the exception of informed consent (as the data was anonymised, and therefore the patients could not be contacted retrospectively). Out of 178 patients who had attended BAG- across 22 groups, 161 met the inclusion criteria (*N*=17 excluded due to not meeting caseness criteria). From the available pool of 161 BAG- patients, 31 were matched to the 31 eligible BAG+ patients to ensure equivalent groups in the final sample (see section 5.2.5.1.1. for full description of matching procedure).

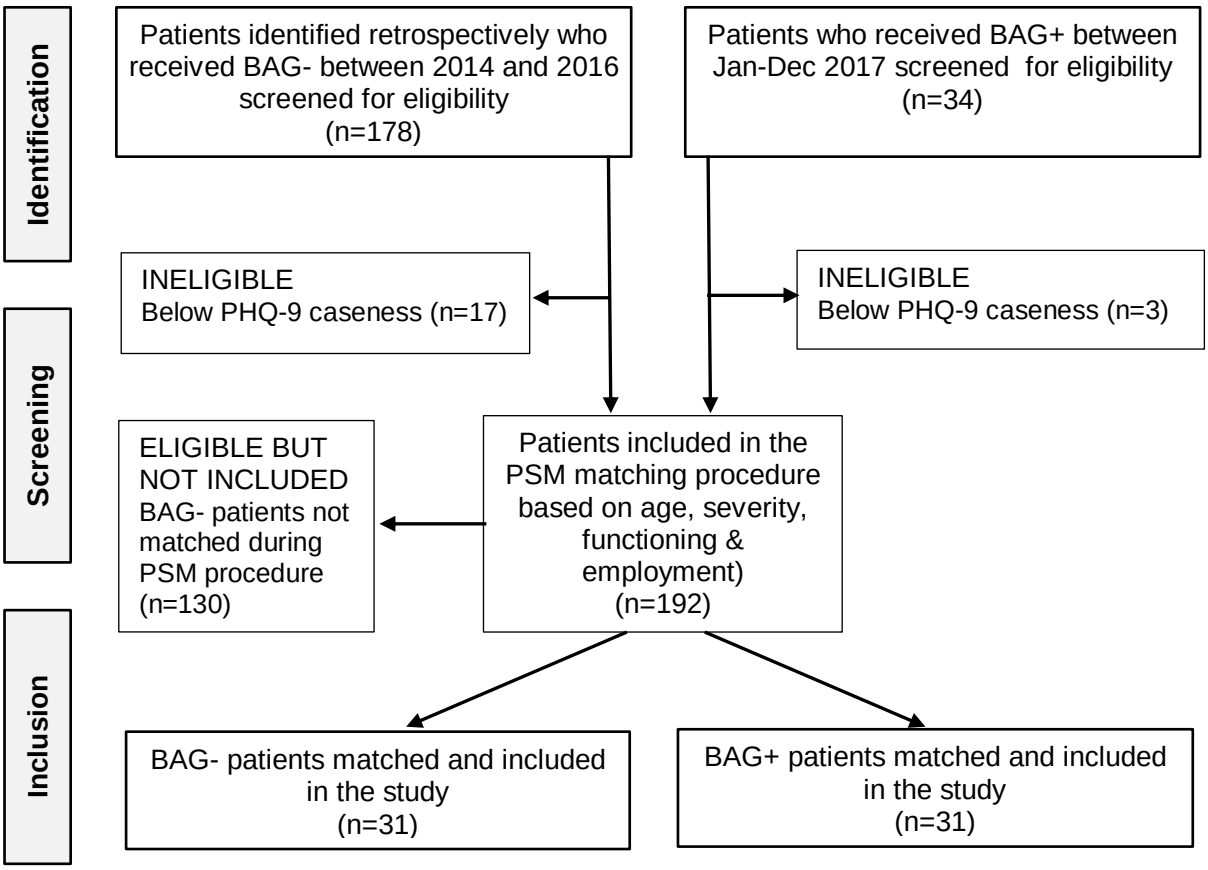


Figure 5.1. STROBE flow diagram of patient selection.

5.2.3 Outcome measures

5.2.3.1 IAPT minimum dataset

The outcome measures consisted of the IAPT minimum dataset (Patient Health Questionnaire-9 [PHQ-9], Generalised Anxiety Disorder-7 [GAD-7] and Work and Social Adjustment Scale [WSAS]). Copies of the outcome measures can be found in Appendix D. These measures are completed at every contact as part of IAPT routine outcome monitoring (see Chapter four, section 4.2.4 for full description of measures previously provided).

5.2.3.2 Demographic information sheet

For the purpose of this study, patients recruited to receive BAG+ were asked to complete a demographic information sheet to capture information about demographics

(age, gender, ethnicity), current antidepressant medication and previous episodes of depression and treatment. The demographic information sheet was not able to be administered to the BAG- sample as the data were accessed from an archived database.

5.2.4 Procedure

5.2.4.1 BAG+ procedure

From January to December 2017, an augmented version of the existing BAG treatment (BAG+) was delivered to all patients referred to BAG therapy. Patients were sent an information pack a week before the group start date, which included group information and a patient information sheet to inform patients about the study prior to attending treatment. At the first BAG+ session, the group facilitators administered the demographic information sheet and obtained informed consent from patients who agreed to have their routine outcomes included in the study. All patients received the same BAG+ treatment regardless of whether they gave consent. As per IAPT protocol, the IAPT minimum dataset was administered at every session and outcomes for consenting patients were passed to the researcher.

5.2.4.2 BAG- procedure

An archived dataset consisting of all patients who received the existing BAG therapy (BAG-) up until December 2016 was accessed. The dataset contained anonymised demographic information (age, gender, ethnicity, index of multiple deprivation [IMD] decile and rank), treatment information (session dates, therapist initials and number of sessions attended) and routine outcomes (PHQ-9, GAD-7 and WSAS) for every BAG session each patient attended. Patients accessing IAPT services are informed that their treatment outcomes may also be used and shared in secondary analyses of treatment delivery and response, but that all data are anonymised and summarized so that it is impossible for any individual patients to be identified (IAPT, 2011).

5.2.4.3 BAG intervention

Both versions of BAG (BAG- and BAG+) were delivered as a step 3 high intensity intervention in the IAPT stepped care model. The intervention was based on the 10-session group protocol developed by Houghton et al (2008) and was adapted for use in Primary Care. BAG consisted of eight two-hour sessions delivered on a weekly basis for 8 weeks. Facilitators used a treatment manual to guide treatment and patients were given a workbook. Patients were given between-session work to complete, which was fed back and reviewed at each session. Each BAG session was based on a different topic relevant to the principles of BA to encourage increased participation in rewarding personally meaningful activities (see Table 5.1).

Table 5.1. *Session outlines for BAG- and BAG+ interventions describing the common BAG- and BAG+ treatment components and additional BAG+ components*

Session	Title	Content (common components across BAG- and BAG+)	Additional BAG+ components (Dose-response psychoeducation augmentation; Implementation intentions augmentation)
1	Learn your patterns and start to change them	Context and symptoms of depression Depression cycle (do less – feel worse/feel worse – do less) Homework: Activity-mood diary	Psychoeducation sheet in workbook. Prompt in treatment manual for facilitators to verbally reiterate the information. Additional homework task: Read psychoeducation sheet 'Achieving your goals' information sheet in workbook 'If-then' plans modelled by facilitators Session specific 'if-then plan' worksheet in workbook to plan and set HW 'If-then' plan silently repeated 3 times and once out loud to partner
2	Values: the guide to who we are	Overcoming mood dependence – concept of 'outside in' Link between values and BA Identify imbalances in action and values using VLQ Homework: Committed-action exercise	Session specific 'if-then plan' worksheet in workbook to plan and set HW 'If-then' plan silently repeated 3 times and once out loud to partner
3	Getting out of the TRAPs and back on TRAC	Common problems in depression Identify TRAPs (trigger, response, avoidance pattern) Develop TRACs (trigger, response, alternative coping) Homework: Apply TRAP/TRAC handouts to tasks	Session specific 'if-then plan' worksheet in workbook to plan and set HW 'If-then' plan silently repeated 3 times and once out loud to partner

4	Taking action: a problem solving approach	'Problem solving approach' for changing behaviour Complete 8-step problem solving handout Homework: Problem solving to change unhelpful behaviours	Session specific 'if-then plan' worksheet in workbook to plan and set HW 'If-then' plan silently repeated 3 times and once out loud to partner
5	Identifying unhelpful thinking, worry and rumination	Difference between thinking, worry and rumination 'Thoughts on a tissue' exercise with group Effect of avoidance, rumination and acceptance on mood Homework: Monitor rumination and use 'two-minute rule'	Session specific 'if-then plan' worksheet in workbook to plan and set HW 'If-then' plan silently repeated 3 times and once out loud to partner
6	Developing responses to thinking, rumination and worry	Techniques for dealing with rumination Rumination cues action (RCA), mindfulness, serenity prayer, self-soothing Homework: RCA, mindfulness and self-soothing handouts	Session specific 'if-then plan' worksheet in workbook to plan and set HW 'If-then' plan silently repeated 3 times and once out loud to partner
7	Making changes one step at a time	Barriers to change Coping with physical symptoms of depression Physical activity exercise Homework: 'Short-term goals' planning worksheet	Session specific 'if-then plan' worksheet in workbook to plan and set HW 'If-then' plan silently repeated 3 times and once out loud to partner
8	Building the relationships you want/tying it all together	Relationships in context of barriers to activation Review content of whole group – ACTION acronym Homework: Apply ACTION to everyday situations	Session specific 'if-then plan' worksheet in workbook to plan and set HW 'If-then' plan silently repeated 3 times and once out loud to partner

Note: Session order was updated during the course of routine delivery between 2014-2018. All patients received the same content, but some sessions were delivered in a different order for the older BAG groups (values at session 4 instead of session 2); HW; Homework task, VLQ; Valued Living Questionnaire, ACTION; assess, choose, try, integrate, observe, never give up.

5.2.4.3.1 Augmented BAG+ delivery

The BAG protocol was enhanced with two augmentations to produce a BAG+ version of the intervention. The treatment augmentations were developed in conjunction with BAG facilitators to ensure the material was suitable for patients and the intervention being delivered.

1. Dose-response psychoeducation augmentation

The first augmentation was a 'bottom-up' data-informed psychoeducation enhancement targeted at poor attendance. The psychoeducation consisted of attendance outcomes taken from the pilot study conducted on a subset of outcome data from existing BAG delivery between 2009-2011 (N=73) (Kellett et al., 2017). The

psychoeducation therefore highlighted the relationship between attendance and outcome. The psychoeducational information sheet (see Appendix E for copy of materials) was included in the pre-treatment information pack informing patients that:

1) Attendance at least 4 BAG sessions was required to support change in depression symptoms
2) BAG was effective regardless of the severity of depression.
3) BAG was also effective at reducing co-existing anxiety symptoms.

In addition, BAG facilitators verbally reiterated the information as part of the session content for the first session (on the agenda), and the psychoeducation information sheet was included in the patient workbook. Patients were asked to read through the information sheet as part of the between-session work from session one.

2. Implementation intentions augmentation

The second augmentation was a 'top-down' theoretically informed enhancement to target stasis outcomes. 'Implementation intentions' were integrated into the process for setting between-session work at the end of each session. An 'Achieving Your Goals' information sheet was included in the patient workbooks explaining the 'if-then' plans process instructions (see Appendix F for copies of materials). 'If-then' plans were introduced and modelled by the facilitators at the end of the first session. 'If-then' planning worksheets were included in the patient workbooks and a session specific 'if-then' between-session plan example provided for every session. Patients made 'if-then' plans using the worksheet for how each of the between-session goals would be implemented. The 'if' section directed patients to identify a potential obstacle or a good opportunity to act on the intention. The 'then' section directed patients to choose a suitable response/action to the identified opportunity oriented towards completing their desired goal. Patients silently repeated their intention to themselves three times and then repeated it once out loud to a partner to verbally commit to the homework. At every

subsequent session, facilitators prompted patients to set between-session work using 'if-then' planning.

5.2.4.4 Facilitators

Each BAG- and BAG+ group was facilitated by two British Association for Behavioural and Cognitive Psychotherapies (BABCP) accredited CBT therapists. Facilitators were experienced one-to-one therapists and received service 'in-house training' specifically in delivering BAG. The research team delivered an additional one hour training workshop for the pool of facilitators (N=8) delivering the augmented BAG+ intervention to introduce the treatment augmentations and outline how to deliver them in practice. Figure 5.2 presents the results for facilitators' understanding and confidence in integrating implementation intentions into BAG, after attending the training workshop.

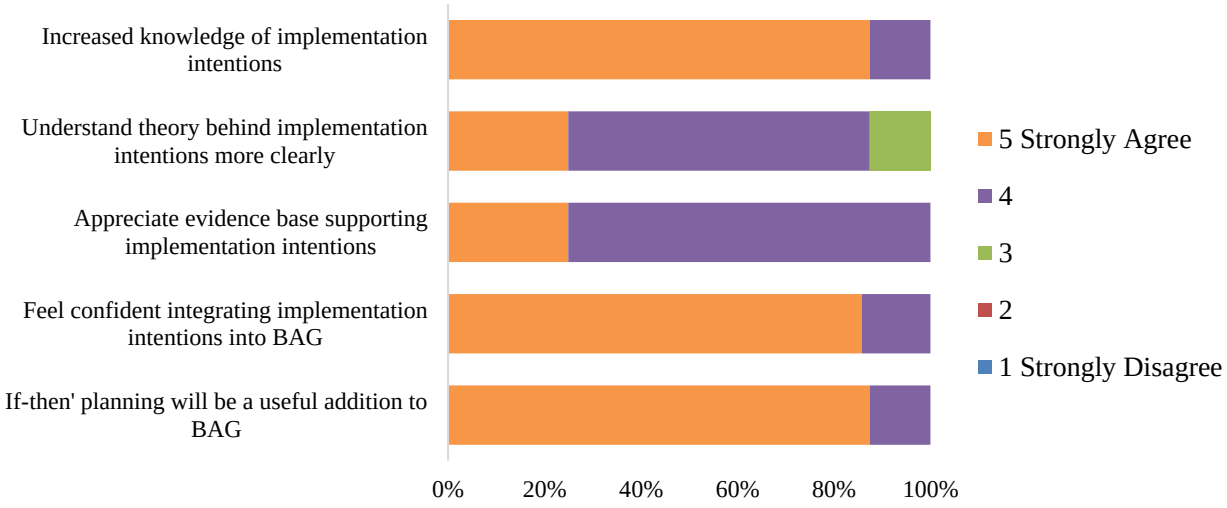


Figure 5.2. Facilitator evaluation of implementation intentions training workshop (N=8)

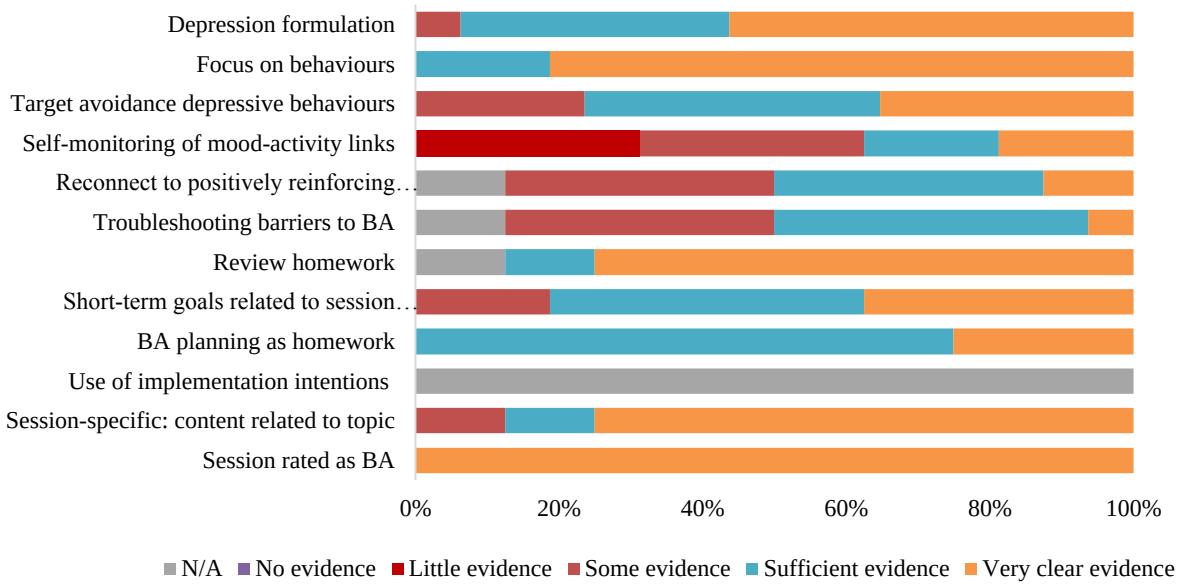
5.2.4.5 Treatment integrity

Three treatment integrity approaches were implemented. First, facilitators attended quarterly BAG working group meetings to provide peer supervision. The researcher also attended the meetings to manage any issues related to the study. Second, both BAG versions used a manualised approach to ensure fidelity to the protocol.

Finally, BAG treatment adherence was assessed using a BAG adherence checklist developed for the purpose of this study (see Appendix G for copy of adherence checklist). The checklist was adapted from an adherence check used in a previous BA trial (Ekers et al., 2011). The checklist included a *general adherence* section (split into items related to the behavioural rationale and items related to between-session work), a *session specific adherence* section, and an overall assessment of whether the session could be rated as BA. An item relating to 'use of implementation intentions' was included to check adherence to the augmentation for BAG+. A customised page of the checklist was adapted for every BAG session to distinguish aspects that would not be expected to be present, given the session content. The session specific *mood dependence* item from the BAG- checklist was changed to a general adherence item in the BAG+ checklist.

After each BAG session, the two facilitators independently completed the checklist to check self-report adherence, before reflecting on their responses together. BAG+ adherence was checked after every session delivered throughout the duration of the study. BAG- adherence was checked for the delivery the final two groups of the existing BAG protocol at end of 2016. Inter-rater reliability between group facilitators was assessed using Cohen's kappa (Cohen, 1960). Adherence agreement was k=.57 and k=.44 for BAG- and BAG+ respectively, indicating moderate agreement (Landis & Koch, 1977).

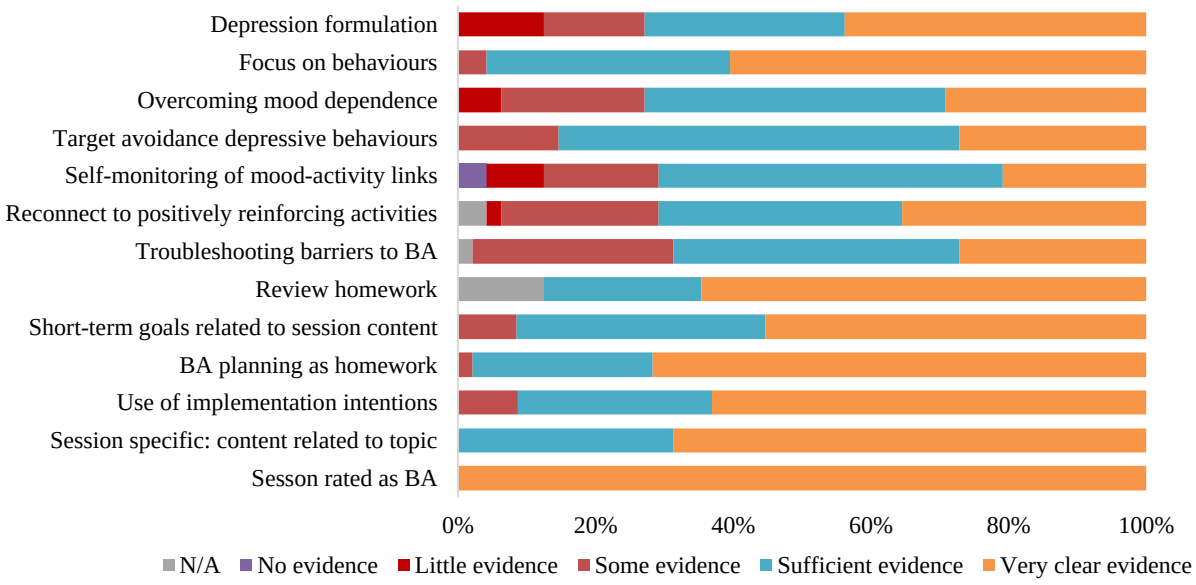
Figures 5.3 and 5.4 present the adherence check outcomes for BAG- and BAG+ respectively, summarising the mean rating for the presence of evidence in each category.



Category mean rating	12%*	0%	3%	14%	29%	44%

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

Figure 5.3. Facilitator level of self-rated adherence to the protocol during BAG-treatment delivery (N=3).



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

Figure 5.4. Facilitator level of self-rated adherence to the protocol during BAG+ treatment delivery (N=8).

All BAG- and BAG+ sessions were rated as representative of BA therapy, indicating that patients received protocol-adherent 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%). Self-monitoring of mood-activity links, reconnecting to positively reinforcing activities and troubleshooting barriers to BA were the least consistent items present in sessions for both BAG- and BAG+. Adherence checks of the implementation intentions augmentation showed they were not used in BAG- delivery (as expected), but were adequately implemented in BAG+ with sufficient or very clear evidence in over 90% of the sessions.

5.2.5 Data analysis

5.2.5.1 Preliminary analysis

5.2.5.1.1 Selection of BAG- subsample

The entire eligible BAG+ sample (*N*=31) were matched to a comparative subsample of BAG- patients (*N*=31) using propensity score matching (PSM). All patients in the existing BAG archived database who met the inclusion requirements were eligible for matching (*N*=161). The samples were matched on variables identified as key to psychotherapy outcome in primary care - age, baseline depression (PHQ-9 score), baseline functioning (WSAS score) and employment status (Delgadillo, Moreea, et al., 2016). A logistic regression model was applied to estimate propensity scores based on the identified baseline covariates. A *one-to-one, nearest neighbour* matching procedure *without replacement* was applied to match BAG+ cases to a BAG- control with a propensity score within a calliper tolerance of 0.2 (Austin, 2011b). Mean difference (standardised 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 the samples prior to and post-matching to ensure adequate matching (Austin, 2009).

5.2.5.1.2 Cluster effects of treatment delivery in groups

To assess the impact of clustering in the data, intraclass correlation coefficients (ICCs) were used to estimate the level of variance attributable to BAG group level factors. ICCs and the associated design effect (DE) for all the outcome measures were calculated using Equations 1 and 2 provided in Chapter four (see section 4.2.5.3). A DE of greater than two was used as an indication of significant co-dependence that would be unsuitable for analysis on a single-level (i.e., would require use of a multi-level model) (Muthen & Satorra, 1995). BAG treatment delivery was conducted via 13 groups in total (BAG- = 10 and BAG+ = 3). The average cluster size was 4.77. ICCs calculated for PHQ-9 (-0.04), GAD-7 (-0.05) and WSAS outcomes (0.06) therefore produced design effects of 0.85, 0.81 and 1.23 respectively. As all the DEs were less than two, single level analyses were deemed appropriate.

5.2.5.1.3 Handling missing data

Data were analysed using the intention-to-treat (ITT) principle, including all patients who entered treatment in the analysis. As outcomes were collected at every session, missing data were accounted for using last observation carried forward (LOCF) imputation. The final available outcome was used as the post score, or if there was only one score available it was assumed that there was no change. Although LOCF has documented statistical limitations (Lachin, 2016), it was deemed clinically applicable to IAPT criteria in that patients are classified as having 'received treatment' if they have attended at least one treatment session.

5.2.5.2 Statistical analysis

All statistical analyses were conducted in SPSS version 24. All analyses testing for a difference between the two interventions used one-tailed tests, as all three

hypotheses specified a clear directional prediction in favour of BAG+ (the reported p-values are the two-tailed p-values halved). To test whether the BAG+ psychoeducation augmentation produced greater attendance at therapy (hypothesis 1), attendance rates were calculated for BAG- and BAG+ based on number of sessions attended. Independent t-tests were used to compare the mean number of sessions attended and the session-by-session attendance rates. Attendees at the all eight sessions were deemed *treatment completers*, attendees at four to seven sessions were deemed *partial attenders,* and attendees at three sessions or fewer were deemed *drop-outs*. Chi-squared tests were performed on the proportions of each attendance category in BAG- compared to BAG+. Odds ratio effect sizes were calculated for significant differences in each attendance category (proportion of attendees to non-attendees in BAG- divided by the proportion of attendees to non-attendees in BAG+).

To test whether the BAG+ implementation intentions augmentation produced greater reductions in depression, anxiety and impaired functioning (hypothesis 2), total scores were calculated for the PHQ-9, GAD-7 and WSAS. Change scores were computed for each outcome for pre to post treatment (session 8 minus 1). To examine outcome change across BAG interventions, two-way mixed ANOVAs were performed on the primary and secondary outcomes, with pre-post scores as the within-group factor and BAG delivery as the between-group factor and a test of the time x BAG delivery interaction effect. Independent t-tests were performed on BAG- and BAG+ mean outcomes for the PHQ-9, GAD-7 and WSAS and for the session-by-session differences in PHQ-9 scores between groups. Within-group (pre-post changes) and between-group (BAG- versus BAG+) Cohen's *d* effect sizes were calculated to estimate the magnitude of the effects for all outcome variables (pre-treatment/BAG- score minus post-treatment/BAG+ score divided by the post-treatment/pooled standard deviation).

Cohen's *d* thresholds of 0.2, 0.5 and 0.8 were considered small, moderate, and large effect sizes respectively (Cohen, 1992).

To test whether the BAG+ implementation intentions augmentation resulted in fewer depression stasis outcomes (hypothesis 3), the proportions of individual outcomes after BAG- and BAG+ were calculated by applying reliable and clinically significant change criteria to depression outcomes (Jacobson & Truax, 1991). Reliable change is deemed to have occurred when change in patients' scores exceeds the measurement error of the measure. For the current study, the reliable change threshold calculated for use with IAPT outcome measures (IAPT, 2014) was applied to establish the recovery categories for depression outcomes defined below;

1) *Deterioration* was recorded when there was a reliable increase in PHQ-9 scores of ≥6.
2) *Improvement* was recorded when there was a reliable decrease in PHQ-9 scores of ≥6.
3) *Recovery* was recorded when there was an decrease in PHQ-9 scores of ≥6 (i.e., improvement), in addition to the score moving from above the clinical cut-off on the measure pre-treatment (≥10 on the PHQ-9) to below the clinical cut-off post-treatment (<10 on the PHQ-9; i.e., *change in caseness*). Accordingly, the 'recovery' and 'improvement' categories were not mutually exclusive.
4) A *stasis* outcome was recorded for cases where no reliable change occurred in either direction on the PHQ-9 (i.e., neither improvement or deterioration was present).

As patients below depression casesness at pre-treatment were excluded from the study, a *harm* outcome was not possible (i.e. reliable deterioration in addition to moving from below to above the clinical cut-off). To compare the rate of each outcome, chi-

squared tests were performed on the proportions of each recovery category in BAG- compared to BAG+. Odds ratio effect sizes were calculated for significant differences in each recovery category (proportion of stasis outcomes to responders in BAG- divided by the proportion of stasis outcomes to responders in BAG+).

5.3 Results

5.3.1 Preliminary analysis

5.3.1.1 PSM matching diagnostics

The matched dataset (N=62) was checked to ensure sufficient distribution of covariates across the samples in comparison to the unmatched sample (N=192). Table 5.2 presents the comparison of the means and frequencies of baseline covariates in BAG- and BAG+ patients in the overall unmatched sample and after the PSM matching. Standardised differences demonstrated that imbalance in all the specified covariates across BAG- and BAG+ were reduced to below the specified threshold (d<0.1) after matching, indicating minimal differences (Austin, 2009).

Table 5.2. *Comparison of baseline covariates in BAG- and BAG+ patients in the overall unmatched sample and after PSM matching.*

Baseline Covariate	**BAG-**	**BAG+**	**Standardised Difference/Proportion**
Unmatched sample	(N=161)	(N=31)	
Age	38.48 (16.27)	41.77 (14.88)	0.20
PHQ-9 score	17.32 (4.27)	18.42 (4.01)	0.26
WSAS score	22.67 (7.28)	24.97 (8.13)	0.32
Employment status			0.37
Employed	41 (25.5%)	13 (41.9%)	
Other	120 (74.5%)	18 (58.1%)	
Matched sample	(N=31)	(N=31)	
Age	42.61 (14.25)	41.77 (14.88)	-0.06*

PHQ-9 score	18.71(4.13)	18.42 (4.01)	-0.07*
WSAS score	24.61 (8.82)	24.97 (8.13)	0.04*
Employment status			0.00*
Employed	13 (41.9%)	13 (41.9%)	
Other	18 (58.1%)	18 (58.1%)	

Note: For continuous covariates mean and SD are presented; for categorical covariate frequencies and percentages are presented. *Standardised differences in sample covariates <0.1 deemed representative of minimal difference between groups.

Table 5.3 and Figure 5.5 present information about the variance and distribution of the continuous covariates before and after matching. After matching, imbalances in variances for age, PHQ-9 and WSAS covariates were reduced (table 5.3). Five number summaries (minimum, 25th percentile, median, 75th percentile, and maximum) of the distribution of the covariates plotted in figure 5.4 demonstrate that the sample distributions became more balanced after matching. Taken together, the matching checks suggest the PSM process was sufficiently specified.

Table 5.3. *Comparison of baseline continuous covariate variances for BAG- and BAG+ in the overall unmatched sample and after PSM matching.*

		Unmatched sample			**Matched sample**		
Covariate	**Variance (BAG+)**	**Variance (BAG-)**	**Ratio: BAG- to BAG+**	**Ratio diff.**	**Variance (BAG-)**	**Ratio: BAG- to BAG+**	**Ratio diff.**
Age	221.38	264.69	0.84	0.16	202.91	1.09	0.09
Pre-PHQ-9	16.05	18.19	0.88	0.12	17.08	0.94	0.06
Pre-WSAS	66.17	52.93	1.25	0.25	77.85	0.85	0.15

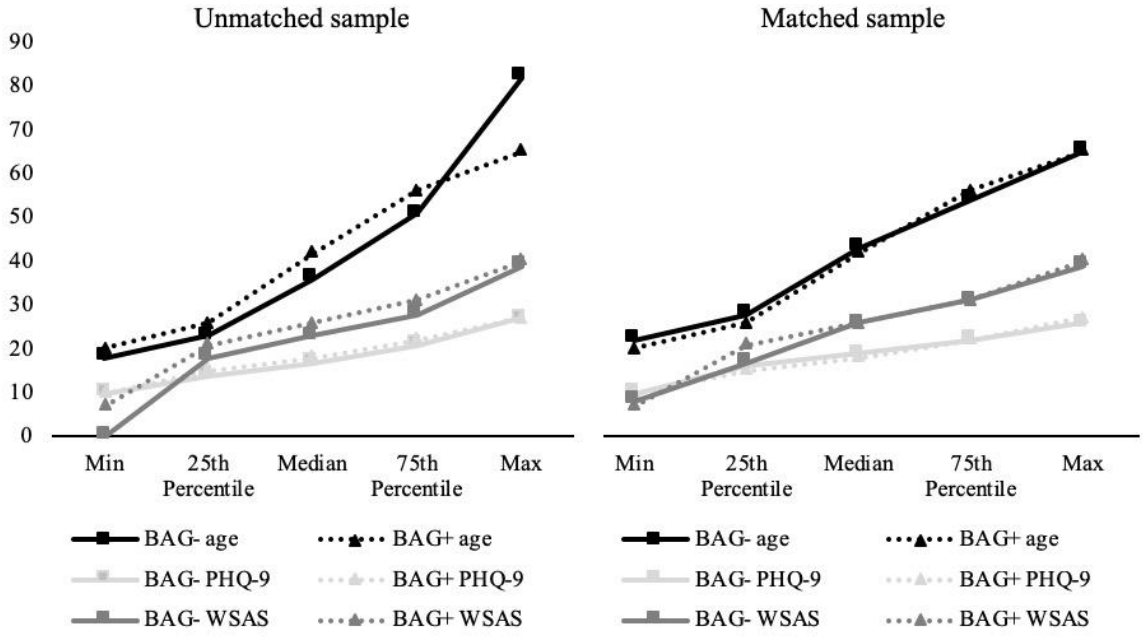


Figure 5.5. Comparison of five-number summaries for baseline continuous covariates in BAG- and BAG+ patients in the overall unmatched sample and after PSM matching.

5.3.1.2 Clinical characteristics

The final sample consisted of 62 patients, 55% (*n*=34) female and 45% (*n*=28) male. Patients ranged from 20 to 65 years old, with a mean age of 42 (SD = 14.45). The majority of the sample identified as White British (87%, *n*=54), two (3%) as Black/Black British Caribbean, one (2%) as Asian/Asian British Indian, one (2%) as Asian Other, one (2%) as mixed White and Black Caribbean, and three (5%) were not asked. Forty-two percent (n=26) were in employment, 18% (n=11) were on long-term sick leave or disabled, 18% (n=11) were unemployed, 10% (n=6) were students, 7% (n=4) were retired, 3% (n=2) were homemakers or full-time carers, and 3% (n=2) were undertaking unpaid voluntary work. Overall, 18% (n=11) were in receipt of welfare benefit payments. Pre-treatment depression severity for the total sample was classified as 37% (n=23) severe, 48% (*n*=30) moderately severe, and 15% (*n*=9) moderate depression. Nearly 89% (*n*=55) also met clinical caseness for anxiety, with 42% (*n*=26)

classified as severe, 37% (*n*=23) as moderate, 18% (*n*=11) as mild and 3% (*n*=2) as experiencing minimal anxiety.

Table 5.4. *Pre-treatment characteristics for the BAG- and BAG+ samples.*

Pre-treatment characteristic	**BAG- (n=31)**	**BAG+ (n=31)**	**X^2 (p value)**
Gender (% female)	54.8%	54.8%	0.00 (p=1.00)
Ethnicity (% White British)	87.1%	87.1%	8.00 (p=.156)
IMD deciles 1-10 (median)	5	6	11.46 (p=.246)
Depression severity			
Moderate	16.1%	12.9%	1.04 (p=.596)
Moderately severe	41.9%	54.8%	
Severe	41.9%	32.3%	
Anxiety severity			
Minimal	0%	6.5%	
Mild	16.1%	19.4%	2.29 (p=.515)
Moderate	38.7%	35.5%	
Severe	45.2%	38.7%	

Table 5.4 reports the breakdown of the pre-treatment characteristics not previously reported in PSM matching procedure across the BAG samples. Chi-square analyses showed BAG- and BAG+ samples did not significantly differ on any of the pre-treatment characteristics.

5.3.2 Hypothesis 1: BAG+ effect on attendance

To test the hypothesis (1) that the BAG+ psychoeducation augmentation would result in greater attendance compared to BAG-, the number of sessions attended were compared. The mean number of sessions attended was 4.58 (SD=2.55) and 5.16 (SD=2.40) for BAG- and BAG+ respectively (t(60) = -0.923, p=.180). Session attendance rates were slightly lower for BAG-, with the BAG+ intervention retaining

13% more patients for at least four sessions (as promoted in the BAG+ psychoeducation augmentation). However, both treatments had a sharp drop-off after seven sessions, resulting in similar rates of patients completing the full treatment. T-tests at each time-point indicated there were no significant differences in session attendance rates for BAG- and BAG+ (minimum 2 sessions: t(60) = -1.074, p=.144 minimum 3 sessions: t(60) = -0.573, p=.285; minimum 4 sessions: t(60) = -1.054, p=.148; minimum 5 sessions: t(60) = -0.514, p=.305; minimum 6 sessions: t(60) = -0.753, p=.227; minimum 7 sessions: t(60) = -0.384, p=.446; all 8 sessions: t(60) = -0.395, p=.347).

Table 5.5 reports the overall rates of treatment completers, partial attenders and drop-outs for BAG- and BAG+ delivery. No significant differences were found for the overall number of patients classified as completing treatment, partially attending treatment or dropping out of BAG- and BAG+. In summary, the addition of the psychoeducation augmentation in BAG+ did not result in significantly greater attendance at treatment.

Table 5.5. *Treatment attendance for BAG- and BAG+ (N = 62)*

Attender status	**BAG- (n = 31)**	**BAG+ (n = 31)**	**Chi-squared (p value)**	**Odds ratio (BAG+:BAG-)**
Treatment completers (8 sessions)	3 (9.7%)	4 (12.9%)	0.16 (p=.344)	1.38
Partial attenders (4-7 sessions)	15 (48.4%)	18 (58.1%)	0.58 (p=.223)	1.48
Drop-outs (1-3 sessions)	13 (41.9%)	9 (29.0%)	1.13 (p=.144)	0.57

5.3.3 Hypothesis 2: BAG+ effect on depression, anxiety and impaired functioning

To test the hypothesis (2) that the BAG+ implementation intentions augmentation will result in greater reductions in overall depression, anxiety and impaired functioning compared to BAG-, two-way mixed ANOVAs were performed on the primary and secondary outcomes. Table 5.6 presents the means and pre-post effect

sizes for every outcome within each sample, in addition to the between-treatment groups effect size and t-test statistics.

In terms of the primary outcome, an ANOVA showed depression symptoms significantly decreased following BAG treatment (pre-post main effect; F(1, 60)=45.224, p<.001). Overall depression scores did not differ significantly between interventions (BAG main effect; F(1, 60)=1.593, p=.106). However, there were significantly greater reductions in depression symptoms following BAG+ compared to BAG- (interaction effect; F(1, 60)=2.911, p=.047). Analysis of pre-post effect sizes demonstrated the post-treatment reductions in depression symptoms represented moderate and large effects for BAG- and BAG+ respectively. The lower post-treatment depression scores after BAG+ compared to BAG- were representative of a significant, small between-groups effect.

Table 5.6. *Means, standard deviations and effect sizes (d) for BAG- and augmented BAG+ on primary and secondary outcomes* (*N* = 62)

	BAG- (n = 31)	**BAG+ (n = 31)**	**Between-group *d* (BAG- vs BAG+)**	**T-score**
Primary outcome				
PHQ-9				
Pre-treatment	18.71 (4.13)	18.42 (4.00)	0.07	0.28 (p=.390)
Post-treatment	15.48 (5.18)	13.00 (6.36)	0.43	1.69 (p=.049)
Pre-post change	-3.23 (4.77)	-5.42 (5.33)	0.43	1.71 (p=.047)
Pre-post *d*	0.69 (*r*=.50)	1.10 (*r*=.55)	-	
Secondary outcomes				
GAD-7				
Pre-treatment	14.55 (4.15)	12.97 (5.01)	0.35	1.35 (p=.091)
Post-treatment	12.68 (4.90)	9.65 (5.16)	0.60	2.38 (p=.011)
Pre-post change	-1.87 (3.63)	-3.32 (4.45)	0.36	1.41 (p=.082)
Pre-post *d*	0.53 (*r*=.69)	0.75 (*r*=.62)	-	
WSAS				
Pre-treatment	24.61 (8.82)	24.97 (8.13)	-0.04	-0.17 (p=.435)

Post-treatment	20.26 (8.94)	19.23 (10.87)	0.10	0.41 (p=.342)
Pre-post change	-4.36 (6.24)	-5.93 (10.60)	0.19	0.71 (p=.240)
Pre-post *d*	0.69 (*r*=.75)	0.56 (*r*=.42)	-	

Note: Pre-post effect sizes (*d*) have been calculated taking the correlation (*r*) between pre-post scores into account (*r* coefficient reported in brackets).

Figure 5.6 displays the session-by-session PHQ-9 scores for BAG- versus BAG+. Both treatments produced early session reductions. However, while the BAG- scores plateaued across the later sessions, BAG+ scores continued to decrease. T-tests at each time-point indicated there were no significant differences in PHQ-9 scores for BAG- and BAG+ until the final session (session 1: t(60) = 0.281, p=.390; session 2: t(60) = 0.077, p=.470; session 3: t(60) = 0.210, p=.417; session 4: t(60) = 1.149, p=.128; session 5: t(60) = 0.676, p=.251; session 6: t(60) = 1.310, p=.098; session 7: t(60) = 1.362, p=.0.089; session 8: t(60) = 1.710, p=.049).

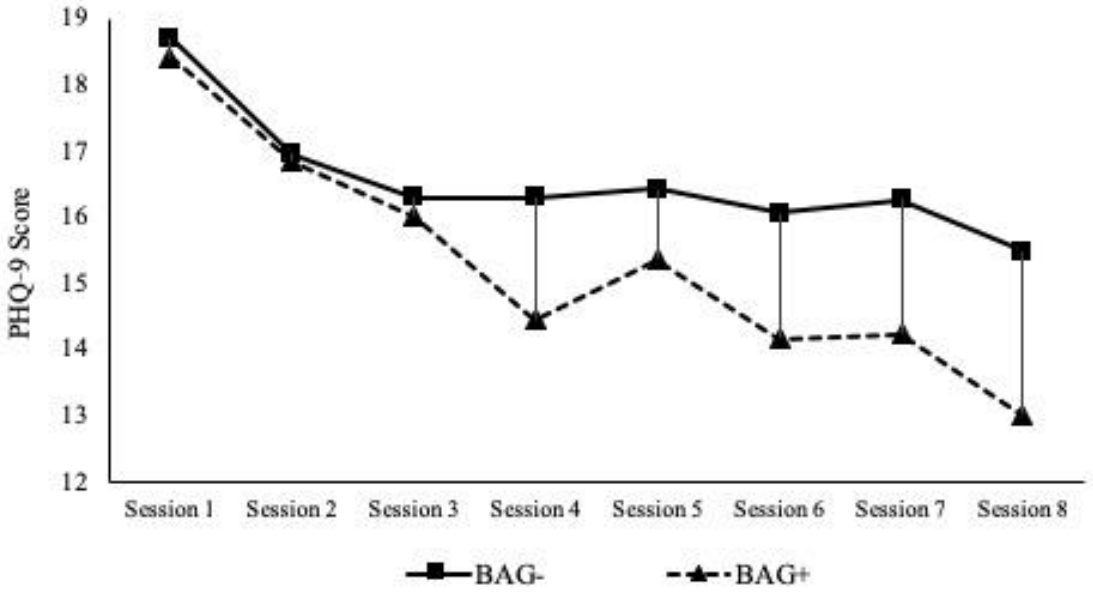


Figure 5.6. Session-by-session PHQ-9 scores for BAG- and BAG+

In relation to the secondary outcomes, an ANOVA showed anxiety symptoms significantly decreased following BAG treatment (pre-post main effect; F(1,

60)=25.383, p<.001). Overall anxiety scores were significantly lower in BAG+ compared to BAG- (BAG main effect; F(1, 60)=4.316, p=.021). However, the difference in amount of pre-post-treatment anxiety reduction between interventions was not significant (interaction effect; F(1, 60)=1.983, p=.082). Effect sizes indicated BAG- and BAG+ both produced moderate pre-post reductions in anxiety symptoms. There was a significant, moderate difference in post-treatment anxiety scores, showing lower anxiety symptoms after BAG+ compared to BAG-. However, BAG- sample had higher pre-treatment anxiety scores (d =0.35). Comparison of anxiety change scores that took pre-treatment levels into account indicated a smaller effect size in favour of BAG+ that was not statistically significant. An ANOVA of WSAS scores showed impaired functioning significantly decreased following BAG treatment (pre-post main effect; F(1, 60)=21.243, p<.001). Overall impaired functioning scores did not differ significantly between interventions (BAG main effect; F(1, 60)=0.027, p=.436). The effect of treatment on functioning was not significantly different for BAG- compared to the BAG+ (interaction effect; F(1, 60)=0.401, p=.265). Improved functioning after both interventions was representative of a moderate effect. Differences in post-treatment scores between the interventions were minimal (d <0.2) and not statistically significant.

In summary, the addition of the implementation intentions augmentation in BAG+ produced significantly greater reductions in depression symptoms after treatment, but did not significantly affect greater reductions in the secondary outcomes of anxiety symptoms or impaired functioning.

5.3.4 Hypothesis 3: BAG+ effect on individual outcomes

To test the hypothesis (3) that the BAG+ implementation intentions augmentation would result in fewer stasis outcomes compared to BAG-, RCI rates were compared across groups. Table 5.7 reports the proportion of individual outcomes across BAG- and BAG+ delivery. In terms of the number of patients who did not benefit from

the BAG intervention, BAG+ produced a significantly lower number of stasis outcomes. Patients who received the augmented BAG treatment were three times less likely to experience a stasis outcome at the end of treatment than those who had received BAG-. Further examination of the proportion of other RCI categories demonstrated that the reduced stasis outcomes were explained by significantly more patients in BAG+ experiencing improvement in their depression symptoms or full recovery (change in caseness in addition to improvement). No patients experienced a change in caseness (moving from above to below the clinical cut-off after treatment) without accompanying improvement (i.e., representing recovery). Therefore, the statistics for *caseness change* and *recovered* were the same. No patients experienced a reliable deterioration in their depression symptoms after attending BAG- or BAG+ interventions. In summary, the addition of the implementation intentions augmentation in BAG+ resulted in significantly fewer patients experiencing a stasis outcome at the end of treatment, due to a greater proportion of patients experiencing symptom improvement or full recovery.

Table 5.7. *Recovery rates for BAG- and BAG+ at treatment completion (N = 62)*

Post-treatment PHQ-9 recovery status	**BAG- (n = 31)**	**BAG+ (n = 31)**	**Chi-squared (p value)**	**Odds ratio (BAG+:BAG-)**
Recovered	9.7% (3)	29.0% (9)	3.72 (p=.027)	3.82
Improved	22.6% (7)	48.4% (15)	4.51 (p=.002)	3.21
Stasis	77.4% (24)	51.6% (16)	4.51 (p=.002)	0.31
Deteriorated	0% (0)	0% (0)	-	-

Note. 'Recovered' represents the proportion of patients who showed clinically significant change in addition to reliable improvement. Therefore, 'Recovered' and 'Improved' categories are not mutually exclusive.

5.4 Discussion

5.4.1 Summary of main findings

This study tested two clinical augmentations aiming to increase attendance and reduce stasis at group BA delivered in routine practice. The findings partially supported the three study hypotheses. Contrary to the first hypothesis, a psychoeducation treatment augmentation did not result in greater attendance rates at BAG. However, in comparison to the existing treatment, patients receiving the augmented intervention were approximately three times more likely to improve after therapy. The addition of the implementation intentions treatment augmentation resulted in greater reductions in depression (albeit only just reaching significance) and fewer stasis outcomes at the end of BAG, providing tentative support for the second and third hypotheses for the primary outcome. Significantly greater reductions in the secondary outcomes, anxiety and impaired functioning, were not evident after augmented BAG treatment.

5.4.2 Empirical relevance for BAG in clinical practice

First, the results provide further evidence that BA is clinically effective when delivered in a group format in routine practice, producing moderate to large reductions in depression, anxiety and impaired functioning (Houghton et al., 2008; Kellett et al., 2017; Porter et al., 2004). Importantly, there were no cases of depression symptom deterioration after BAG treatment, despite it often being reported that approximately up to 10% of patients may get worse after psychotherapy (Cahill et al., 2010; Lambert, 2013).

Second, the findings show preliminary evidence to suggest the BAG intervention can be enhanced to produce better depression outcomes. With the exception of the two treatment augmentations, all BAG- and BAG+ patients received the same intervention. However, BAG+ produced greater reductions in depression symptoms, albeit as the difference only just reached significance this finding should be treated

cautiously. On average, patients ended BAG+ with moderate depression symptoms, in comparison to moderately severe depression symptoms after BAG-. Benefits of the augmentation were more evident in the individual outcomes, with stasis outcomes reduced from approximately 77% to 52%. The fewer patients experiencing no change in their symptoms were a consequence of 26% more patients experiencing reliable improvement, with 19% of those also moving below the clinical cut-off for depression. The added benefit of BAG+ for depression outcomes therefore appears attributable to the treatment enhancements. Few prior studies have attempted quality improvement through integrating augmentations into existing depression treatments. Studies that have enhanced treatments have commonly used pharmacological or technological adjuncts, which require rigorous testing or technical expertise and ongoing maintenance costs (Donker et al., 2013). In contrast, the present preliminary results were produced through small tweaks to an extant treatment protocol using a research-informed approach.

5.4.3 Theoretical interpretation of the quality improvement effect

The existing BAG treatment was augmented with two simple, low-cost strategies that were easily integrated into the existing BA group structure. One was a data-driven augmentation targeted at drop-out, while the other was a theory-driven augmentation targeted at stasis outcomes. The lack of improved treatment retention, while depression outcomes appeared simultaneously improved, disaggregates the impact of the two augmentations. The divergent findings suggest forming implementation intentions to set between-session work improved the quality of the therapy. Patients experienced greater improvements despite attending the same number of treatment sessions, thereby the benefits did not come from receiving a higher treatment dose, but from the nature of the therapy working better. The beneficial effect of the implementation intentions and not the treatment dose psychoeducation

enhancement, suggests theory-driven, rather than data-driven augmentations might be more effective at improving outcomes.

5.4.3.1 Implementation intentions mechanism for reducing stasis

These findings build on the promising use of implementation intentions in mental health contexts (Lucock et al., 2018; Sheeran et al., 2007; Toli et al., 2016; Varley et al., 2011), by providing the first indication that 'if-then' planning techniques can be integrated into existing psychological treatments without affecting the integrity of the therapy. As suggested by Toli et al. (2016), it is likely the theoretically-coherent behavioural principles of BA aided the integration of implementation techniques into the process of setting between-session activities. This would suggest similar enhancements could be made to other CBT-based treatments that have a focus on setting between-session tasks, such as exposure or behavioural experiments.

The current findings present some evidence that patients may have been better able to gain benefit from the BAG+ intervention compared to BAG-. Although homework completion was not measured, previous findings have showed 'if-then' plans helped people with depression to engage more in personal activities (Fritzsche et al., 2016). Fritzche et al. (2016) also reported depression symptoms appeared to reduce, a finding that has been tentatively replicated in this study. Additional support for the implementation intentions mechanism accounting for this improvement (through more effective homework completion), is provided by the digression in depression outcome trajectories for BAG+ compared to BAG- at session three. Reduced BAG+ depression outcomes appeared at the first outcome measure after patients were tasked with the TRAP/TRAC between-session activities, comprising the first practical activity focused at changing behaviours (by targeting avoidance through elicitation of positive replacement behaviours). This suggests patients using 'if-then' plans benefitted more after receiving this core principle of BA than those who did not. Given that the main

difference between the two samples was the use of implementation intentions to set the between-session work, it appears that the improvement may have come from an improved ability to put those principles into action.

Theoretically, this outcome could fit with how the mechanisms of BA and 'if-then' plans map onto each other. The underpinnings of BA highlight the importance of context in both the maintenance of depression, and the breaking of depressive cycles (Martell, Addis & Jacobson, 2001). Consequently, BA treatment targets are directed at the recognition of contextual factors which are maintaining avoidant and depressive behaviours. Similarly, implementation intentions promote the use of contextual cues to initiate pre-planned action Sheeran & Webb, 2016). The mechanisms of 'if-then' plans interlink with the principles of BA. Therefore, it is feasible that they stimulate a stronger link between contextual cues and behaviour change, which has a knock on effect on depression outcomes. Consequently, more patients were able to experience a benefit from therapy.

5.4.3.2 Treatment-dose education mechanism for increasing treatment retention

The prior literature on improving treatment retention through education about treatment attendance and realistic rate of improvement had found mixed results (Delgadillo et al., 2015; Reis & Brown, 2006; Swift & Callahan, 2011). It was thought that the variable effects might be in part due to difficulties in ensuring patients read the orientation leaflets when sent prior to therapy. Yet, in spite of an attempt to increase the salience of the information by using applicable practice-based dose-response evidence embedded in the treatment content of the first BAG+ session, attendance rates were not affected. This is in contrast to the beneficial effect of orientation leaflets found by Swift & Callahan (2011), but is in line with the null findings from Delgadillo et al. (2015). Questions are raised as to whether the dose-response education in Delgadillo et al.'s study and the present study failed to influence patient therapy expectations. It may just

be that dose-response information has minimal effect on patients who are prone to drop-out regardless. Depression is known to have a considerable effect on attention and memory (Otte et al., 2016), so it is possible patients found it difficult to process and remember the psychoeducation information.

It is also possible that other differences in the therapy or context prevented improved expectations having an influence on attendance in the way they did in the Swift & Callahan (2011) study. Two differences that were evident are i) the use of briefer interventions (typically not exceeding 8 sessions), in contrast to recommendations of 13-15 sessions, and ii) group therapy (a mix of individual and group interventions in Delgadillo et al., 2015) in contrast to exclusively individual treatment. With regards to brief treatments, it may be that differences in treatment duration expectations and the reality are not that significant. Any potential influence on attendance rates is subsequently diluted. On the other hand, group therapies face a different set of challenges in relation to treatment attendance and retention (i.e. attitudes about group therapy and its effectiveness). Reasons for drop-out might not be sufficiently addressed with dose-response education alone. Nonetheless, the improvement in individual outcomes without an increase in attendance observed in this study demonstrates that finding a way to successfully increase attendance is a mechanism that could help to improve outcomes even further.

5.4.4 Limitations and future research directions

The major limitations relate to the methodological design and were largely a result of conducting the study within routine practice. The use of practice-based data ensured these effects for BAG are highly relevant to clinical practice. Confidence in the applicability of the results is further strengthened by methods used to limit threats to internal validity that often come with naturalistic studies. PSM and ITT analyses were employed to emulate RCT procedures, while multiple attempts to monitor treatment

integrity provided assurances that BAG was delivered as intended. However, the inevitable impacts on internal validity do require consideration in terms of the results.

The use of a historical control group means patients were not able to be randomised to treatments or treatments be fully controlled. Although PSM helped to reduce the impact of selection biases, the historical nature of the control means it cannot be definitively stated that slight improvements in outcomes were entirely due to the treatment augmentations. Over the time-period of BAG- compared to BAG+, small changes in service delivery such as session order, different therapists or even the delivery of BAG simply getting better from experience might have accounted for improvements in outcomes. Similarly, there were no checks of augmentation compliance (due to considerations about patient and therapist burden in routine practice). Specifically, the lack of a measure of homework completion means the proposed process of outcome improvement cannot be confirmed by these results. Although it is suggested that it occurred through the use of implementation intentions increasing homework completion, a process evaluation using mediation analyses would provide a clearer indication of the mechanism of 'if-then' plans. The medium and long-term effects of BAG+ over BAG- were not able to be assessed, as no follow-up was conducted. Long-term assessment of outcomes after BAG+ is needed to investigate the durability of the effects and whether the implementation intentions technique can continue to be useful for patient's activation of non-depressed behaviours after therapy has ended.

There were a number of weaknesses relating to the nature of practice-based data. First, missing data was unavoidable due to the routine service setting. Treatment completion rates of the full eight sessions were sub-optimal in general, so session-by-session scores were utilised to ensure a pre-post score was available for the whole sample. Nonetheless, the use of the LOCF method does have acknowledged faults and

reduces accuracy of the results (Lachin, 2016). Second, the results are based solely on self-report data that are known to be at risk of validity issues, such as social desirability when providing sensitive information (Tourangeau & Yan, 2007). Third, session-by-session scores were attained using the same outcome measure meaning the pre- and post-treatment assessment is subject to testing effects (Wampold, 2015). A combination of self-report and therapist assessment measures would strengthen the interpretation of future studies. Finally, although checks were performed to assess the suitably of single-level analyses for the clustered data (by nature of group delivery), some argue ICCs as low as 0.01 can still violate dependency assumptions and have an impact on analyses (Baldwin et al., 2011). WSAS outcomes had the highest ICC (0.06), therefore power to detect reductions in impaired functioning may have been affected. It should also be noted that there was an inflated risk of a Type 1 error due to the use of one-tailed tests. Taking these limitations into account, it is imperative that these results are replicated before firm conclusions can be made. Full testing of the efficacy, follow-up and cost-effectiveness of an implementation intentions enhanced treatment using an additive component type RCT would facilitate this (Cuijpers, Cristea, Karyotaki, Reijnders, & Hollon, 2017).

5.4.5 Clinical and research implications

Given the high frequency of stasis outcomes, researchers, service managers and clinicians have a responsibility to improve depression treatments. Large scale organisational change or developing new treatments are clearly not the only methods for improving outcomes. Small tweaks to clinical practice appear to have the potential to produce meaningful change for individual patients at far less cost and with quicker time frames to implementation. The present results advocate the utility of small, low-cost, theory-informed augmentations to currently available therapies as a method for improving patient outcomes. These kind of strategies, once sufficiently empirically

supported, could be easily implemented by clinicians and would see patients experiencing the benefits relatively quickly, without the need for slow and complex organisational change. While the current findings are only generalisable to group-based BA treatments, the techniques themselves are not specific to use in BAG. It would be interesting if the same effect could be seen for individual BA and therefore provide a useful quality improvement approach for potential mass therapy dissemination. Considering in the UK, 560,000 patients a year receive psychological treatment in IAPT services (Clark, 2018), scaling up the 26% increased rate of improvement observed here could see up to 140,000 more people benefitting. In effect, better resource utilisation within the treatment content could produce small effects that manifest into a big impact for patients.

It should be noted however, that although BAG+ reduced the rate of stasis, approximately half the patients receiving treatment still failed to experience meaningful change in their depression. This is indicative of the scale of variability in outcomes experienced by patients with depression, especially in routine practice (Hansen et al., 2002). Clearly BAG can be effective for certain patients (and potentially be made more effective for others). However, the large proportion of non-response even after attempts to reduce it might reflect that there is a subset of patients that are not going to benefit from BAG regardless of treatment optimisation. Stasis needs tackling on multiple fronts. Individual and clinical factors may interact to make patients more suited to certain psychotherapies; the subset that do not improve after BAG might be able to gain benefit from a more cognitive-focused treatment, and vice versa. Treatment matching could reduce this mismatch (Kwan, Dimidjian, & Rizvi, 2010) and research is required to get a better understanding of how patient preferences and symptom clusters may respond better to specific treatment components.

5.4.6 Conclusions

In conclusion, quality improvement strategies integrated directly into treatment components may have the potential to improve outcomes and reduce stasis rates. Treatment response psychoeducation integrated into a brief group BA treatment for depression does not seem to affect treatment retention. However, the use of implementation intentions as a technique for reducing stasis outcomes, potentially by increasing treatment compliance, showed promise. Low-cost quality improvement augmentations of existing therapies through theoretical analyses of outcome predictors and suitable enhancement techniques offer a simple and direct way for services to improve outcomes. The variety of processes that help to produce positive change during psychotherapy for depression provide multiple targets for these types of enhancements. In aid of these strategies, future research should continue to establish the processes that enable treatments to exert their positive effects and reduce the impact of their negative effects. Studies should adopt theory-driven augmentations of existing therapies as a method of treatment optimisation in routine practice. The high rates of stasis, even after improvement attempts, require considerable attention going forward, but as shown in the present study, small effects can result in meaningful benefits for patients.

CHAPTER 6

Values and Depression Outcomes During Group Behavioural Activation Therapy

The previous chapter demonstrated outcomes of an existing group BA treatment could be improved, resulting in more patients benefitting from treatment. Treatment was enhanced using a theoretically-informed strategy targeted at a crucial element of therapy (between-session homework engagement). To build on these results and identify other theoretically-informed strategies, there needs to be a better understanding about what are the active ingredients of group BA that produce the reductions in depression. The final empirical study therefore attempted to establish a mediator of outcome in group behavioural activation (BA) for depression. Increasing behaviours that are in accordance with life values (termed *valued living*) was proposed as a mediator of change. The role of *valued living* was investigated by 1) analysing the association between patient values and depression prior to treatment, 2) exploring whether depression change can be predicted by pre-treatment values, 3) establishing if group BA therapy produces changes in *valued living* and 4) attempting to evaluate the mediating effect of increased *valued living* on depression outcomes. Methodological limitations hampered the mediation investigation and those results are therefore presented as exploratory findings in Appendix H.

5.5 Introduction

5.5.1 Values and depression

Values are the personal importance placed on various domains of life that any person seeks to life their life by (Hayes, Strosahl, & Wilson, 1999). Values are freely-chosen, ongoing, dynamic and unique to each individual and create and then reinforce behaviours associated with values (Wilson & DuFrene, 2009). They can apply to any

domain of life, including family ('father figure'), relationships ('kind'), work ('team worker'), education ('life-long learner'), health ('caring for self'), spirituality ('compassionate') or leisure ('player'). Unlike goals values do not have an achievable end-point, and more act as psychological blueprints to how people want to live their lives, so informing how they behave and the decisions they make (Schwartz, 1992).

Values can be viewed across two dimensions, (i) *values importance*, the level of meaning attributed to aspects of life and (ii) *valued living*, the extent to which behaviour is in line with values. Valued behaviours are therefore differentiated from purely pleasurable activities due to the meaning that underpins them (Wilson & Murrell, 2004). Living in accordance with personal values provides meaning and a sense of purpose, and has been shown to have a positive effect on wellbeing, functioning and life satisfaction (Ferssizidis et al., 2010; Lyubomksky, Sheldon, & Schkade, 2005; Rogers, 1964).

Depression obstructs individuals from living a life they value (Stangier, Ukrow, Schermelleh-Engel, Grabe, & Lauterbach, 2007). In behavioural models of depression, the avoidant behaviour patterns that contribute to the maintenance of symptoms act as a barrier to engagement with *valued living* (Jacobson, Martell, & Dimidjian, 2006). Finding ways to relieve feelings of distress through avoidance becomes a prominent focus. Although distressing situations may be avoided, it often comes at the expense of values. In other words, their action is not in line with their value. Therefore, in depression there is often a dislocation between personal values and the consistent behavioural expression of these values.

Avoidance is associated with increased discrepancies in *valued living* (Smout, Davies, Burns, & Christie, 2013). As depression intensifies, 'inactivity breeds inactivity' and the discrepancy between what is valued as important and behavioural expression widens. Consequently, people suffering with depression often report less

frequent or effective engagement with *valued living* (Wilson & Murrell, 2004; Wilson, Sandoz, Kitchens, & Roberts, 2010). In turn, actions that often sharply contrast with life values (e.g. not watching a child's end of year assembly due to lack of motivation) only serve to produce more distress and depression (Plumb & Hayes, 2008; Plumb, Stewart, Dahl, & Lundgren, 2009).

5.5.2 Incorporating values-based interventions into behavioural activation treatment

Given the impact depression appears to have on *valued living*, increasing behaviour that is consistent with values might help relieve depression symptoms. On this basis, newer, more complex behavioural interventions have incorporated values work into treatment protocols. Acceptance and commitment therapy (ACT; Hayes et al., 1999) first pioneered values as a treatment focus. Epilepsy patients receiving ACT that engage with *valued living* have increased well-being and quality of life (Lundgren, Dahl, & Hayes, 2008), and increases in *valued living* have been associated with reductions in depression and distress following ACT (Bramwell & Richardson, 2018; Vowles & McCracken, 2008).

The development of behavioural activation (BA) as a stand-alone treatment for depression is based around the core technique of increasing activation (Martell et al., 2001). The relevance of considering what a patient values when trying to increase their activation has led to the integration of ACT-based values work into newer, more intensive BA protocols (Hopko & Lejuez, 2007; Martell et al., 2010). Incorporating values-driven behaviour into the process of increasing activity, encourages focus on personally meaningful behavioural change. In BA, the patterns of avoidance that impact on engagement with *valued living* are identified and approach behaviours are elicited that realign patients' actions with their values (Martell et al., 2010).

In BA, values act as both 'direction' for goals and 'motivation', through stimulating engagement. Patients in a depressive episode need motivation to engage in a non-depressive activity that exceeds the short-term comfort provided by withdrawal (Wilson & Murrell, 2004). Theoretically, a behavioural target focused on what is valued has the potential to provide more motivation (Dahl, Plumb, Stewart, & Lundgren, 2009). By acting in line with what is valued, the degree of positive reinforcement is also likely to be stronger than merely pleasurable activity. Acting according to life values was shown to be associated with increased activation and achievement of scheduled activities and in turn greater reward (Doi, Yokomitsu, & Sakano, 2016). The inclusion of values-based work in more complex protocols is therefore thought to facilitate patients to implement more meaningful behaviour change strategies.

5.5.2.1 Values assessment for clinical use

Implementation of values-based work in BA treatment requires therapists to gain a picture of the patient's life values and the consistency of their current behaviour. Several tools have been developed to assess values, including the Schwartz value inventory (SVI; Schwartz, 1992), survey of life principles (SLP; Ciarrochi & Bailey, 2008) and the valuing questionnaire (VQ; Smout et al., 2013). However, the most widely implemented tool for behavioural analysis values assessment is the Valued Living Questionnaire (VLQ). The VLQ was developed as a clinical tool to help clinicians assess the level of values-behaviour imbalance and elicit associated behavioural targets (Wilson et al., 2010). The VLQ captures both Values Importance and Values Consistency dimensions across ten domains of life, with the difference between values and importance ratings representing the size of the discrepancy in *valued living*. The total scores on the *importance* and *consistency* dimensions and the *valued living* score produces a 'values profile'. Therapists use values profiles to help

guide treatment by identifying suitable behavioural targets for BA activation strategies (Wilson & Murrell, 2004).

5.5.3 Mechanisms and mediators of outcome change

Symptom improvement during psychological therapies occur due to a therapeutic mechanism that produces change (Kazdin, 2007). Mediators are variables that account for the associations between the therapy and resulting symptom reductions. Intervening mediator variables are not a mechanisms in themselves, but can provide information about how possible change mechanisms work within a therapy (Kazdin, 2007). Understanding the psychological processes of change that produce the clinical benefits of therapy plays a valuable role in the evaluation of treatments and so has far reaching implications for tackling stasis (Hopko et al., 2011). Treatment effectiveness can be optimised if key mediators of change can be identified, in that techniques that enable patients to produce the most change can be prioritised.

Although BA has been shown to be effective at reducing depression, the mechanism by which it produces change is still to be determined At the heart of BA theory is change is brought about by increased positive reinforcement for non-depressive behaviours (Lewinsohn et al., 1980). Increased activation (and subsequent positive reinforcement) have been the favoured primary mechanism of action, with temporal investigations attempting to confirm this more conclusively (Santos et al., 2017). The majority of work to produce change in BA takes place in the between-session activities (i.e., when the patient engages in the valued activities) , therefore it seems logical that a mechanism that brings about change is present within this process. However, despite the long-known positive correlation between mood and activity (Lewinsohn & Libet, 1972), sheer quantity of activities has not been shown to mediate BA treatment benefits (Ryba, Lejuez, & Hopko, 2014). Instead proportion of activities completed appeared more influential for improved outcomes. Ryba et al. (2014)

suggested this conflicting finding may be explained by greater quantities of activities becoming less in line with life values and therefore, less salient.

BA protocols that promote values work have continually been shown to be effective for reducing depression symptoms (Dimidjian et al., 2006; Hopko et al., 2003; Richards et al., 2016). However, the explicit role values play in contributing to the BA treatment effect has not been established (Hopko et al., 2011). The existing literature suggests values-based activities (termed *valued living*) may have a mediating effect. The proposed motivating benefit (Dahl et al., 2009) and greater experiential reward of activities in line with values (Doi et al., 2016) may influence completion and salience of activities, and in turn reduce depression symptomology. The added association of living life in accordance with your values with less depressed mood (Jarden, 2010; Plumb & Hayes, 2008) makes it reasonable to propose *valued living* as a mediating variable for depression in BA.

Despite the link between values and depression, *valued living* has as of yet only been examined as an active ingredient within ACT-based treatments. Increases in *valued living* have been associated with improvements in outcomes for depression (Bramwell & Richardson, 2018), generalized anxiety disorder (Hayes, Orsillo, & Roemer, 2010), panic disorder (Wersebe et al., 2017) and chronic pain (Vowles & McCracken, 2008). Systematic longitudinal investigations have shown increased *valued living* accounts for a significant proportion of outcome change (Lundgren et al., 2008) and implicated changes in *valued living* preceding reductions in outcomes (Gloster et al., 2017). It would appear there is evidence for a mechanism by which through increasing *valued living*, suffering can be reduced. Whether there is a similar effect for BA and depression symptom reduction remains to be seen.

5.5.4 Relationship between values and depression outcome in BA

Values provide an avenue for investigation into the potential processes by which BA exerts its treatment effect. There has been relatively limited empirical investigation into values during BA treatment. This study therefore sought to further explore the relationship between values and depression during group BA treatment, in addition to investigating whether *valued living* mediates depression reductions.

Depression is a heterogenous disorder and patients do not all present with the same sets of symptoms. Symptoms can typically be clustered into affective (cognitive-based) and somatic (physical-based) symptoms (Elhai et al., 2012). Affective symptoms consist of low mood, lack of interest in activities and feelings of worthlessness, whereas somatic symptoms are physiological such as trouble sleeping, issues with eating or fatigue. Patients presenting with different depression profiles (e.g. high versus low somatic symptoms) may respond to treatment in a different manner (Zimmerman, Ellison, Young, Chelminski, & Dalrymple, 2015). Neither BA, nor the construct of values has been investigated in relation to symptom clusters of depression. BA has a treatment model that encourages behaviour changes, in spite of low mood or physical symptoms. If the greater presence of one symptom cluster over the other responds to the behavioural strategies differently, it could provide additional information about the mechanism of change in BA interventions. Likewise, the relationship between values and depression symptom clusters could provide insight into BA mechanisms. *Valued living* provides individuals with meaning in their lives, so behaviour that is incongruent with values might contribute more strongly to affective symptoms such as feelings of worthless and low mood rather than somatic symptoms.

The clinical use of the VLQ provides opportunity to further explore the different dimensions of values, in terms of *values importance* and *valued living.* Studies that have investigated values changes after psychotherapy have tended to find changes vary for different values dimensions. Measures of values importance generally tend to show less

change after values-based interventions (Bramwell & Richardson, 2018; Wersebe et al., 2017), while measures of *values consistency*, *valued living* or *discrepancy* have been more receptive to change following ACT (Hayes et al., 2010; Michelson, Lee, Orsillo, & Roemer, 2011; Vowles & McCracken, 2008; Wersebe et al., 2017). Moreover, what is valued as important has been shown to be less consequential to positive wellbeing than the extent of engagement with values (Bahraini et al., 2013). This finding would fit with development of the VLQ, whereby importance of values is expected to remain relatively stable. On the other hand, engagement with values is thought to be more fluid and receptive to targeted interventions (Wilson et al., 2010). Taken together, this would suggest BA would have an effect of closing the values-behaviour gap, whereas values importance would be relatively unaffected.

The level of importance placed in values and the extent of engagement with *valued living* patients begin treatment with could also have implications for the amount of symptom change experienced during BA. Low overall levels of values importance may reduce the benefit of BA. As a lack of positive reinforcement predicts depression severity (Carvalho, Trent, & Hopko, 2011), BA aims to bring about change through increased positive reinforcement from increased activity. The benefit of valued activity is posited in the stronger response-contingent positive reinforcement it produces over a less valued activity (Doi et al., 2016). If patients place minimal importance on all values in their life, they may not be able to experience the additional positive reinforcement from valued activity to the extent of a patient who places great importance on valued areas. The resulting BA treatment effect could then be diminished.

In addition, smaller discrepancies between values and action at the start of treatment could affect outcomes. A study of an acceptance-based behavioural therapy with patients suffering with generalised anxiety disorder (GAD) found treatment non-responders began treatment with higher levels of engagement with *valued living*

compared to responders, but did not go on to experience as much change (Hayes, Orsillo, & Roemer, 2010). One interpretation could be that if patients perceive their behaviour to already be in line with their values (regardless of whether that is the case or they do not have the insight to recognise it), interventions aimed at increasing *valued living* have less of an impact on symptoms. If this is the case, it would be expected that patients who begin treatment with acknowledged greater discrepancies in *valued living* would experience greater benefit and resulting change from the values-based intervention within BA.

5.5.5 Aims and hypotheses

In summary, depression appears to be linked to reduced *valued living* and vice versa. Values work aimed at increasing *valued living* have been integrated into BA treatments as a method for improving depression symptoms. However, it is not known precisely how values in BA contribute to the treatment outcome and whether change in values enable change in certain depression symptom clusters. Values seem theoretically relevant to the processes of change in BA and could provide insight into how BA exerts its treatment effect. By understanding the predictors and mediators of change that occur in BA, optimised interventions could be developed.

The objective of this study was to conduct a pilot investigation into the relationship between values and depression during routinely delivered BA group treatment. The study had four aims: (1) to examine the pre-treatment relationship between values and depression symptoms, (2) to identify whether pre-treatment values levels predict clinical outcome, (3) to investigate changes in values following group BA treatment and (4) to explore *valued living* as a mechanism of change for depression reduction during BA.

Study hypotheses were therefore as follows;

(1) patients will present at the start of treatment with a discrepancy between what they value and their current behaviour. Pre-treatment values levels will be related to pre-treatment depression symptom severity, and affective rather than somatic symptoms clusters.

(2) pre-treatment levels of values will predict the amount of change in depression following treatment;

(3) importance of values will remain stable, but discrepancies between values importance and action will decrease following treatment (i.e., increased *valued living*).

(4) changes in *valued living* will mediate change in depression outcomes.

5.6 Method

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*). Information and evidence about ethical approval can be found in Appendix C (research protocol, ethical approval confirmation).

5.6.1 Design

A repeated-measures design was employed to utilise session by session outcomes from the routine clinical practice delivery of behavioural activation group (BAG) therapy. Data for this study were collected as part of the BAG+ arm in the intervention study described in the previous chapter. The details of the intervention, facilitators and treatment process are therefore presented in the previous chapters. Additional methodological information specific to the current study is described below.

5.6.2 Participants

5.6.2.1 Inclusion criteria

All patients who attended BAG in a UK Improving Access to Psychological Therapies (IAPT) service between January and December 2017 were invited to take part in the study. It was a study of routine practice so inclusion criteria were minimal. Inclusion criteria were (a) seeking treatment for a primary presenting problem of depression (co-morbid anxiety symptoms were accepted as long as depression was the primary diagnosis), (b) self-referred to BAG treatment or referred following assessment by Psychological Wellbeing Practitioners (PWP), (c) at least 18 years old, and (d) gave informed consent for their data to be used in the study.

5.6.2.2 Sample size

A power analysis using G*Power indicated a sample size of 52 patients would be required to detect a small effect (d = 0.35) in change in *valued living* with .80 power using a paired t-test at p = 0.05. Meanwhile, a sample size of 55 patients would be required to detect a moderate effect (f^2 = 0.15) for pre-treatment values predicting depression change with .80 power, using multiple regression at p = 0.05. Six BAG groups were scheduled to be delivered during the data collection period, and service data about previous delivery of six BAG groups in the same time period suggested a sample size of 55 seemed attainable. However, due to unforeseen service issues, only three BAG groups were delivered in 2017. Following the data collection period, the sample consisted of 28 patients, falling short of the requirements for adequate power. Due to service factors, BAG treatment delivery was suspended. Therefore, further data collection to increase the sample size was not possible. The power of the available sample size was .56 for the t-test and .47 for the regression.

5.6.3 Outcome measures

5.6.3.1 Valued Living Questionnaire (VLQ)

The VLQ is a 20-item measure designed to measure the extent that patients are engaging with the valued areas of their lives (Wilson et al., 2010) and was used as the

primary outcome measure for evaluating values (see Appendix D for copies of the outcome measures).. It consists of Importance and Consistency scales, both of which are rated on a 10-point Likert scale for 10 domains of valued living: (1) family, (2) marriage/couples/intimate relationships, (3) parenting, (4) friendships/social relationships, (5) work, (6) education, (7) recreation, (8) spirituality, (9) citizenship, and (10) physical self-care.

Individuals are first asked to rate how important they personally deem each of the 10 domains to be in their life. This produces a *Values Importance* total score out of 100. Individuals are then asked to rate how consistently they have engaged with each domain in terms of how important they rated it (in the previous Values Importance section). For example, if education is rated as of little importance in terms of what an individual values (scored towards the lower end of the importance scale) and that individual has undertaken no action that week that relates to education, then they have been living fairly consistently with that value (score towards the higher end of the consistency scale). Whereas, if friendships are rated as a highly important value (scored towards the higher end of the importance scale), but the individual has not engaged in any social activities, then they have not been living consistently with that value (score towards the lower end of the consistency scale). This produces a *Values Consistency* total score out of 100.

To quantify how an individual's actual behaviour matches up with the valued areas of their life, a weighted composite *Valued Living* score is produced. The *Valued Living* score is calculated by multiplying each domain's Importance and Consistency scores and then averaging the cross products to produce a score ranging from 1-100. Higher *Valued Living* scores signify higher levels of *valued living*. Finally, a *Discrepancy* score gives an indication of the size of the discrepancy between behaviour and what is valued as important. The discrepancy score ranges from -90 to +90 and is

calculated by subtracting the total Importance score from the Consistency score. Larger negative scores signify greater discrepancy between Values Importance and Consistency. Preliminary results have shown the VLQ *Valued Living* score has adequate internal consistency (α= .65 - .74), good test-retest reliability (.75) and adequate construct validity (Wilson et al., 2010). The internal consistency of the measure in the present study sample was acceptable to good (α= .65 - .83).

5.6.3.2 Patient Health Questionnaire-9 (PHQ-9)

The Patient Health Questionnaire-9 (PHQ-9) was used to measure depression symptoms. See Chapter 4 (section 4.2.4) for full description of the PHQ-9 measure and psychometric properties. Depression was assessed using both the total scale score and two sub-scale scores (referring to affective and somatic symptoms). The most commonly supported model from factor analyses of the PHQ-9 has defined a two-factor structure, comprising of affective and somatic symptom clusters (Elhai et al., 2012; Petersen et al., 2014). The affective symptoms cluster comprises four items, asking about 'anhedonia', 'low mood', 'feelings of worthlessness' and 'suicidal thoughts'. The somatic symptom cluster comprises five items asking about 'trouble sleeping', 'fatigue', 'appetite problems', 'trouble concentrating' and 'slow or agitated psychomotor skills".

5.6.3.3 Demographic information sheet

For the purpose of this study, patients were asked to complete a demographic information sheet to capture information about demographics (age, gender, ethnicity), current antidepressant medication and previous episodes of depression and treatment.

5.6.4 Procedure

All patients received manualised BAG therapy consisting of eight, weekly, two-hour sessions facilitated by two British Association of Behavioural and Cognitive Psychotherapy (BABCP) accredited cognitive behavioural therapists (see Chapter 6 for full intervention details). At the first BAG session, facilitators invited patients to

participate in the study. Informed consent and demographic information was obtained from patients willing to allow their data to be included. Patients completed the PHQ-9 at each session attended. The VLQ was administered at the first session (pre-test), the second session (as part of the BA values session content) and at the final session (post-test).

5.6.5 Data analysis

5.6.5.1 Handling missing data

Data were analysed according to the intention to treat principle (ITT) using the entire sample entering treatment, regardless of data completion. Missing data were originally planned to be handled using multiple imputation in SPSS. However, the multiple imputation dataset was not compatible with the SPSS macro required for mediation analysis of repeated measures data. As BAG treatment delivery was suspended in the service following the initial period of data collection, increasing the sample size was not an option. To maximise the use of data, last observation carried forward (LOCF) imputation was used instead to produce scores from the last available measure. If there was only one score available, it was assumed there was no change. Although LOCF has documented statistical limitations (Lachin, 2016), it was deemed the best available option. LOCF is clinically applicable to IAPT criteria in that patients are classified as having 'received treatment' if they have attended at least one treatment session. The number of missing total scores was low, with missing PHQ-9 and VLQ totals for one patient at pre-treatment, seven patients at session two, and nine patients post-treatment.

5.6.5.2 Cluster effects of treatment delivery in groups

To assess the impact of clustering in the data, intraclass correlation coefficients (ICCs) were used to estimate the level of variance attributable to BAG group level factors. ICCs and the associated design effect (DE) for all the outcome measures (total

and subscale scores) were calculated using Equations 1 and 2 provided in Chapter four (see section 4.2.5.3). A DE of greater than two was used as an indication of significant co-dependence that would be unsuitable for analysis on a single-level (i.e., would require use of a multi-level model) (Muthen & Satorra, 1995). BAG treatment delivery was conducted via three groups. The average cluster size was nine. Table 6.1 reports the ICCs and DEs calculated for the VLQ and PHQ-9. All the DEs were less than two, with the exception of the VLQ Importance subscale. Single level analyses were deemed appropriate due to the DEs suggesting low dependence in the majority of outcomes. However, the power to detect an effect in analyses involving the VLQ Importance variables was affected and will have to be considered when interpreting the results.

Table 6.1. *ICC and DE estimates for study outcome measures*

Measure (*N*=28)	ICC	DE
VLQ Importance	0.18	2.44
VLQ Consistency	-0.04	0.68
VLQ Valued Living	0.02	1.16
VLQ Discrepancy	-0.08	0.36
PHQ-9 Total	-0.11	0.12
PHQ-9 Affective	-0.10	0.20
PHQ-9 Somatic	-0.08	0.36

Abbreviations: ICC: intraclass correlation coefficient; DE: design effect; VLQ: Valued Living Questionnaire; PHQ-9: Patient Health Questionnaire-9.

5.6.5.3 Preliminary analysis

Total scores were calculated for the overall PHQ-9, Affective and Somatic subscales and the VLQ Importance and Consistency subscales. VLQ Discrepancy scores were calculated by subtracting the Importance total from the Consistency total. An overall VLQ Valued Living score was computed from the mean product of each

domain's Importance and Consistency scores. Descriptive statistics were performed on all VLQ and depression variables. Change scores were computed for each outcome for pre to post BAG (session 1 minus 8). Change was specified so that a positive score indicated improvement in the variable (e.g., reduction in depression or increase in *valued living*) and a negative score indicated a deterioration in the variable (e.g. an increase in depression or decrease in *valued living*).

5.6.5.4 Statistical analysis

All statistical analyses were conducted in SPSS version 24. A paired t-test compared the importance and consistency scores to establish whether there was a discrepancy. Pearson's correlations were performed on VLQ and PHQ-9 pre-treatment scores to investigate the relationship between depression severity and symptom clusters and the values profiles of patients at the start of BAG. Hierarchical linear regression was performed to investigate pre-treatment levels of VLQ Importance and Discrepancy scores as predictors of depression change after treatment. Pre-treatment depression score was entered as a control variable, with pre-post change in depression as the outcome variable in step 1. Pre-treatment VLQ Importance or Discrepancy score was entered as the predictor variable in step 2. Separate analyses were performed to investigate prediction of overall depression symptoms, as well as the affective and somatic symptom clusters.

Paired *t*-tests were performed on the VLQ Importance, Consistency, Valued Living and Discrepancy scores pre- and post-treatment to examine the effect of BAG on values. Reliable change criteria were applied to the VLQ Valued Living score to assess individual outcomes (Jacobson & Truax, 1991). Significance of changes in depression outcomes were reported in Chapter 5, so were not the focus of this study.

Due to the reduced sample size the planned mediational analyses were substantially underpowered, restricting interpretation of the results. As a result the

mediation analyses were performed on an exploratory basis and are presented in Appendix H rather than in the main results. Repeated-measures mediation analysis using the MEMORE SPSS macro (Montoya & Hayes, 2017) was performed to explore whether pre-post change in the VLQ Valued Living score mediated pre-post change in depression. Again, separate analyses were performed for overall depression symptoms and affective and somatic symptom clusters. A path-analysis repeated-measures mediation model was employed using bootstrapped confidence intervals to establish significance of the indirect effect (a*b path).

Due to the lack of power in the sample size, effect sizes are reported where possible to give an indication of the effect magnitude regardless of significance. For correlation analyses, Pearson's *r* is reported, where 0.1, 0.3 and 0.5 are considered small, moderate, and large effect sizes respectively (Cohen, 1988). For pre-post analyses, Cohen's *d* was calculated, where 0.2, 0.5 and 0.8 are considered small, moderate, and large effect sizes respectively (Cohen, 1992). For regression analyses, Cohen's f^2 was calculated (using the equation $f^2 = R^2 / 1 - R^2$), where 0.02, 0.15 and 0.35 are deemed small, moderate, and large respectively (Cohen, 1988).

5.7 Results

5.7.1 Preliminary analysis

5.7.1.1 Clinical characteristics

The final sample consisted of 28 patients, 57% (*n*=16) female and 43% (*n*=12) male. Patients ranged from 20 to 65 years old, with a mean age of 41 (SD = 15.42). The majority of the sample identified as White British (89%, *n*=25), one (3.6%) as Indian, one (3.6%) as White Other and one (3.6%) as Other. Previous episodes of depression were prevalent, with 82% (*n*=23) of the sample reporting having at least one previous depressive episode (range 1 to 5+) and 75% (*n*=21) having received prior treatment for

depression. In addition, 71% reported currently taking antidepressant medication. Pre-treatment depression severity for the sample was classified as 32% (*n*=9) severe, 46% (*n*=13) moderately severe, 14% (*n*=4) moderate and 7% (*n*=2) mild depression. Nearly 80% (*n*=22) also met clinical caseness for anxiety with 36% (*n*=10) classified as severe, 32% (*n*=9) as moderate, 25% (*n*=7) as mild and 7% (*n*=2) as experiencing minimal anxiety.

5.7.1.2 Descriptive statistics

The means and standard deviations (SD) for all values and symptom variables at pre- and post-BAG treatment are presented in Table 6.2. Within-group effect sizes taking into account the pre-post correlation are also reported for each variable. Figure 6.1 plots the shape of change for *valued living* and depression session-by-session outcomes.

Table 6.2. *Means, SDs and effect sizes for values and depression variables at pre- and post-treatment*

	Pre		**Post**		**Change**	**Pre-post correlation**	**Effect size (*d*)**
	Mean	**(SD)**	**Mean**	**(SD)**			
Values variables (VLQ) (N=28)							
Importance	64.75	(17.34)	65.18	(18.41)	0.43	.85	-0.04
Consistency	51.96	(19.80)	56.25	(21.95)	4.29	.61	-0.23
Valued Living	35.31	(16.29)	39.61	(18.12)	4.30	.60	-0.28
Discrepancy	-12.79	(24.30)	-8.93	(23.07)	3.86	.64	-0.19
Symptom variables (PHQ-9) (N=28)							
Depression	17.50	(4.43)	12.75	(6.06)	-4.75	.31	0.77
Affective	7.29	(2.71)	5.21	(2.83)	-2.08	.40	0.69
Somatic	10.21	(2.38)	7.54	(3.72)	-2.67	.26	0.72

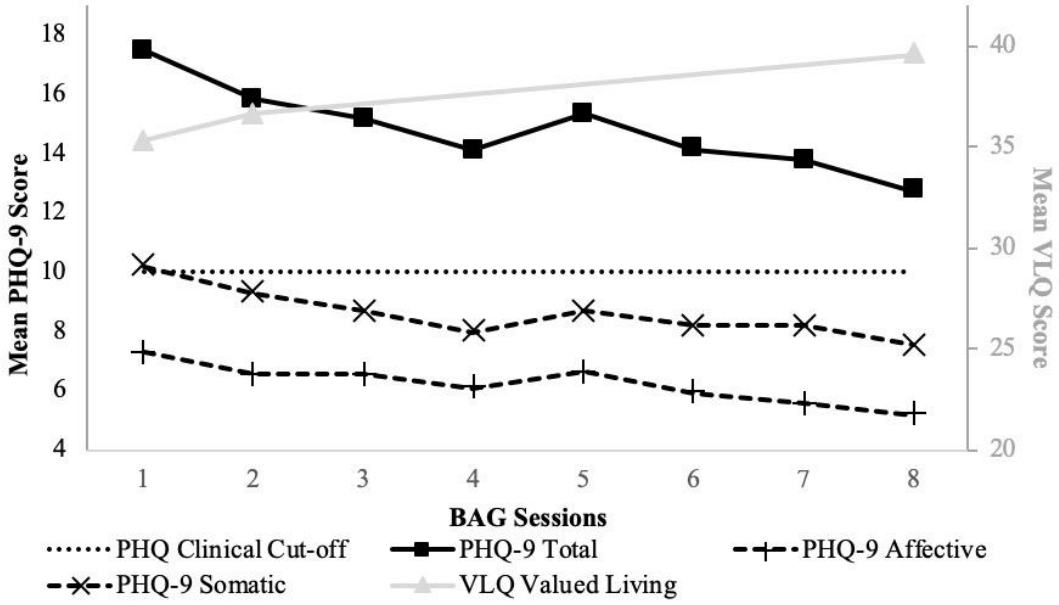


Figure 6.1. BAG Session-by-Session Scores for VLQ Valued Living and PHQ-9 Total and Symptom Cluster Scores

5.7.2 Relationship between depression and values at pre-treatment

To test the hypothesis (1) that patients will present at the start of treatment with a discrepancy between what they value and their behaviour, the Importance and Consistency subscale scores were compared. The importance of values were rated significantly higher than the corresponding level of values consistent behaviour ($t(27) = 2.784$, $p = 0.01$). Bivariate correlations further investigated the relationship between pre-treatment values levels and pre-treatment depression severity and affective and somatic symptom clusters and are presented in Table 6.3. The directions of the correlations between PHQ-9 and VLQ Importance, Consistency, Valued Living and Discrepancy pre-treatment scores were negative (with the exception of somatic symptoms relationship with VLQ Importance, Consistency and Discrepancy). However, there were no significant effects evident apart from within measure sub-scales. Interpretation of the magnitude of Pearson's *r* effect size indicated a small negative effect between severity

of depression symptoms (total and affective) and degree of *valued living* (VLQ Valued Living score).

Table 6.3. *Correlations between pre-treatment PHQ-9 and VLQ scores*

Outcome Measure (N=28)	PHQ-9 Total	PHQ-9 Affective	PHQ-9 Somatic	VLQ Importance	VLQ Consistency	VLQ Valued Living
PHQ-9 Affective	.887***					
PHQ-9 Somatic	.851***	.514**				
VLQ Importance	-.023	-.061	.026			
VLQ Consistency	-.085	-.176	.043	.149		
VLQ Valued Living	-.220	-.246	-.129	.668***	.683***	
VLQ Discrepancy	-.053	-.100	.016	-.592***	.708***	.080

Note. *$p < .05$, **$p < .01$, ***$p < .001$

5.7.3 Predicting depression change from pre-treatment values

To test the hypothesis (2) that pre-treatment levels of values will predict the amount of change in depression following treatment, a series of multiple regression analyses were performed. Pre-treatment levels of Values Importance and Discrepancy were included as predictors for changes in depression symptoms when pre-treatment depression severity was controlled for. Separate analyses were performed to investigate predicting overall depression symptoms, as well as the affective and somatic symptom clusters. Table 6.4 reports the results of the regression model for Values Importance predicting overall depression change. Once pre-treatment depression severity was controlled, Importance did not significantly improve the model, only accounting for an additional 9% of the variance in depression symptom change. The association was indicative of a small, but non-significant effect ($f^2 = .10$) in the direction of higher

overall levels of importance placed on values at pre-treatment being related to greater improvement in depression symptoms after treatment.

Table 6.4. *Multiple Regression Model for Pre-treatment Values Importance Predicting Overall Depression Symptom Change, Controlling for Pre-treatment Severity (N=28)*

	B	*SE B*	β
Step 1			
Constant	-5.29	4.60	
Pre-treatment depression score	0.57	0.26	.40*
Step 2			
Constant	-12.64	6.07	
Pre-treatment depression score	0.58	0.25	.41*
Pre-treatment Values Importance	0.11	0.06	.31

Note. R^2 = .16, f^2 = .19 for Step 1; ΔR^2 = .09, f^2 = .10 for Step 2 (p=.089). *p < .05

Table 6.5 reports the regression models for the breakdown of affective and somatic depression symptom clusters. Once pre-treatment depression severity was controlled for, Values Importance did not significantly improve the affective model, only accounting for an additional 2% of the variance in affective symptoms. The association was indicative of a very small, but non-significant effect (f^2 = .02) in the direction of higher overall levels of importance placed on values at pre-treatment being related to increased improvement in affective symptoms after treatment. However, Values Importance did significantly improve the model for somatic symptoms, accounting for an additional 15% of variance in somatic symptoms. Higher overall levels of importance placed on values at pre-treatment was associated with larger improvement in somatic symptoms after treatment, indicative of a moderate effect (f^2 = .18).

Table 6.5. *Multiple Regression Models for Pre-treatment Values Importance Predicting Affective and Somatic Depression Symptom Change, Controlling for Pre-treatment Severity (N=28)*

	B	*SE B*	β
Affective depression symptoms			
Step 1			
Constant	-2.15	1.46	
Pre-treatment depression score	0.58	0.19	.52**
Step 2			
Constant	-3.78	2.48	
Pre-treatment depression score	0.59	0.19	.53**
Pre-treatment Values Importance	0.02	0.03	.14
Somatic depression symptoms			
Step 1			
Constant	-3.40	3.10	
Pre-treatment depression score	0.60	0.30	.37
Step 2			
Constant	-8.85	3.72	
Pre-treatment depression score	0.58	0.27	.36*
Pre-treatment Values Importance	0.09	0.04	.39*

Note. Affective symptoms $R^2 = .27$, $f^2 = .37$ for Step 1; $\Delta R^2 = .02$, $f^2 = .02$ for Step 2 (p=.423). Somatic symptoms $R^2 = .14$, $f^2 = .16$ for Step 1; $\Delta R^2 = .15$*, $f^2 = .18$ for Step 2 (p=.030)*. *$p < .05$, **$p < .01$

Table 6.6 reports the results of the regression model for discrepancy in *valued living* predicting overall depression change. Once pre-treatment depression severity was controlled, Discrepancy did not significantly improve the model, only accounting for an additional 8% of the variance in depression symptom change. The association was indicative of a small, but non-significant effect ($f^2 = .09$) in the direction of greater discrepancies in *valued living* (represented by larger negative scores) at pre-treatment being related to greater improvement in depression symptoms after treatment.

Table 6.6. *Multiple Regression Model for Pre-treatment Valued Living Discrepancy Predicting Overall Depression Symptom Change, Controlling for Pre-treatment Severity (N=28)*

	B	*SE B*	β
Step 1			
Constant	-5.29	4.60	
Pre-treatment depression score	0.57	0.26	.40*
Step 2			
Constant	-5.85	4.48	
Pre-treatment depression score	0.55	0.25	.39*
Pre-treatment Values Discrepancy	-0.07	0.05	-.28

Note. R^2 = .16, f^2 = .19 for Step 1; ΔR^2 = .08, f^2 = .09 for Step 2 (p=.122). *p < .05

Table 6.7 reports the regression models for the breakdown of affective and somatic depression symptom clusters. Once pre-treatment depression severity was controlled for, *valued living* discrepancy did not significantly improve the affective model, only accounting for an additional 2% of the variance in affective symptoms. The association was indicative of a very small, but non-significant effect (f^2 = .02) in the direction of greater discrepancies in *valued living* (represented by larger negative scores) at pre-treatment being related to increased improvement in affective symptoms after treatment. However, again Discrepancy did significantly improve the model for somatic symptoms, accounting for an additional 13% of variance in somatic symptoms. Greater discrepancies in *valued living* at pre-treatment were associated with larger improvement in somatic symptoms after treatment, indicative of a moderate effect (f^2 = .15).

Table 6.7. *Multiple Regression Models for Pre-treatment Valued Living Discrepancy Predicting Affective and Somatic Depression Symptom Change, Controlling for Pre-treatment Severity (N=28)*

	B	*SE B*	β
Affective depression symptoms			
Step 1			
Constant	-2.15	1.46	
Pre-treatment depression score	0.58	0.19	.52**
Step 2			
Constant	-2.24	1.48	
Pre-treatment depression score	0.57	0.19	.51**
Pre-treatment Values Discrepancy	-0.02	0.02	.13
Somatic depression symptoms			
Step 1			
Constant	-3.40	3.10	
Pre-treatment depression score	0.60	0.30	.37
Step 2			
Constant	-4.22	2.95	
Pre-treatment depression score	0.61	0.28	.37*
Pre-treatment Values Discrepancy	-0.06	0.03	-.36*

Note. Affective symptoms R^2 = .27, f^2 = .37 for Step 1; ΔR^2 = .02, f^2 = .02 for Step 2 (p=.463). Somatic symptoms R^2 = .14, f^2 = .16 for Step 1; ΔR^2 = .13*, f^2 = .15 for Step 2 (p=.049)*. *p < .05, **p < .01

5.7.4 Effect of BAG treatment on values

To investigate the hypothesis (3) that importance of values would remain stable, whereas the discrepancy between importance and action would decrease during BAG treatment, separate paired t-tests were conducted on each VLQ dimension score pre- and post-treatment. Significant changes were not evident for Values Importance (t(27)=-0.23, p=0.82), Values Consistency (t(27)=-1.23, p=0.23), Valued Living (t(27)=-1.48, p=0.15) or Discrepancy (t(27)=-1.01, p=0.32).

As predicted, minimal change was observed in Values Importance. However, contrary to the hypothesis, BAG did not significantly increase the degree with which patients engaged with *valued living*. Pre-post effect sizes taking into account the

correlation of scores indicated no effect for change in Importance. Small, unreliable effect sizes were apparent for changes in Consistency and Valued Living scores. Analysis of individual outcomes found only 14% of patients showed reliable change in *valued living*, with 7% showing reliable deterioration in *valued living*. The remaining 79% experienced no change in their level of values-consistent behaviour.

5.7.5 Valued living as a mediator of depression outcome during BAG

Due to the low sample size, testing the hypothesis that change in *valued living* would mediate change in depression outcomes during BAG (4) was limited to exploratory analyses. In addition, the minimal change in valued living over the course of BAG therapy demonstrated by the previous analysis (section 6.3.4) annulled the possibility that changes in valued living would mediate the observed reductions in depression. Therefore, the exploratory mediation analyses are presented in Appendix H.

5.8 Discussion

The purpose of this study was to investigate the relationship between values and depression during routinely delivered BAG treatment, to determine whether increased values-based behaviour produces associated reductions in depression. This goal was achieved by first examining the relationship between patient 'values profiles' (incongruent values and behaviour) and depression presentations at the start of treatment. Second, pre-treatment values profiles were explored as predictors of post-treatment BAG depression outcomes. Third, pre-post changes in values profiles were compared, to establish whether BAG affected patients' engagement with values-based behaviours. Finally, exploratory mediation analyses looked at the role of increased values-based behaviour as a potential contributor to depression reductions after BAG (presented in Appendix H).

5.8.1 Summary of results

Patients presenting for BAG treatment reported a discrepancy in their behaviour and their life values. This is seen as a key insight during BA and the goal of treatment is to close the value-behaviour gap. However, neither the magnitude of the discrepancy or the extent to which patients placed importance in life values was related to severity of depression or separate symptom clusters (Table 6.3) and this is therefore a challenge to BA theory and practice. In addition, pre-treatment values profiles did not predict overall post-treatment change in depression (Tables 6.4 & 6.6). In contrast, there were associations between values sub-scales and changes in depression symptom clusters. Higher levels of importance placed on values (Table 6.5) and greater discrepancies between values and behaviour (Table 6.7) predicted larger improvements in somatic depression symptoms. However, the BAG intervention did not produce any significant pre-to-post changes in *valued living*, which is a core aim of the groups. As change in *valued living* was minimal, mediation analyses exploring the potential link to observed depression reductions were redundant. Analyses were therefore conducted on an exploratory basis and provided in Appendix H.

5.8.2 Values as a mediator of change

The findings largely failed to support the hypotheses and BA theory. It is proposed in BA that defining and articulating values is integral to the process of increasing activation to produce change that is meaningful to a patient (Ryba et al., 2014). However, unlike the mediating effects of *valued living* seen in ACT-based treatments (Bramwell & Richardson, 2018; Gloster et al., 2017; Lundgren et al., 2008), the exploratory findings did not find evidence to suggest increased values-based behaviour mediate subsequent reductions in depression symptoms following BAG. As only 14% of patients experienced reliable change in *valued living* over the course of

treatment, the majority of patients who experienced reductions in depression did so without improving their values-consistent behaviour.

It is not clear why more changes in valued living were not observed. Interpreting the results is complicated by the small sample size and ICC cluster effects. Whether the non-significant results reflect a true null effect or are the product of insufficient statistical power or a poor measurement tool is impossible to establish from this study. As a result, these findings should be considered preliminary, but worthy of prompting further investigation.

5.8.3 Values and depression symptom clusters

Participants appeared to present at the start of treatment with a discrepancy between what they valued and how they were behaving (Wilson & Murrell, 2004). Although these results suggest that values levels prior to treatment do not relate to variance in depression symptoms, they did predict the amount of change in somatic symptoms (Hayes et al., 2010). Given that the underlying change process of BA posits activation in the face of low mood (Martell et al., 2010), those who ascribe higher levels of importance to values or have lower levels of values-consistent behaviour appear better able to implement this approach. Drawing awareness to ongoing behaviour that is inconsistent with values provides opportunity for change (Hayes, Strosahl, & Wilson, 2012). Patients who perceive their behaviour to be inconsistent with their values may benefit from a greater motivation to change (Wersebe et al., 2017). Likewise, those who place a great importance on the functional outcome of certain behaviours receive stronger response-contingent positive reinforcement when they re-engage with those activities (Doi et al., 2016). The current study suggests the resulting outcome is that patients then experience the physical benefit of reduced somatic symptoms. So, although patients are initially still feeling the effects of the affective symptoms (such as low mood), they gain benefit in terms of improvements in the more physical, somatic

symptoms. This finding would be indicative of the 'outside-in' change process exemplified during BA treatment (Curran et al., 2012), but specific to the physical symptomology of depression.

5.8.4 Implications

The main implication of the present findings is that the link between values and depression needs to investigated in a larger sample size with adequate power in order to be able confidently to interpret the findings. In addition, the uncertainty in the findings and variability in responses suggest that values may only change in subtle ways. Therefore, measures of values need to be suitably sensitive to detect small changes and shifts in values and behaviours. It has been suggested that the VLQ may not be an ideal measurement tool for capturing such changes (Baer, 2010; Wersebe et al, 2017). Research on values in BA treatments would benefit from first focusing on effectively measuring and establishing the extent to which values-based work integrated into treatment can affect values-consistent behaviour. Once the process of valued-living change is better established, investigations into its role as a mediator will become clearer.

In the meantime, one should be careful about assuming that values-based behaviour is a key mechanism in depression and a contributing factor to depression changes during BA. If changes in *valued living* can be detected in a larger sample or with a different measure, then *valued living* should not yet be dismissed as a potential contributor to the process of change. However, if the lack of change in *valued living* is replicated, then it would demonstrate that realigning behaviour with values is not a key mediator of outcome during BAG, meaning that other mechanisms need to be considered.

There are also implications relating to the association between pre-treatment values and changes in somatic depression symptoms. If the finding can be replicated

that pre-treatment levels of values importance and size of discrepancies in *valued living* each predict post-treatment somatic symptoms change, it would point to potential treatment-matching theories that could be tested. Patients who have a high level of importance in what they value in life or who begin treatment with a greater discrepancy in their *valued living* may be more receptive to the treatment techniques used in BA. Although preliminary, this theory implies that the trajectories of symptom change may differ between symptom clusters during BA. To expand on this idea, it would be informative to investigate whether changes in somatic symptoms precede changes in affective symptoms, or even mediate subsequent changes in affective symptoms. Such findings might be useful in identifying those patients who will respond the most to BA treatment strategies (e.g.., high somatic symptoms benefitting from behavioural therapy, and high cognitive symptoms benefiting from cognitive therapy).

5.8.5 Limitations and future research

The large number of findings in this study that do not support the hypotheses might in part be accounted for by a number of methodological limitations. First, as referred to throughout the study, the smaller sample size resulted in a lack of statistical power. Some outcomes were affected by ICC effects from correlated outcomes within the group treatment. Large sample sizes are particularly recommended for investigations of mediators (Baron & Kenny, 1986). The large number of null findings therefore have to be considered uncertain, as true effects may not have been captured. Attempts were made in the study design to ensure there was adequate power to detect effects, but the service issues that arose were uncontrollable (i.e., patient referrals siphoned into a large research trial on depression, cancelled BAG groups during study period, service decision to stop offering BAG as a treatment option). The difficulties experienced highlight the difficulties of conducting research in routine practice settings (Barkham, Stiles, Lambert, & Mellor-Clark, 2010). The increased external validity often comes at

the expense of experimental rigour. Research going forward needs to investigate whether these results are replicated in a larger sample with sufficient statistical power to provide clarification.

Second, results were solely based on self-report measures. The use of self-report measures on their own has been criticised. Patients can be susceptible to demand characteristics or a tendency to complete measures according to what they think the therapist wants (Tourangeau & Yan, 2007). In particular, the validity and sensitivity of the VLQ as a measurement tool has been questioned. It has been suggested the strength of the VLQ is in being a clinical tool to aid reflections about patient's values, rather than an assessment tool for change (Baer, 2010). There were no checks of whether the VLQ was clinically well explained to patients, and the variation in facilitators delivering the groups may have resulted in different groups getting different levels of explanation. The evidence of clustering in some of the VLQ outcomes would support this assertion that different BAG delivery groups may have had different interpretations about how to fill in the measure. Although the VLQ has undergone a degree of psychometric validation, the capacity of the measure to detect changes in a clinical sample undergoing treatment is not clear. To be able to investigate values in treatment studies effectively, future research needs to clarify the suitability of the VLQ. Behavioural measures (such as percentage of time engaged in valued activities; Hayes et al., 2010) should be used in conjunction with self-report measures.

Third, the study was conducted on routine practice delivery, so there was no control group and the sample were not screened for suitability. While this enabled a high level of external validity, patient characteristics such as complex co-morbidity or unsuitability for treatment might have confounded the results. Within IAPT services, waitlists for one-to-one treatment can be very long, whereas access to BAG treatment was available every two months. Some patients may have been referred or self-referred

to BAG when they were not entirely suitable, in order to access the first available treatment. Furthermore, the lack of a control group means that the effects observed cannot be compared against the general process and trajectory of values in relation to depression in patients who did not receive BAG.

Fourth, there was no measure of patient treatment compliance, especially in relation to implementing between-session work in relation to values. If BAG did not result in any changes in *valued living* (as the null results suggest), is that due to the treatment methods not being effective or is it because patients did not comprehend and adequately implement the strategies?

Finally, although there were no significant findings, the mediation analysis could not adequately take into account temporal change in the mediator and outcome variables (Kazdin, 2009). A causal relationship cannot be established if both variables are measured at the same time-points. Any change in the mediator could be a result of the improvement in symptoms, rather than an active mechanism to produce clinical outcome. To be confident in conclusions about mediators, change in the mediator needs to be shown to precede change in the symptoms. It also needs to be demonstrated that the mediator has had sufficient time to change. This was complicated by the repeated-measures mediation model used, meaning the independent variable had to be expressed via the use of change scores. Both variables had to be measured at two separate time points (to enable a change score), while still allowing for a time lag between measurements of the mediator and outcome variable. Although attempts were made to collect temporal measurements, this was difficult to implement in a relatively short treatment over the space of 8 weeks. Consequently, as VLQ measurements were not taken at every session, the change scores had to be calculated from pre and post scores. Therefore, all variables were measured at the same time points. Future research into treatment mediators should ensure the sequence of variable measurements allows

temporal relationships to be established, in order to be confident in identifying causality. In short-treatment protocols in particular, mediation analyses would be made easier if both the mediator and outcome variable were measured at every treatment session.

5.8.6 Conclusion

This study has provided a preliminary investigation into the relationship between what a patient values and depression outcomes during BAG treatment - the first such study to do so. The values-based work that has been integrated into BAG does not seem to produce meaningful changes in *valued living*. Nor do *valued living* changes then mediate depression outcome at the end of treatment. However, the study was subject to several methodological limitations. Currently, it should not be assumed that behaviours that are specifically values-based play a critical role in depression reductions after BA treatment. Further research should build on these exploratory findings to clarify whether BA can produce change in *valued living*, before dismissing it as a crucial mechanism. However, if these results are replicated in larger samples, it would suggest attempts to increase *valued living* as they are currently delivered are not adding any clinical benefit to BA treatments, and other mechanisms need to be considered.

CHAPTER 7

General Discussion

The objective of the final chapter is to distinguish results and themes from the four empirical studies presented in the thesis, and discuss the overall clinical, theoretical and organisational implications. First, a summary of the thesis aims will be provided, before an overview of the findings from each chapter are summarised. Second, an interpretation of the results will be discussed in relation to 1) understanding of depression, 2) behavioural activation (BA) as a treatment for depression, 3) BA treatment mechanisms and 4) predicting and reducing stasis. Third, the implications of the findings will be outlined, divided into recommendations for clinical practice, services and policy, and theory. Fourth, strengths and limitations of the thesis will be reviewed. Finally, recommendations for future research will be made, before summing up with the thesis conclusions.

6.1 Summary of thesis aims

This thesis sought to investigate stasis outcomes in routine practice after a BA group treatment for depression and also try to reduce stasis outcomes for depressed patients engaging in BA treatment. First, the lack of clarity in the evidence base for group-based BA interventions was addressed with a meta-analysis of group BA treatment outcomes in trial (efficacy) and naturalistic (effectiveness) contexts (Chapter two). Second, a review of stasis as a concept, prevalence of stasis within treatment completion outcomes and associated factors was conducted to define a metric for capturing stasis (Chapter three). Third, the stasis metric was applied to an empirical analysis of BA routine outcome data to investigate treatment response after stepped-care delivery of BA interventions (Chapter four). The effect of intervention intensity, format

and duration on stasis outcomes was explored, in addition to stasis risk predictors. Fourth, an intervention to enhance an existing group BA treatment was developed and tested to see if treatment retention and outcomes could be improved (therefore reducing rates of stasis; Chapter five). Finally, a mediation study was conducted to explore what mediates depression symptom cluster reductions (affective and somatic symptoms) after group BA (Chapter six). Increased behaviour in accordance with life values was a proposed and evaluated as an active change mechanism.

6.2 Summary of thesis findings

The first empirical study (Chapter two) showed group BA produces aggregated post-treatment depression outcomes that are superior to controls (g = 0.72) and equivalent to other active therapies (g = 0.14), which are maintained at follow-up. Group BA is effective both in trial contexts (g = 0.82) and when translated to naturalistic settings (g = 0.63), albeit with a slightly smaller effect. It is also acceptable to patients, with a drop-out rate of 14-16%.

While summary outcomes evidence the positive effects of interventions, individual patient outcomes highlight that many patients do not experience any change in their symptoms, experiencing a stasis outcome. However, processes for assessing therapy outcomes are not designed to capture minimal change. Chapter three demonstrated the disparity in the lack of research on stasis as a concept, in contrast with the prevalence of the issue. The review determined the reliable and clinically significant improvement (RCSI; Jacobson & Truax, 1991) method would be best to capture and investigate the stasis phenomenon.

Analysis of stepped-care BA treatment response in the second empirical study (Chapter four) demonstrated BA was effective at reducing depression (in 4-9 sessions) regardless of format, with larger effects seen for more intensive versions. Applying the

stasis metric identified considerable rates of stasis across BA treatment delivery, with those at risk of a stasis outcome distinguishable from improvers after two sessions. Risk of an overall BA stasis outcome was predicted by attending fewer sessions, greater impaired functioning prior to treatment and less severe depression.

The third empirical study (Chapter five) showed group BA could be enhanced using a theory-driven treatment augmentation ('if-then' plans) to produce fewer stasis outcomes and greater reductions in depression and anxiety. However, a data-driven augmentation (attendance psychoeducation) did not affect group BA drop-out rates.

The final empirical study (Chapter six) found that despite patients appearing to present for treatment with a discrepancy in what they value and how they are behaving, the magnitude of the discrepancy was not related to depression severity or predict group BA treatment outcome, nor did valued-living increase as a result of treatment. Subsequent, exploratory mediation analyses reflected that changing behaviour to reduce the discrepancy did not mediate reductions in depression after group BA. However, that study was underpowered to reach a definitive conclusion that there is no such effect.

6.3 Interpretation of results

Discussion around how the results from this thesis can be interpreted has been divided into four sections; 1) first, what the findings say about depression, 2) second, what the results show regarding the effectiveness of BA as a treatment for depression, 3) third, understanding about how BA works, and 4) fourth, what can be learned about depression stasis and whether it can be reduced?

6.3.1 Understanding of depression

The overall findings from the thesis support the role of behavioural inactivation and avoidance in maintaining low mood. This interpretation of the cycle of depression has links to both the behavioural and learned helplessness theories of depression

(Ferster, 1973; Seligman, 1973). Both theories postulate depression is the result of a learned response that is shaped by our environment and driven by limited attempts at change. By attempting to reverse these maintaining factors, the strategies used in the four studies were able to alleviate depression symptoms. Making changes to patients' environments had beneficial effects on their depression symptoms, showing our environment plays a pivotal role in shaping depressive behaviours. The behavioural theory that depression is maintained by removal of positive reinforcement for non-depressive behaviour was reflected in the results from Chapter six. Patients presented at treatment with discrepancies between their behaviour and life values, suggesting an impaired relationship between behaviour and positive reinforcement in depression (Lewinsohn & Graf, 1973). However, variability in depression severity was not accounted for the extent that behaviours had become incongruent from life values. This suggests that detachment from life values is a by-product of the initial development of depression, but is not a mechanism that contributes to severity of symptoms.

6.3.2 Effectiveness of BA for depression

6.3.2.1 Expanding the evidence base

The overall findings drawn from each study in this thesis have added to the evidence base for the effectiveness of BA as a treatment for depression when used in routine practice. The Salkovkis 'hourglass model' (1995) for psychological therapy research advocates the development of an evidence base by using several research phases. Initial theoretical development and uncontrolled evaluation should be followed by controlled clinical trials, before finally conducting real world research and service implementation. These results therefore fit within the final research phase for the BA evidence base, by demonstrating the effectiveness of real-world delivery of BA and treatment optimisation. BA, delivered in both simple or complex versions and individually or in groups, produces reductions in depression, anxiety and impaired

functioning. Effective treatment in a variety of delivery formats supports BA as a flexible and adaptable intervention (Sturmey, 2009). Meanwhile, half of all treatment improvers can be identified after 4-6 sessions of BA, with the majority needing 7-9 sessions to experience improvement, showing that BA interventions only need to be brief.

The range of improvement rates seen across the studies (17-48%) are reflective of the variable rates reported for psychotherapy interventions in routine services (Hansen et al., 2002). However, in contrast to typically reported 5-10% who may get worse after therapy, very little deterioration was observed after all versions of BA (0-3%). BA as a therapeutic model appears to cause patients minimal harm (even if they do not benefit). The lower than usual deterioration rate might be explained by the simple principles that BA is based on. It may be that potential harm as a result of the intervention is reduced due to the behavioural focus on external contexts. Opportunities to become fixated on negative internal cognitions and dwell on past experiences are minimised. This focus may also help people to remain more stable when they would otherwise get worse despite treatment (i.e., get classified as a stasis outcome, rather than deterioration).

6.3.2.2 Group delivery of BA

It is often thought that individual therapy is more effective than group delivery, although the clinical relevance of the difference has been debatable (Cuijpers & Straten, 2008). The present evaluation of BA delivered in groups has provided clarification that group delivery is not detrimental to BA outcomes. Meta-analytic findings demonstrated the equivalence of group BA with other active therapies (Chapter 2), while the real-world comparison of high intensity (HI) one-to-one versus group BA produced very similar outcomes (Chapter 4). These finding show a direct therapeutic bond is not necessary for a good outcome in BA. It appears patients can draw equal benefit from the

group support and normalising of problems (Yalom & Leszcz, 2005). There have also been concerns that groups are not as acceptable to patients, with many patients preferring one-to-one over group treatment (Brown et al., 2011). However, group BA retained the same amount of people in treatment than the HI one-to-one version. This comparable rate of retention suggests that once patients attend the group treatment, it does not put them off. The BA group protocol aims to utilise group members' sense of commitment to each other to encourage homework completion and attendance, which may facilitate treatment retention. However, during data collection for Chapter five, several BA groups had to be cancelled due to low referrals (influenced by several service issues). Therefore, the greater challenge may actually be in setting up services to promote group treatments and then getting people to consider attending a group in the first place.

In light of apparent acceptability and comparable outcomes with one-to-one treatment, group BA provided a number of service benefits. Treatment was accessible to more patients at a time and, in general, more frequently (a new group started every eight weeks, as opposed to waiting until a therapist has a one-to-one slot free). In addition, treatment was delivered over a more condensed delivery period (8 weeks compared to 14-16 weeks). As a result, effects of BA could be disseminated more widely and faster in the group context. Taken together, the results of this thesis show that group BA can be a valuable treatment option for services, providing an effective and scalable way to treat depression.

6.3.3 BA treatment mechanisms

6.3.3.1 Distinct 'outside-in' approach

In the context of psychological therapies for depression, BA is quite distinct. The 'outside-in' BA philosophy and clinical approach is generally at odds with how most therapies approach depression, as such therapies tend to emphasis an 'inside-out'

approach (Curran et al., 2012), of using therapy to feel better, to then make changes. Unlike many psychotherapies, BA is not overly concerned with focusing on 'why' there is a problem. BA is immediately directed at 'how' the problem can be solved through 'doing' rather than 'talking' within a supportive and compassionate therapeutic relationship. When experiencing the desperation and hopelessness of depression, approaching the problem with an obviously pragmatic and problem-solving stance is likely to be easier to engage with for many people (Coffman et al., 2007). As shown by these findings, even simple activation can be effective for people with depression. Reliable improvements in depression can start to be seen after only two sessions. This progress supports previous findings that show patients report behaviour change as being a crucial element of therapy for their recovery (Finning et al., 2017). The benefits produced from actively only trying to make behavioural changes reiterate the standpoint that cognitive work is not necessary for change (Jacobson et al., 1996; Longmore & Worrell, 2007). Therefore, BA not only provides an alternative way to approach depression, but it does so in a way that makes better use of resources than other treatments. BA techniques are simple and parsimonious, enabling shorter training and easier dissemination. Both these factors show BA can be a more cost-effective treatment option (Richards et al., 2016).

6.3.3.2 Key role of activation

Given the core role of behavioural change in helping people live a more meaningful life, increased activation is thought to be a key mechanism in how BA works. This thesis further explored how the 'doing' aspect of BA produces reductions in depression. Barriers to change have been shown to come from difficulties in engaging with the homework (Barnes et al., 2013). Facilitating plans for completing the homework tasks using implementation intentions produced greater reductions in depression and fewer stasis outcomes. The implementation intention plans encouraged

patients to pre-empt barriers and plan how they would respond. This finding supports the critical role of between-session activation for treatment response and suggests pre-planning can help overcome some of the barriers to homework engagement.

Despite both low intensity (LI) and high intensity (HI) versions of BA emphasising between-session activation strategies, the larger effects for more intensive versions of BA imply there are other treatment factors that influence depression change in BA. Looking at the differences in the protocols, the differential effects could be due to level of training of the therapist, longer treatment or the additional treatment components in HI BA. The final study looked at whether the additional contextualisation of the activation strategies around patient values was able to explain reductions in depression symptoms after group BA. It was proposed that group BA enables patients to engage in increased values-based behaviour, which in turn reduces depression symptoms (thereby the additional values work in HI BA enables more targeted values-based activation, enabling larger treatment effects). However, increased values-driven behaviour did not appear to be as key as proposed, as it did not mediate depression changes and methodological limitations hampered interpretation. If additional values work does not provide a treatment benefit over standard activation, it calls into question the inclusion of values in the HI BA treatment protocol. BA is built on the premise of parsimony, so redundant elements would seem contradictory. The next step would appear to be a component analysis comparing full protocol values-driven BA versus BA with the values aspect removed (Bell, Marcus, & Goodlad, 2013). These results highlight that more investigation is needed to establish BA mechanisms of change.

6.3.4 Predicting and reducing stasis

6.3.4.1 Scale of the stasis issue

Historically, reporting of treatment outcomes have not always clearly quantified the amount of patients not experiencing any change. Instead the focus has been on rates of responders, making it difficult to confidently distinguish when and why patients have not benefitted from treatment. Given the varied and heterogenous presentations of depression, patients experiences of treatment will be different. It is important the methods for evaluating treatment outcomes are able to adequately capture that. Following the lead of recent investigations into routine practice treatment response, the use of a defined stasis metric has enabled such outcomes to be clearly quantified throughout this body of work (Delgadillo, Moreea, et al., 2016; Kellett et al., 2017; Pybis et al., 2017). This was implemented through the use of an existing standardised system for establishing reliable change (Jacobson & Truax, 1991). The reliable change criteria were applied to identify the limits of change that could construe no meaningful benefit, using the psychometric properties of the measure used. Consequently, the metric is usefully compatible with the current processes for capturing treatment outcomes. Fitting the stasis classification within the reliable and clinically significant change (RCSC) method creates a graded continuum to capture differing extents of treatment benefit (or lack of). Rather than being restricted to a binary choice, outcomes were better distinguished from those who had fully recovered, those who had had some sort of worthwhile gain, those who had not had any, and those who had got worse.

Despite the empirical support for BA treatments, the outcomes reported here confirm that stasis is still a common occurrence after routine treatment (in the Improving Access to Psychological Therapies [IAPT] service model). IAPT services are set targets to achieve 50% recovery rates, which quarterly IAPT performance data suggests is achieved (IAPT, 2011; NHS England, 2016). The rates of stasis in the present findings consistently appear to exceed 50%, reaching as high as ~66% of patients. These figures suggest IAPT services do not always meet their outcome targets

and outcomes are known to vary from service to service (Clark et al., 2017). However, critiques of IAPT suggest that independent findings like these show there is a discrepancy between what information IAPT releases (50% recovery rate achieved) and what rates are actually being produced more generally (Scott, 2018).

In the wider context of mental health services, recovery rates are often not recorded (Clark et al., 2017), and when they are, they often fall short of the rates reported by IAPT services (Hansen et al., 2002). It is possible the state of stasis in other types of services is even worse. It is important, however, to consider that not all stasis outcomes may necessarily represent a negative outcome. Stasis could occur in cases where a patient would have been worse off without treatment (i.e., although their symptoms did not change, treatment stopped them deteriorating) (Cuijpers, 2018). BA stasis rates may, to some extent, be inflated as a function of prevented deterioration cases. Nonetheless, the current results reiterate the scale of the problem, with minimal improvement in outcomes since the seminal Hansen et al. (2002) paper, which highlighted the issue of stasis outcomes across disorders nearly 20 years ago.

6.3.4.2 Understanding stasis

Chapter three reviewed the factors associated with stasis outcomes. In light of inconsistencies in the predictors implicated, the review concluded that the identification of stasis predictors for specific treatments could provide more clarity. The second empirical study (Chapter four) identified three variables that predict risk of having a stasis outcome after BA treatment - greater impaired functioning, less severe depression, and attending fewer sessions. In contrast to what is often assumed, demographic factors (such as age and gender) did not affect risk of stasis. Depression severity has previously been identified as a treatment response predictor, although with inconsistent directions of effects. Some findings have tended to indicate that higher levels of baseline depression are associated with poorer outcomes (Delgadillo, Moreea,

et al., 2016; Thibodeau et al., 2015), whereas in line with the current findings, others have found lower levels of depression predict nonresponse. The present finding here may be attributable to the suggestion that BA can have more beneficial effects for severe depression in comparison to cognitive behavioural therapy (CBT; Coffman et al., 2007).

Meanwhile, the association between an increased risk of stasis and greater impaired functioning, may reflect the between-session engagement required in BA. Those with more impaired functioning might find it harder to implement the activation strategies, and therefore get less benefit from treatment. Similar variation in outcomes when working with patients with more functional impairment have been seen for other interventions with a focus on implementing coping strategies in between sessions (Firth et al., 2015).

The strongest predictor of stasis was attendance at fewer sessions, showing that treatment retention is crucial (Firth, Barkham, Kellett, et al., 2015; Saxon et al., 2017). When patients attend a full course of treatment, they are more likely to experience some form of improvement. Furthermore, probability of end-of-treatment improvement or stasis was distinguishable after two sessions of BA, (regardless of format or intensity), showing that early change can be a useful indication of likelihood of a good outcome.

6.3.4.3 Reducing stasis outcomes

Although the rates of stasis can make it seem that the prospects for depression treatment are rather bleak, a promising outcome from this research is that there are ways to intervene. Two approaches were tested in Chapter five - one targeted at reducing stasis outcomes directly, and one aimed at improving treatment retention to reduce patients risk of stasis (as suggested by the findings in Chapter four). The results showed stasis can be directly reduced through a 'theory-informed' treatment enhancement. A 25% reduction in stasis was achieved by augmenting the existing group BA treatment.

The improved outcomes came from a low-cost strategy (implementation intentions) to enhance patients' ability to put their homework into action.

The indirect 'data-driven' approach was less successful. The attempt to increase treatment retention used psychoeducation about treatment outcomes and attendance to align patient expectations with what to expect from therapy. However, the augmentation had no effect on treatment attendance. Given the importance of treatment attendance for reducing stasis (shown in Chapter four) and the failure to improve retention (in Chapter five), it is important to consider why patients did not complete treatment. Looking at the attendance at BA treatment across the studies in the thesis, the rates of full treatment completers were largely sub-optimal. There are several ways to interpret the patterns of BA attendance seen in these studies. In the case of LI BA (where attendance was the poorest), treatment termination could be because the patient has been stepped up to a higher intensity treatment due to lack of improvement. On the other hand, it would be more concerning if termination was regularly occurring through patients not attending, without communicating with their therapist/the service. If patients disengage from LI treatments, they risk being 'lost' from the system and miss out on additional support that might be available to help them.

Reasons why patients drop-out could be because they do not take to the BA model (in which case they need to be given a different treatment option). Alternatively, a study of reasons for non-attendance in IAPT services found the themes that emerged were less related to the treatment models, but mainly based around the rigid way services were set up and communication with patients (Marshall et al., 2016). Inflexibility with session arrangements and poor communication about the options available may be making IAPT services quite difficult for patients to navigate. The consequence may be that, while they want to engage with treatment, they are not always able to fit around the service set-up. This interpretation could explain the generally good

partial attendance rates, despite the poor full treatment completion. In summary, the results from Chapters four and five highlight that more varied strategies are needed to improve treatment retention as a way to reduce stasis.

6.4 Implications

6.4.1 Clinical practice implications

The findings have several implications for the clinical practice of BA for depression. First, the core behavioural element of BA takes place between-sessions, with the engagement with activation strategies. The beneficial effect of 'if-then' plans for setting homework, demonstrated by this thesis, highlights that clinicians should emphasise the importance of completing the homework tasks. They should try to support patients with how they will implement their activation strategies and overcome barriers to change to encourage sustainable change.

Second, the consistent finding that treatment attendance was the best predictor for reducing risk of a stasis outcome should be disseminated clearly to patients. Although psychoeducation at the start of treatment was not able to increase treatment retention, efforts should still be made to promote attendance throughout the treatment process. The longer a patient remains in a course of treatment, the more likely it is they will benefit from treatment. Even if they complete treatment without seeing improvement, remaining within the system will enable them to be better placed to receive additional support for their depression. Otherwise, they risk being left isolated and alone with continued depression symptoms. Communicating this concept to patients could enable them to navigate mental health systems more effectively and access the right treatment to help them.

Finally, clinicians should utilise outcome feedback to guide treatment. Stasis patients can be distinguished from improvers after as few as two sessions. If indications

of change are not evident after a few sessions, clinicians should consider what could be done differently. Earlier intervention could prevent later stasis outcomes, and as a result, reduce dejection about future treatment options.

Although these findings are specific to BA for depression, there is a wider picture of potential implications in terms of the uses of BA for helping people live a meaningful life. BA has been adapted for use with excessive worry (Chen, Liu, Rapee, & Pillay, 2013), substance misuse (Daughters et al., 2008), smoking cessation (MacPherson et al., 2010), alcohol use (Reynolds, MacPherson, Tull, Baruch, & Lejuez, 2011), and increasing exercise for people with diabetes (Schneider et al., 2016). The underpinning principles of BA (activation strategies, reduction of avoidance, positive reinforcement, between-session tasks) will be relatively standard in the treatment for these conditions. The present recommendations for BA therefore may also be applicable to other patient populations.

6.4.2 Service and policy implications

Pulling together the conclusions about group BA as a treatment for depression and how to tackle stasis, also has implications for service provision. There is a gap between the demand for depression treatment and the resources available to provide it (Shidhaye et al., 2015; World Federation for Mental Health, 2012). In addition, this thesis has shown even when patients access treatment, their symptoms are not guaranteed to improve. The longer they have to wait for treatment, reduces the likelihood of a positive post-treatment outcome (Clark et al., 2017). Taking these issues into consideration, points to turnover as critical to improving the situation. First, if cognitive work is not necessary for a beneficial outcome (as these outcomes from just behavioural strategies suggest), then spending a surplus of sessions on it in later phases of treatment directs resources away from other patients who are waiting for treatment. Behavioural change is key to outcome so treatments should emphasise the behavioural

elements. Phone applications could be useful tools for delivering and supporting the use of techniques such as implementation intentions to facilitate engagement with behavioural changes.

Second, the BA groups used in these studies resulted in treatment effects wider and quicker than one-to-one versions. If enough patients are referred and remain in treatment, group treatments can also be a more efficient use of resources (i.e., disseminating treatment to multiple patients at once). Services could utilise these benefits of groups to increase turnover. Services could also consider adopting the group BA model for one-to-one versions of BA, and delivering the protocol over a more condensed period. Delivering behaviourally focused treatments and utilising group formats/more condensed treatment would enable increased turnover and could reduce the time that patients spend on waiting lists. If more people are able to get treatment quicker, they will in turn have a better chance of responding. This recommendation supports the stepped-care model used by IAPT services, which to some extent embodies this approach with differing intensities of treatment. However, more could be done to target turnover. For example, group BA could be offered to patients while on waiting lists for individual assessments and treatment.

The issue of stasis also needs to be tackled on a wider scale. The development of new treatments requires considerable time and financial costs, yet are unlikely to greatly improve on the options already available (Cuijpers, 2017). Alternatively, improving the treatments that are already available provides a more practical way to tackle stasis. The evidence base is already established, therapists are trained in the models, services are set up to deliver the intervention, and treatment is already rolled out in routine practice. While the idea of a new, more effective treatment is appealing, we should not let that detract from what we already know works. The improved stasis outcomes observed here had an immediate impact for real-world patients and were achieved over a relatively

short time-frame, using low-cost strategies and minimal resources. These results advocate similar approaches as a scalable way to improve outcomes. IAPT services need to be able to look beyond simple 50% recovery notions. Even if half of patients benefit from the therapy they receive, that still leaves plenty of room for treatment improvements. Patients who do not benefit from the current treatments should not be left overlooked and unacknowledged.

6.4.3 Theory and research implications

Several measurement issues were encountered during the project. There are no validated measures of treatment adherence specific to BA. It is difficult to establish whether BA has been delivered as intended without an assessment tool. A measure of adherence was developed for the purpose of this study, but generalisability and comparability with other studies is limited. BA research would benefit from the development of a standardised adherence measure that can be used as a self-rated or expert-rated tool. Measurement of patient values and values-based behaviour was also hampered by inadequate measurement tools. The Valued Living Questionnaire (VLQ) is more suited to use as a clinical tool, rather than for precise measurement. A more appropriate measure of *valued living* is needed to study its mediating properties.

Although this thesis was not able to implicate *valued living* as a mediator of change, understanding how BA produces reductions in depression can still provide insight into ways to tackle stasis. As stated, there were methodological limitations in relation to the study design and measurement of patient values-based behaviour that need to be clarified before the influence of values can be disregarded. If *valued living* is not shown to be a mechanism of change in BA, theoretical consideration is needed to propose other areas for investigation. The differential treatment effects between LI and HI versions of BA could be used to identify potential mechanisms that are heightened by HI protocols (and facilitate larger treatment effects). For example, reduced

avoidant/increased approach behaviours (formulated by functional analysis in HI BA) or reduced effects of rumination inhibiting behaviour (targeted with behavioural-based rumination work in HI BA) are possible mechanisms to be investigated.

More generally, while this body of work has provided a small insight into the phenomena of stasis, stasis remains an ongoing problem. There are still a lot of unanswered questions that require investigating. The present findings have focused on current treatment stasis outcomes, but there needs to be a better understanding of the prior experiences of these patients (e.g., are they 'frequent fliers' in mental health services, and if so, what treatments have they had before?) and the subsequent long-term experiences (e.g., do they return after a failed treatment, and if they do, are they offered a different option?). Nonresponse to treatment is also applicable to a variety of contexts outside the scope of this thesis. The stasis investigation here has been based around BA for depression, but stasis occurs after all types of treatment. Stasis is also not specific to depression - it is evident in all mental health disorders. Treatment outcome research needs to address the disparity in the evidence base to reflect the scale of stasis. Capturing and reporting stasis as a mainstream outcome, alongside treatment response rates, would help provide parity.

6.5 General strengths and limitations of the thesis

6.5.1 Strengths

The investigation conducted within this thesis had several strengths with regards to the approaches that were used. A range of methods were used across the studies to investigate the research questions - meta-analysis of existing treatment effects, retrospective analysis of longitudinal outcomes, a prospective intervention study, and mediation analysis. All the studies utilised practice-based evidence. As stasis outcomes are more of an issue in routine practice, the findings are more applicable to real world

delivery of depression treatment. The resulting implications relate to how therapy actually works when it is delivered to people in the real world. Likewise, the investigation benefitted from the advantages of using large routine outcome datasets, in that they provided a wealth of archived longitudinal session-by-session data on a range of outcomes. Routine outcome monitoring, in particular, ensured data was available for the last session attended so analysis did not have to depend on having a pre and post-treatment score. Consequently, recommendations as a result of the research are able to be implementable by services within the current set-up of systems. Although, practice-based research reduces the internal validity and increases the risk of bias, methods were employed to reduce the impact. For example, the Becker method (1988) was used to allow comparison of uncontrolled practice-based evidence with controlled trial evidence in the meta-analysis (Chapter two), and propensity score matching was employed to match patients to historical controls in Chapter five. These techniques aimed to emulate control trial conditions while retaining the generalisability to clinical practice.

6.5.2 Limitations

There were also a number of limitations across the studies. Those limitations relate to the robustness of the findings, restrictions due to the type of data and measures used, and wider generalisability of the results.

First, in relation to the robustness of the findings, the practice-based nature meant using less rigorous study designs that are more susceptible to threats to internal validity. The studies were not randomised, largely uncontrolled, and were susceptible to power issues (data collection was not able to be designed to suit the research). Treatment effects could be due to confounding factors such as unknown influences or changes in treatment, spontaneous recovery, or biases in patient treatment allocations in routine practice. Some of the studies were also hampered by small (Chapter five and seven) or uneven sample sizes (Chapter four). Therefore, we need to be wary of placing

too much weight on the conclusions. The use of archived data and routine practice delivery meant there were limited checks of treatment fidelity and competence. Treatment adherence was assessed for the BA groups used for the Chapter five and six analyses, but it is not known whether BA interventions were delivered adherently and competently in the stepped-care analysis in Chapter four. Future stasis investigation centred around BA would benefit from using the Behavioural Activation Treatment Scale (BATS; Jacobson et al., 1996). Likewise, homework completion and use of implementation intentions was not checked in Chapter five, so it is not known whether the augmentations were used as intended.

Second, regarding outcome measure and data restrictions, there were limitations due to issues inherent to practice-based data. Missing data were commonplace in the routine datasets, and reasons for 'missingness' were not always apparent. Treatment attendance was generally poor, although amount of sessions initially offered, and reasons for treatment termination were not available. Therefore, classifying patients as drop-outs was difficult. Session-by-session scores from the last available session were used to provide post-treatment scores using the last observation carried forward (LOCF) method. The LOCF method does have acknowledged faults and reduces accuracy of the results (Lachin, 2016). Future analysis would benefit from using more advanced multiple imputation methods. The nature of the data meant assumptions of independence were at risk of violation from the clustering effects in outcomes (i.e., patients clustered in groups or within therapists). Although checks were performed to assess the suitably of single-level analyses for the clustered data (using the magnitude of the design effects), it has been argued that intraclass correlations (ICCs) as low as 0.01 can still impact analyses (Baldwin et al., 2011). To increase confidence that data dependency is not influencing analyses, multi-level models would be best employed (when sample sizes/number of clusters are large enough).

Although the studies benefitted from the use of large routine datasets, they also came with restrictions. The outcome measures used and type of data that is collected, is for policy and monitoring purposes. It is not designed by theory or set up to test a hypothesis, therefore studies were limited to the information that is collected for practice. In terms of stasis investigation, only routinely collected information could be investigated as stasis predictors. Future stasis investigation would benefit from looking at potential sources of predictors outside of routinely collected information. Similarly, the outcome measures had to be the IAPT minimum dataset. Consequently, the results are based solely on self-report data that are known to be at risk of validity issues, such as social desirability when providing sensitive information (Tourangeau & Yan, 2007). The self-report measures also make diagnostic certainty difficult, as there was no specific diagnostic assessment. It is possible patients allocated to BA for depression would not be diagnosed as having clinical depression if diagnostically assessed. Session-by-session scores were attained using the same outcome measure meaning the pre- and post-treatment assessment is subject to testing effects (Wampold, 2015). A combination of self-report and therapist assessment measures would strengthen the interpretation of future studies. Meanwhile, only summary scores were available for analysis (with the exception of Chapter six that involved data collected for the purpose of the study). As a result, item data or information about different symptom clusters was not accessible to enhance the analyses.

Third, in terms of the wider generalisability of the results, the findings are limited to the IAPT minimum dataset and IAPT service delivery in England. With the exception of the Valued Living Questionnaire in Chapter seven, the only measures used were the Patient Health Questionnaire-9 (PHQ-9), Generalised Anxiety Disorder-7, and the Work and Social Adjustment Scale (WSAS). These scales are only brief and designed for use in Primary Care. Therefore, the present findings can only tell us about

depression that presents to IAPT. It cannot generalise to other types or presentations of depression. Also, this thesis did not assess the medium or long-term effects of BA interventions (group BA, LI BA or HI BA), as no follow-up was conducted. Long term assessment is needed to establish the durability of BA treatment effects, to determine patients' continued use of the implementations intentions technique, and to explore what happens to stasis patients after treatment has ended. Finally, the means of defining stasis is very challengeable. Although the investigation in this thesis adopted the reliable and clinically significant change criteria to classify when no change had occurred, some may argue whether a change of six or more on the PHQ-9 is clinically relevant to a patient (especially for severe depression). Patients may meet this criterion for improvement, yet feel like they have not benefitted from treatment. Conversely, some patients may only experience a small reduction in their outcomes, but feel they benefitted greatly from treatment. More patient-focused research, potentially utilising qualitative methods yoked to quantitative outcomes, would shed more light on these issues.

6.6 Recommendations for future research directions

The present thesis has added to the group BA evidence base and provided an initial investigation into stasis. In terms of further progressing the BA evidence base, future research should look to conduct longitudinal follow-up studies of BA intervention effects to establish the long-term treatment effects on depression. In particular, it would be useful to compare the durability of group versus individual BA and LI versus HI versions of BA.

These findings have shown some people are able to elicit a greater benefit from the delivery of a more intensive version of BA. The ability to match patients to the intensity of treatment they are most likely to benefit from, whilst still making the most

economical and effective use of resources, would be a valuable tool. Innovative machine learning algorithms could be used to create predictive models of BA treatment benefit. These models could then be used to establish who will get a significant benefit from HI BA over LI BA or individual therapy over groups, in order to optimise treatment allocation and maximise resources.

The current results were unable to provide any firm conclusions about the active mechanisms of change in BA. Controlled, longitudinal mediation studies should look to establish the factors that mediate depression changes after BA. Potential mechanisms could include reduced avoidant/increased approach behaviours, or reduced effects of rumination inhibiting behaviour.

In terms of continuing the stasis research, a controlled trial comparing BA with and without implementation intentions incorporated into the homework setting is needed to see whether the current findings can be replicated. Wider research should emulate the method used in this thesis and work to enhance the effectiveness of existing treatments using theory-informed treatment augmentations.

Research should also continue to look for strategies to increase treatment retention, as session attendance is critical for reducing risk of stasis. Techniques could make use of technology to facilitate attendance. The ability to book, rearrange and manage appointments online could make attendance easier for patients and alleviate some of the worries they have about cancelling or returning after missing a session. Alternatively, phone applications could give patients a sense of mastery over their treatment and encourage engagement, enabling patients to track their own progress in comparison to projected outcomes, receive feedback or personalised psychoeducation based on their progress, input homework progress, and receive appointment reminders.

Finally, additional investigation is also needed to provide a better understanding of what we still do not know about stasis. Studies should aim to look at the prior

experiences stasis patients have had with treatment, and whether there are differences between people who have continued stasis outcomes and those experiencing it for the first time. Likewise, investigation into what happens after experiencing a stasis outcome would be useful to reduce the risk of patients who do not benefit from treatment being 'lost' from the system.

6.7 Conclusion

This thesis has defined and explored stasis outcomes in routine practice after a BA group treatment for depression, in order to understand whether stasis can be predicted, and ultimately reduced. The resulting findings have expanded understanding of BA treatment for depression and the issue of stasis after evidence-based treatment.

First, the findings have demonstrated the utility of the BA model. Behavioural techniques alone are sufficient to produce changes in depression, meaning that treatments can be delivered more parsimoniously and over shorter durations (do not need additional sessions for other strategies). BA interventions can be brief (4-9 sessions), and can be effective using simple and complex protocols (with half as many people needing to be treated to see additional benefits after the more complex versions). BA produces similar outcomes regardless of delivery in groups or one-to-one. Potential added benefits of group BA are shorter waiting times and wider (plurality of intervention) and quicker (condensed protocol) disseminations. These thesis findings support BA delivered in groups as an effective treatment for depression, and advocate its use as a valuable treatment option for clinical services.

Second, this thesis has provided an initial investigation into the issue of stasis after depression treatments. It is clear that tackling stasis is a challenging prospect and one that clinical services can collude with ignoring, if consumed with the 50 % recovery rate. Depression stasis in routine services appears to be widely prevalent, with in excess

of 50% of patients not benefitting from the treatment they receive. Progression in stasis research has been hampered by a lack of a consensus in reporting and capturing inadequate treatment outcomes. The findings from this thesis have provided much needed investigation into the substantial subset of patients that do not benefit from depression treatment, but have historically been overlooked.

Two suggested ways of targeting stasis are: to identify what predicts who is likely to fail to respond to a specific treatment to inform more effective treatment allocation; and to try to improve the actual treatments. Encouragingly, very few patient characteristics affected the likelihood of not benefitting from BA treatment, with only greater impaired function and lower depression severity predicting stasis outcomes. What was most influential for reducing risk of stasis was attending enough sessions of treatment. Fewer sessions was the strongest stasis predictor and distinguishable feature for stasis versus improver patients, providing an area for possible intervention. However, attempts to increase group BA treatment retention with attendance-outcome psychoeducation did not have an effect. More promisingly, the use of a theoretically-informed strategy integrated into the group BA homework setting (implementation intentions) produced fewer stasis outcomes and more patients benefitting. Finally, an attempt to pinpoint additional key mechanisms for the reduction of depression during group BA was not able to implicate increased values-based behaviour as a mediator of change.

Taking these all the thesis findings together has several implications for how to tackle stasis going forward. Turnover is critical in meeting the demand for depression treatment and reducing time patients have to wait, thereby increasing their chances of a positive outcome. BA interventions, especially delivered in groups, provide a way to administer treatment widely and quickly, with patients needing to attend fewer than 10 sessions to get an adequate dose. Notably, the present research also highlights that the

enhancement of existing treatments can improve outcomes without the need to develop and evidence new treatment methods. It is therefore recommended that the improvement of extant treatment protocols is extensively adopted as a method for reducing stasis. Investigation into mechanisms of change would enable the identification of appropriate targets in existing treatments for the integration of theory-informed improvement strategies. Furthermore, treatment retention is vital, and strategies for increasing attendance remains an untapped area that could be generalised to multiple interventions.

In conclusion, depression stasis is an important issue that has potentially quietly devastating consequences, if left unchecked and unnoticed. Although there is clearly a long way to go, the encouraging findings from this thesis show that small changes can produce meaningful benefits for patients suffering with depression.

REFERENCES

References marked with an asterisk indicate group BA treatment studies included in the meta-analysis in Chapter 2.

Addis, M. E., & Jacobson, N. S. (2000). A closer look at the treatment rationale and homework compliance in cognitive-behavioral therapy for depression. *Cognitive Therapy and Research*, *24*, 313–326. https://doi.org/10.1023/A:1005563304265

Addis, M. E., & Martell, C. R. (2004). *Overcoming depression one step at a time: The new behavioral activation approach to getting your life back*. Oakland, CA: New Harbinger.

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

Ahern, E., Kinsella, S., & Semkovska, M. (2018). Clinical efficacy and economic evaluation of online cognitive behavioral therapy for major depressive disorder: a systematic review and meta-analysis. *Expert Review of Pharmacoeconomics & Outcomes Research*, *18*, 25–41. https://doi.org/10.1080/14737167.2018.1407245

Al-Harbi, K. S. (2012). Treatment-resistant depression: Therapeutic trends, challenges, and future directions. *Patient Preference and Adherence*, *6*, 369–388. https://doi.org/10.2147/PPA.S29716

Almond, S., & Healey, A. (2003). Mental health and absence from work: New evidence from the UK quarterly labour force survey. *Work, Employment and Society*, *17*, 731–742. https://doi.org/10.1177/0950017003174007

Altman, D. G., Gore, S. M., Gardner, M. J., & Pocock, S. J. (1983). Statistical guidelines for contributors to medical journals. *British Medical Journal*, *286*, 1489–1493. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1547706/pdf/bmjcred00552-0033.pdf

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Washington DC: Author.

Arbour, K. P., & Martin Ginis, K. A. (2009). A randomised controlled trial of the effects of implementation intentions on women's walking behaviour. *Psychology and Health*, *24*, 49–65. https://doi.org/10.1080/08870440801930312

Ardito, R. B., & Rabellino, D. (2011). Therapeutic alliance and outcome of psychotherapy: Historical excursus, measurements, and prospects for research. *Frontiers in Psychology*, *2*, 1–11. https://doi.org/10.3389/fpsyg.2011.00270

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

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, *28*, 3083–3107. https://doi.org/10.1002/sim.3697

Austin, P. C. (2011a). 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

Austin, P. C. (2011b). Optimal caliper widths for propensity-score matching when

estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, *10*, 150–161. https://doi.org/10.1002/pst.433

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

Baguley, C., Farrand, P., Hope, R., Leibowitz, J., Lovell, K., Lucock, M., … Williams, C. (2010). *Good practice guidance on the use of self-help materials within IAPT services*. *Technical Report, IAPT*. Retrieved from www.iapt.nhs.uk

Bahraini, N. H., Devore, M. D., Monteith, L. L., Forster, J. E., Bensen, S., & Brenner, L. A. (2013). The role of value importance and success in understanding suicidal ideation among Veterans. *Journal of Contextual Behavioral Science*, *2*, 31–38. https://doi.org/10.1016/j.jcbs.2013.03.001

Baldwin, S. A., Murray, D. M., Shadish, W. R., Pals, S. L., Holland, J. M., Abramowitz, J. S., … 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., Hardy, G. E., & Mellor-Clark, J. (2010). *Developing and delivering practice-based evidence: A guide for the psychological therapies*. Chichester, UK: Wiley-Blackwell. https://doi.org/10.1002/9780470687994

Barkham, M., & Mellor-Clark, J. (2003). Bridging evidence-based practice and

practice-based evidence: Developing a rigorous and relevant knowledge for the psychological therapies. *Clinical Psychology and Psychotherapy, 10*, 319–327. https://doi.org/10.1002/cpp.379

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

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

Barlow, D. H. (2010). Negative effects from psychological treatments: A perspective. *American Psychologist, 65*, 13–20. https://doi.org/10.1037/a0015643

Barnes, M., Sherlock, S., Thomas, L., Kessler, D., Kuyken, W., Owen-Smith, A., … Turner, K. (2013). No pain, no gain: Depressed clients' experiences of cognitive behavioural therapy. *British Journal of Clinical Psychology*, *52*, 347–364. https://doi.org/10.1111/bjc.12021

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

*Barrera, M. (1979). An evaluation of a brief group therapy for depression. *Journal of Consulting and Clinical Psychology, 47*, 413–415. https://doi.org/10.1037/0022-006X.47.2.413

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

Batalden, P. B., & Davidoff, F. (2007, February 1). What is "quality improvement" and how can it transform healthcare? *Quality and Safety in Health Care*, *16*, 2–3. https://doi.org/10.1136/qshc.2006.022046

Bayer, U. C., Achtziger, A., Gollwitzer, P. M., & Moskowitz, G. B. (2009). Responding to subliminal cues: Do if-then plans facilitate action preparation and initiation without conscious intent? *Social Cognition*, *27*, 183–201. https://doi.org/10.1521/soco.2009.27.2.183

Beck, A. T. (1964). Thinking and depression: II. Theory and therapy. *Archives of General Psychiatry*, *10*, 561–571. https://doi.org/10.1001/archpsyc.1964.01720240015003

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

Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond* (2nd ed.). New York, NY: Guilford Press.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*, 257–278. https://doi.org/10.1111/j.2044-8317.1988.tb00901.x

Beidas, R. S., & Kendall, P. C. (2010). Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science and Practice*, *17*, 1–30. https://doi.org/10.1111/j.1468-

2850.2009.01187.x

Bell, E. C., Marcus, D. K., & Goodlad, J. K. (2013). Are the parts as good as the whole? A meta-analysis of component treatment studies. *Journal of Consulting and Clinical Psychology*, *81*, 722–736. https://doi.org/10.1037/a0033004

Bergin, A. E. (1963). The empirical emphasis in psychotherapy: A symposium. The effects of psychotherapy: Negative results revisited. *Journal of Counseling Psychology*, *10*, 244–250. https://doi.org/10.1037/h0043353

*Besyner, J. K. (1978). The comparative efficacy of cognitive and behavioral treatments of depression: A multi-assessment approach. *Dissertation Abstracts International*, *39*, 4568B.

Bohart, A. C., & Wade, W. A. (2013). The client in psychotherapy. In *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 219–257). Hoboken, NJ: John Wiley & Sons Ltd.

Boland, R. J., & Keller, M. B. (2002). Course and outcome of depression. In I. H. Gotlib & C. L. Hammen (Eds.), *Handbook of depression* (pp. 43–60). New York, NY: Guilford Press. Retrieved from https://psycnet.apa.org/record/2002-01778-002

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: Access, effectiveness and efficiency. *British Journal of Psychiatry*, *186*, 11–17. https://doi.org/10.1192/bjp.186.1.11

Bramwell, K., & Richardson, T. (2018). Improvements in depression and mental health after acceptance and commitment therapy are related to changes in defusion and values-based action. *Journal of Contemporary Psychotherapy*, *48*, 9–14. https://doi.org/10.1007/s10879-017-9367-6

*Brand, E., & Clingempeel, W. G. (1992). Group behavioral therapy with depressed geriatric inpatients: An assessment of incremental efficacy. *Behavior Therapy*, *23*, 475–482. https://doi.org/10.1016/S0005-7894(05)80170-4

Brown, I., Sheeran, P., & Reuber, M. (2009). Enhancing antiepileptic drug adherence: A randomized controlled trial. *Epilepsy and Behavior*, *16*, 634–639. https://doi.org/10.1016/j.yebeh.2009.09.014

Brown, J. S. L., Sellwood, K., Beecham, J. K., Slade, M., Andiappan, M., Landau, S., … Smith, R. (2011). Outcome, costs and patient engagement for group and individual CBT for depression: A naturalistic clinical study. *Behavioural and Cognitive Psychotherapy*, *39*, 355–358. https://doi.org/10.1017/S135246581000072X

Bruijniks, S. J. E., Franx, G., & Huibers, M. J. H. (2018). The implementation and adherence to evidence-based protocols for psychotherapy for depression: The perspective of therapists in Dutch specialized mental healthcare. *BMC Psychiatry*, *18*, 1–10. https://doi.org/10.1186/s12888-018-1768-x

Burns, D. D., & Spangler, D. L. (2000). Does psychotherapy homework lead to improvements in depression in cognitive-behavioral therapy or does improvement lead to increased homework compliance? *Journal of Consulting and Clinical Psychology*, *68*, 46–56. https://doi.org/10.1037/0022-006X.68.1.46

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review*, *26*, 17–31. https://doi.org/10.1016/j.cpr.2005.07.003

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

https://doi.org/10.1348/014466503321903553

Cahill, J., Barkham, M., & Stiles, W. B. (2010). Systematic review of practice-based research on psychological therapies in routine clinic settings. *British Journal of Clinical Psychology*, *49*, 421–453. https://doi.org/10.1348/014466509X470789

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*, 111–119. https://doi.org/10.1037/a0012901

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

Carvalho, J., Trent, L. R., & Hopko, D. R. (2011). The impact of decreased environmental reward in predicting depression severity: Support for behavioral theories of depression. *Psychopathology*, *44*, 242–252. https://doi.org/10.1159/000322799

*Catanese, R. A., Rosenthal, T. L., & Kelley, J. E. (1979). Strange bedfellows: Reward, punishment and interpersonal distraction strategies in treating dysphoria. *Cognitive Therapy and Research*, *3*, 299–305.

Chan, A. T. Y., Sun, G. Y. Y., Tam, W. W. S., Tsoi, K. K. F., & Wong, S. Y. S. (2017). The effectiveness of group-based behavioral activation in the treatment of depression: An updated meta-analysis of randomized controlled trial. *Journal of Affective Disorders*, *208*, 345–354. https://doi.org/10.1016/j.jad.2016.08.026 (Retraction published December 1, 2018, Journal of Affective Disorders, 241, p. 643. doi.org/10.1016/j.jad.2018.07.084).

Chandola, T., & Zhang, N. (2018). Re-employment, job quality, health and allostatic load biomarkers: Prospective evidence from the UK Household Longitudinal Study. *International Journal of Epidemiology*, *47*, 47–57. https://doi.org/10.1093/ije/dyx150

Chen, J., Liu, X., Rapee, R. M., & Pillay, P. (2013). Behavioural activation: A pilot trial of transdiagnostic treatment for excessive worry. *Behaviour Research and Therapy*, *51*, 533–539. https://doi.org/10.1016/j.brat.2013.05.010

Chesney, E., Goodwin, G. M., & Fazel, S. (2014). Risks of all-cause and suicide mortality in mental disorders: A meta-review. *World Psychiatry*, *13*, 153–160. https://doi.org/10.1002/wps.20128

Ciarrochi, J., & Bailey, A. (2008). *A CBT practitioners guide to ACT: How to bridge the gap between cognitive behavioral therapy and acceptance and commitment therapy*. Oakland, CA: New Harbinger Publications.

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. https://doi.org/10.3109/09540261.2011.606803

Clark, D. M. (2018). Realizing the mass public benefit of evidence-based psychological therapies: The IAPT program. *Annual Review of Clinical Psychology*, *14*, 159–183. https://doi.org/10.1146/annurev-clinpsy-050817-084833

Clark, D. M., Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2017). Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *The Lancet*, *391*, 679–686. https://doi.org/10.1016/S0140-6736(17)32133-5

Cochrane Collaboration. (2011). Cochrane Handbook for Systematic Reviews of

Interventions. In J. P. T. Higgins & S. Green (Eds.) (version 5). Retrieved from www.handbook.cochrane.org.

Coffman, S. J., Martell, C. R., Dimidjian, S., Gallop, R., & Hollon, S. D. (2007). Extreme nonresponse in cognitive therapy: Can behavioral activation succeed where cognitive therapy fails? *Journal of Consulting and Clinical Psychology, 75*, 531–541. https://doi.org/10.1037/0022-006X.75.4.531

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. (1988). Statistical power analysis for the behavioral sciences. *Statistical Power Analysis for the Behavioral Sciences*. https://doi.org/10.1234/12345678

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

Cohen, Z. D., & Derubeis, R. J. (2018). Annual review of clinical psychology treatment selection in depression. *Annual Review of Clinical Psychology*, *14*, 209–245. https://doi.org/10.1146/annurev-clinpsy-050817

*Comas-Diaz, L. (1981). Effects of cognitive and behavioral group treatment on the depressive symptomatology of Puerto Rican women. *Journal of Consulting and Clinical Psychology*, *49*, 627–632. https://doi.org/10.1037//0022-006X.49.5.627

Constantino, M. J., Ametrano, R. M., & Greenberg, R. P. (2012). Clinician interventions and participant characteristics that foster adaptive patient expectations for psychotherapy and psychotherapeutic change. *Psychotherapy*, *49*, 557–569. https://doi.org/10.1037/a0029440

Cooper, A. A., & Conklin, L. R. (2015). Dropout from individual psychotherapy for

major depression: A meta-analysis of randomized clinical trials. *Clinical Psychology Review*, *40*, 57–65. https://doi.org/10.1016/j.cpr.2015.05.001

Corey-Lisle, P. K., Nash, R., Stang, P., & Swindle, R. (2004). Response, partial response, and nonresponse in primary care treatment of depression. *Archives of Internal Medicine*, *164*, 1197–1204. https://doi.org/10.1001/archinte.164.11.1197

Coyne, J. C. (2016). Toward an interactional description of depression. *Psychiatry*, *39*, 28–40. https://doi.org/10.1080/00332747.1976.11023874

Craske, M. G., Hermans, D., & Vervliet, B. (2018). State-of-the-art and future directions for extinction as a translational model for fear and anxiety. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*, 20170025. https://doi.org/10.1098/rstb.2017.0025

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

Cuijpers, P. (2018). The challenges of improving treatments for depression. *JAMA - Journal of the American Medical Association*, *320*, 2529–2530. https://doi.org/10.1001/jama.2018.17824

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *The Canadian Journal of Psychiatry*, *58*, 376–385. https://doi.org/10.1177/070674371305800702

Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Hollon, S. D. (2017). Component studies of psychological treatments of adult depression: A systematic review and meta-analysis. *Psychotherapy Research*, *0*, 1–15.

https://doi.org/10.1080/10503307.2017.1395922

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

Cuijpers, P., Smit, F., Oostenbrink, J., De Graaf, R., Ten Have, M., & Beekman, A. (2007). Economic costs of minor depression: A population-based study. *Acta Psychiatrica Scandinavica*, *115*, 229–236. https://doi.org/10.1111/j.1600-0447.2006.00851.x

Cuijpers, P., & Straten, A. van. (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

Cuijpers, P., Van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychological Medicine*, *40*, 211–223. https://doi.org/10.1017/S0033291709006114

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

Cuijpers, P., Ven Straten, A., Warmerdam, L., & Andersson, G. (2009). Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: A meta-analysis. *Depression and Anxiety*, *26*, 279–288. https://doi.org/10.1002/da.20519

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

Dahl, J. C., Plumb, J. C., Stewart, I., & Lundgren, T. (2009). *The art and science of valuing in psychotherapy: Helping clients discover, explore, and commit to valued action using acceptance and commitment therapy*. Oakland, CA: New Harbinger Publications.

*Daughters, S. B., Braun, A. R., Sargeant, M. N., Reynolds, E. K., Hopko, D. R., Blanco, C., & 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. https://doi.org/10.4088/JCP.v69n0116

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

Delgadillo, J., de Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., … McMillan, D. (2018). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: A multisite, open-label, cluster randomised controlled trial. *The Lancet Psychiatry*, *5*, 564–572. https://doi.org/10.1016/S2215-0366(18)30162-7

Delgadillo, J., & Groom, M. (2017). Using psychoeducation and role induction to improve completion rates in cognitive behavioural therapy. *Behavioural and Cognitive Psychotherapy*, *45*, 170–184. https://doi.org/10.1017/S1352465816000643

Delgadillo, J., Kellett, S., Ali, S., McMillan, D., Barkham, M., Saxon, D., … Lucock,

M. (2016). A multi-service practice research network study of large group psychoeducational cognitive behavioural therapy. *Behaviour Research and Therapy*, *87*, 155–161. https://doi.org/10.1016/j.brat.2016.09.010

Delgadillo, J., McMillan, D., Lucock, M., Leach, C., Ali, S., & Gilbody, S. (2014). Early changes, attrition, and dose-response in low intensity psychological interventions. *British Journal of Clinical Psychology*, *53*, 114–130. https://doi.org/10.1111/bjc.12031

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 bandomized controlled trial. *Journal of Clinical Psychology*, *71*, 1139–1152. https://doi.org/10.1002/jclp.22228

DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young, P. R., Salomon, R. M., … Gallop, R. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry*, *62*, 409–416. https://doi.org/10.1001/archpsyc.62.4.409

Dimidjian, S., Arch, J. J., Schneider, R. L., Desormeau, P., Felder, J. N., & Segal, Z. V. (2016). Considering meta-analysis, meaning, and metaphor: A systematic review and critical examination of "third wave" cognitive and behavioral therapies. *Behavior Therapy*, *47*, 886–905. https://doi.org/10.1016/j.beth.2016.07.002

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

clinpsy-032210-104535

Dimidjian, S., & Hollon, S. D. (2011). What can be learned when empirically supported treatments fail? *Cognitive and Behavioral Practice*, *18*, 303–305. https://doi.org/10.1016/J.CBPRA.2011.02.001

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., … Jacobson, N. S. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology*, *74*, 658–670. https://doi.org/10.1037/0022-006X.74.4.658

Doi, S., Yokomitsu, K., & Sakano, Y. (2016). Relationships among valued action, behavioral activation, avoidance, and reinforcement in a sample of college students. *Psychological Reports*, *118*, 5–22. https://doi.org/10.1177/0033294115625272

Dold, M., Bartova, L., Fugger, G., Kautzky, A., Souery, D., Mendlewicz, J., … Kasper, S. (2018). Major depression and the degree of suicidality: Results of the European Group for the Study of Resistant Depression (GSRD). *International Journal of Neuropsychopharmacology*, *21*, 539–549. https://doi.org/10.1093/ijnp/pyy009

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M. R., & Christensen, H. (2013). Smartphones for smarter delivery of mental health programs: A systematic review. *Journal of Medical Internet Research*, *15*, e247. https://doi.org/10.2196/jmir.2791

Downs, S., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomised studies of health care interventions. *Journal of Epidemiololgy and Community Health*, *52*, 377–384. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1756728/pdf/v052p00377.pdf

Driessen, E., Cuijpers, P., Hollon, S. D., & Dekker, J. J. M. (2010). Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. *Journal of Consulting and Clinical Psychology*, *78*, 668–680. https://doi.org/10.1037/a0020570

Driessen, E., & Hollon, S. D. (2010). Cognitive behavioral therapy for mood disorders: Efficacy, moderators and mediators. *Psychiatric Clinics of North America*, *33*, 537–555. https://doi.org/10.1016/j.psc.2010.04.005

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel- plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

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

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

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

Elhai, J. D., Contractor, A. A., Tamburrino, M., Fine, T. H., Prescott, M. R., Shirley, E., … Calabrese, J. R. (2012). The factor structure of major depression symptoms: A test of four competing models using the Patient Health Questionnaire-9. *Psychiatry Research*, *199*, 169–173. https://doi.org/10.1016/J.PSYCHRES.2012.05.018

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

Falconnier, L. (2009). Socioeconomic Status in the Treatment of Depression. *American Journal of Orthopsychiatry*, *79*, 148–158. https://doi.org/10.1037/a0015469

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

Fava, M., & Kendler, K. S. (2000). Major depressive disorder. *Neuron*, *28*, 335–341. https://doi.org/10.1016/S0896-6273(00)00112-4

*Fereidooni, S., Gharaei, B., Birashk, B., Sahraeian, A., & Hoseini, S. (2015). Efficacy of group behavioral activation treatment as supplemental treatment to pharmacotherapy in inpatients with depressive disorders. *Journal of Mood Disorders*, *5*, 104. https://doi.org/10.5455/jmood.20140913061428

Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J. L., … Whiteford, H. A. (2013). Burden of depressive disorders by country, sex, age, and year: Findings from the Global Burden of Disease Study 2010. *PLoS Medicine*, *10*. https://doi.org/10.1371/journal.pmed.1001547

Ferssizidis, P., Adams, L. M., Kashdan, T. B., Plummer, C., Mishra, A., & Ciarrochi, J. (2010). Motivation for and commitment to social values: The roles of age and gender. *Motivation and Emotion*, *34*, 354–362. https://doi.org/10.1007/s11031-010-9187-4

Ferster, C. B. (1973). A functional analysis of depression. *American Psychologist, 28,* 857–870. https://doi.org/10.1037/h0035605

Fife-Schaw, C., Sheeran, P., & Norman, P. (2007). Simulating behaviour change interventions based on the theory of planned behaviour: Impacts on intention and action. *British Journal of Social Psychology, 46,* 43–68. https://doi.org/10.1348/014466605X85906

Finning, K., Richards, D. A., Moore, L., Ekers, D., McMillan, D., Farrand, P. A., … Wray, F. (2017). Cost and outcome of behavioural activation versus cognitive behavioural therapy for depression (COBRA): A qualitative process evaluation. *BMJ Open*, *7*, e014161. https://doi.org/10.1136/bmjopen-2016-014161

Firth, N., Barkham, M., & Kellett, S. (2015). The clinical effectiveness of stepped care systems for depression in working age adults: A systematic review. *Journal of Affective Disorders*, *170*, 119–130. https://doi.org/10.1016/j.jad.2014.08.030

Firth, N., Barkham, M., Kellett, S., & 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. https://doi.org/10.1016/j.brat.2015.04.001

Fowler, J. C., Patriquin, M., Madan, A., Allen, J. G., Frueh, B. C., & Oldham, J. M. (2015). Early identification of treatment non-response utilizing the Patient Health Questionnaire (PHQ-9). *Journal of Psychiatric Research, 68,* 114–119. https://doi.org/10.1016/j.jpsychires.2015.06.018

Frank, E., Cassano, G. B., Rucci, P., Thompson, W. K., Kraemer, H. C., Fagiolini, A., … Forgione, R. N. (2011). Predictors and moderators of time to remission of major depression with interpersonal psychotherapy and SSRI pharmacotherapy. *Psychological Medicine*, *41*, 151–162.

https://doi.org/10.1017/S0033291710000553

Freud, S. (1917). Mourning and melancholia. *Standard Edition*, *14*, 237–259.

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

*Fuchs, C. Z., & Rehm, L. P. (1977). A self-control behavior therapy program for depression. *Journal of Consulting and Clinical Psychology*, *45*, 206–215. https://doi.org/10.1037/0022-006X.45.2.206

*Gallagher-Thompson, D., Lovett, S., Rose, J., McKibbin, C., Coon, D., Futterman, A., & Thompson, L. W. (2000). Impact of psychoeducational interventions on distressed family caregivers. *Journal of Clinical Geropsychology*, *6*, 91–110. https://doi.org/10.1023/A:1009584427018

*Gallagher, D. (1981). Behavioral group therapy with elderly depressives: An experimental study. In D. Upper & S. Ross (Eds.), *Behavioral group therapy* (pp. 187–224). Champaign, IL: Research Press.

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

Gelhart, R. P., & King, H. L. (2001). The influence of comorbid risk factors on the effectiveness of cognitive-behavioral treatment of depression. *Cognitive and Behavioral Practice*, *8*, 18–28. https://doi.org/10.1016/S1077-7229(01)80039-0

Gelhorn, H. L., Sexton, C. C., & Classi, P. M. (2011). Patient preferences for treatment of major depressive disorder and the impact on health outcomes: A systematic

review. *The Primary Care Companion for CNS Disorders*, *13*, PCC.11r01161. https://doi.org/10.4088/PCC.11r01161

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

Gibbons, C. R., Wiltsey Stirman, S., Derubeis, R. J., Newman, C. F., & Beck, A. T. (2013). Research setting versus clinic setting: Which produces better outcomes in cognitive therapy for depression? *Cognitive Therapy and Research*, *37*, 605–612. https://doi.org/10.1007/s10608-012-9499-7

Gilbody, S., Richards, D., Brealey, S., & 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. https://doi.org/10.1007/s11606-007-0333-y

Gloster, A. T., Klotsche, J., Ciarrochi, J., Eifert, G., Sonntag, R., Wittchen, H. U., & Hoyer, J. (2017). Increasing valued behaviors precedes reduction in suffering: Findings from a randomized controlled trial using ACT. *Behaviour Research and Therapy*, *91*, 64–71. https://doi.org/10.1016/j.brat.2017.01.013

Goldberg, S. B., Babins-Wagner, R., Rousmaniere, T., Berzins, S., Hoyt, W. T., Whipple, J. L., … Wampold, B. E. (2016). Creating a climate for therapist improvement: A case study of an agency focused on outcomes and deliberate practice. *Psychotherapy*, *53*, 367–375. https://doi.org/10.1037/pst0000060

Goldstein, D. J., Potter, W. Z., Ciraulo, D. A., & Shader, R. I. (2011). Biological theories of depression and implications for current and new treatments. *Pharmacotherapy of Depression: Second Edition*, 1–32.

https://doi.org/10.1007/978-1-60327-435-7_1

Gollan, J. K., Hoxha, D., Hunnicutt-Ferguson, K., Norris, C. J., Rosebrock, L., Sankin, L., & Cacioppo, J. (2016). The negativity bias predicts response rate to Behavioral Activation for depression. *Journal of Behavior Therapy and Experimental Psychiatry*, *52*, 171–178. https://doi.org/10.1016/j.jbtep.2015.09.011

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

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology*, *38*, 69–119. https://doi.org/10.1016/S0065-2601(06)38002-1

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

Hamilton, K. E., & Dobson, K. S. (2002). Cognitive therapy of depression: Pretreatment patient predictors of outcome. *Clinical Psychology Review*, *22*, 875–893. https://doi.org/10.1016/S0272-7358(02)00106-X

Hans, E., & Hiller, W. (2013). Effectiveness of and dropout from outpatient cognitive behavioral therapy for adult unipolar depression: A meta-analysis of nonrandomized effectiveness studies. *Journal of Consulting and Clinical Psychology*, *81*, 75–88. https://doi.org/10.1037/a0031080

Hansen, A. M. D., Hoogduin, C. A. L., Schaap, C., & de Haan, E. (1992). Do drop-outs differ from successfully treated obsessive-compulsives? *Behaviour Research and Therapy*, *30*, 547–550. https://doi.org/10.1016/0005-7967(92)90040-N

Hansen, N. B., & Lambert, M. J. (2003). An evaluation of the dose-response

relationship in naturalistic treatment settings using survival analysis. *Mental Health Services Research, 5*, 1–12. https://doi.org/10.1023/A:1021751307358

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

Hardy, G. E., Bishop-Edwards, L., Chambers, E., Connell, J., Dent-Brown, K., Kothari, G., … Parry, G. D. (2017). Risk factors for negative experiences during psychotherapy. *Psychotherapy Research*, *29*, 403–414. https://doi.org/10.1080/10503307.2017.1393575

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

Harvey, S. B., Modini, M., Joyce, S., Milligan-Saville, J. S., Tan, L., Mykletun, A., … Mitchell, P. B. (2017). Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. *Occupational and Environmental Medicine*, *74*, 301–310. https://doi.org/10.1136/oemed-2016-104015

Hawton, K., Casañas I Comabella, C., Haw, C., & Saunders, K. (2013). Risk factors for suicide in individuals with depression: A systematic review. *Journal of Affective Disorders*, *147*, 17–28. https://doi.org/10.1016/j.jad.2013.01.004

Hayes, S. A., Orsillo, S. M., & Roemer, L. (2010). Changes in proposed mechanisms of action during an acceptance-based behavior therapy for generalized anxiety disorder. *Behaviour Research and Therapy*, *48*, 238–245. https://doi.org/10.1016/j.brat.2009.11.006

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change*. New York, NY: Guilford Press.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2012). *Acceptance and commitment therapy: The process and practice of mindful change* (2nd ed.). New York, NY: Guilford Press.

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

Hepgul, N., King, S., Amarasinghe, M., Breen, G., Grant, N., Grey, N., … Cleare, A. J. (2016). Clinical characteristics of patients assessed within an Improving Access to Psychological Therapies (IAPT) service: Results from a naturalistic cohort study (Predicting Outcome Following Psychological Therapy; PROMPT). *BMC Psychiatry*, *16*. https://doi.org/10.1186/s12888-016-0736-6

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

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

Hiller, W., Schindler, A. C., & Lambert, M. J. (2012). Defining response and remission in psychotherapy research: A comparison of the RCI and the method of percent improvement. *Psychotherapy Research*, *22*, 1–11. https://doi.org/10.1080/10503307.2011.616237

Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The

efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research*, *36*, 427–440. https://doi.org/10.1007/s10608-012-9476-1

Hölzel, L., Härter, M., Reese, C., & Kriston, L. (2011). Risk factors for chronic depression - A systematic review. *Journal of Affective Disorders*, *129*, 1–13. https://doi.org/10.1016/j.jad.2010.03.025

Hopko, D. R., & Lejuez, C. W. (2007). *A cancer patient's guide to overcoming depression and anxiety: Getting through treatment and getting back to your life*. Oakland, CA: New Harbinger Publications. Retrieved from https://www.kobo.com/us/en/ebook/a-cancer-patient-s-guide-to-overcoming-depression-and-anxiety

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

Hopko, D. R., 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

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

*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

House of Commons Library. (2018). *Mental health statistics for England: Prevalence,*

services and funding. London: House of Commons Library. https://doi.org/10.1117/12.511423

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist, 41*, 159–164. https://doi.org/10.1037/0003-066X.41.2.159

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

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

Improving Access to Psychological Therapies (IAPT). (2011). *The IAPT data handbook*. Retrieved from http://www.iapt.nhs.uk/silo/files/iapt-data-handbook-appendicies-v2.pdf

Improving Access to Psychological Therapies (IAPT). (2014). *Measuring improvement and recovery (Version 2)*. Retrieved from 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., … Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology*, *64*, 295–304. https://doi.org/10.1037//0022-006X.64.2.295

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, *15*, 336–352. https://doi.org/10.1016/S0005-7894(84)80002-7

Jacobson, N. S., Martell, C. R., & Dimidjian, S. (2006). Behavioral activation treatment for depression: Returning to contextual roots. *Clinical Psychology: Science and Practice*, *8*, 255–270. https://doi.org/10.1093/clipsy.8.3.255

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

Jarden, A. J. (2010). *Relationship between personal values, and depressed mood and subjective well-being*. Unpublished disseration. University of Canterbury. Retrieved from https://ir.canterbury.ac.nz/handle/10092/5205

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

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

Kanter, J. W., & Puspitasari, A. J. (2016). Global dissemination and implementation of behavioural activation. *The Lancet*, *388*, 843–844. https://doi.org/10.1016/S0140-6736(16)31131-X

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

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology, 3*, 1–27. https://doi.org/10.1146/annurev.clinpsy.3.022806.091432

Kazdin, A. E. (2009). Understanding how and why psychotherapy leads to change. *Psychotherapy Research, 19*, 418–428. https://doi.org/10.1080/10503300802448899

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

Kellar, I., & Abraham, C. (2005). Randomized controlled trial of a brief research-based intervention promoting fruit and vegetable consumption. *British Journal of Health Psychology, 10*, 543–558. https://doi.org/10.1348/135910705X42940

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

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

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., … Wang, P. S. (2003). The epidemiology of major depressive disorder. *JAMA - Journal of the American Medical Association*, *289*, 3095–3105. https://doi.org/doi:10.1001/jama.289.23.3095

Kessler, R. C., Greenberg, P. E., Mickelson, K. D., Meneades, L. M., & Wang, P. S. (2001). The effects of chronic medical conditions on work loss and work cutback. *Journal of Occupational and Environmental Medicine*, *43*, 218–225. https://doi.org/10.1097/00043764-200103000-00009

Killip, S., Mahfoud, Z., & Pearce, K. (2004). What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. *Annals of Family Medicine*, *2*, 204–208. https://doi.org/10.1370/AFM.141

Knapp, M., & Ilson, S. (2002). Economic aspects of depression and its treatment. *Current Opinion in Psychiatry*, *15*, 69–75. https://doi.org/10.1097/00001504-200201000-00012

Knoth, R. L., Bolge, S. C., Kim, E., & Tran, Q.-V. (2010). Effect of inadequate response to treatment in patients with depression. *The American Journal of Managed Care*, *16*, e188-96. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20690785

*Kornblith, S. J., Rehm, L. P., O'Hara, M. W., & Lamparski, D. M. (1983). The contribution of self-reinforcement training and behavioral assignments to the efficacy of self-control therapy for depression. *Cognitive Therapy and Research*, *7*, 499–527. https://doi.org/10.1007/BF01172888

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

Krämer, L. V., Helmes, A. W., Seelig, H., Fuchs, R., & Bengel, J. (2014). Correlates of reduced exercise behaviour in depression: The role of motivational and volitional deficits. *Psychology and Health*, *29*, 1206–1225. https://doi.org/10.1080/08870446.2014.918978

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

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

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

Lam, R. W., Kennedy, S. H., McIntyre, R. S., & Khullar, A. (2014). Cognitive dysfunction in major depressive disorder: Effects on psychosocial functioning and implications for treatment. *Canadian Journal of Psychiatry*, *59*, 649–654. https://doi.org/10.1177/070674371405901206

Lambert, M. J. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, *17*, 1–14. https://doi.org/10.1080/10503300601032506

Lambert, M. J. (2011). What have we learned about treatment failure in empirically supported treatments? Some suggestions for practice. *Cognitive and Behavioral Practice*, *18*, 413–420. https://doi.org/10.1016/j.cbpra.2011.02.002

Lambert, M. J. (2013). Outcome in psychotherapy: The past and important advances.

Psychotherapy, *50*, 42–51. https://doi.org/10.1037/a0030682

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

Layard, R. (2006). The case for psychological treatment centres. *British Medical Journal*, *332*, 1030–1032. https://doi.org/10.1136/bmj.332.7548.1030

Layard, R., & Clark, D. M. (2014). *Thrive: The power of psychological therapy*. London, UK: Allen Lane, Penguin Group.

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

Lemeshow, S., & Hosmer, D. W. (1982). A review of goodness of fit statistics for use in the development of logistic regression models. *American Journal of Epidemiology*, *115*, 92–106.

Lewinsohn, P. M., & Graf, M. (1973). Pleasant activities and depression. *Journal of Consulting and Clinical Psychology*, *41*, 261–268. https://doi.org/10.1037/h0035142

Lewinsohn, P. M., & Libet, J. (1972). Pleasant events, activity schedules, and depressions. *Journal of Abnormal Psychology*, *79*, 291–295. https://doi.org/10.1037/h0033207

Lewinsohn, P. M., Sullivan, J. M., & Grosscup, S. J. (1980). Changing reinforcing events: An approach to the treatment of depression. *Psychotherapy: Theory,*

Research & Practice, *17*, 322–334. https://doi.org/10.1037/h0085929

Lim, G. Y., Tam, W. W., Lu, Y., Ho, C. S., Zhang, M. W., & Ho, R. C. (2018). Prevalence of depression in the community from 30 countries between 1994 and 2014. *Scientific Reports*, *8*, 2861. https://doi.org/10.1038/s41598-018-21243-x

Limosin, F., Loze, J.-Y., Zylberman-Bouhassira, M., Schmidt, M. E., Perrin, E., & Rouillon, F. (2004). The course of depressive illness in general practice. *Canadian Journal of Psychiatry*, *49*, 119–123. https://doi.org/10.1177/070674370404900207

Linden, M. (2013). How to define, find and classify side effects in psychotherapy: From unwanted events to adverse treatment reactions. *Clinical Psychology and Psychotherapy*, *20*, 286–296. https://doi.org/10.1002/cpp.1765

Longmore, R. J., & Worrell, M. (2007). Do we need to challenge thoughts in cognitive behavior therapy? *Clinical Psychology Review*, *27*, 173–187. https://doi.org/10.1016/j.cpr.2006.08.001

López, M., Hurtado, G. D., Santiago-Rivera, A. L., West, P., Santos, M. M., Kanter, J. W., & Nagy, G. (2014). A randomized hybrid efficacy and effectiveness trial of behavioral activation for Latinos with depression. *Behavior Therapy*, *46*, 177–192. https://doi.org/10.1016/j.beth.2014.09.011

Lorentzen, S., Hoglend, P., Martinsen, E. W., & Ringdal, E. (2011). Practice-based evidence: Patients who did not respond to group analysis. *International Journal of Group Psychotherapy*, *61*, 366–395.

Lorenzo-Luaces, L., Derubeis, R. J., & Webb, C. A. (2014). Client characteristics as moderators of the relation between the therapeutic alliance and outcome in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, *82*, 368–373. https://doi.org/10.1037/a0035994

*Lovett, S., & Gallagher, D. (1988). Psychoeducational interventions for family caregivers: Preliminary efficacy data. *Behavior Therapy, 19*, 321–330. https://doi.org/10.1016/S0005-7894(88)80005-4

Lucock, M., Bartys, S., Cupac, J., Delgadillo, J., Denton, C., Gaines, S., … 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

Lundgren, T., Dahl, J., & Hayes, S. C. (2008). Evaluation of mediators of change in the treatment of epilepsy with acceptance and commitment therapy. *Journal of Behavioral Medicine, 31*, 225–235. https://doi.org/10.1007/s10865-008-9151-x

Lunnen, K. M., & Ogles, B. M. (1998). A multiperspective, multivariable evaluation of reliable change. *Journal of Consulting and Clinical Psychology, 66*, 400–410. https://doi.org/10.1037/0022-006X.66.2.400

Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders*, *118*, 60–68. https://doi.org/10.1016/j.jad.2009.01.019

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

Lyubomksky, S., Sheldon, K. M., & Schkade, D. (2005). Pursuing happiness: The architecture of sustainable change. *Review of General Psychology, 9*, 111–131. https://doi.org/10.1037/1089-2680.9.2.111

Macaskill, P., Walter, S. D., & Irwig, L. (2001). A comparison of methods to detect

publication bias in meta-analysis. *Statistics in Medicine, 20*, 641–654. https://doi.org/10.1002/sim.698

Maciejewski, P. K., Prigerson, H. G., & Mazure, C. M. (2000). Self-efficacy as a mediator between stressful life events and depressive symptoms: Differences based on history of prior depression. *British Journal of Psychiatry, 176*, 373–378. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10827887

MacPherson, L., Tull, M. T., Matusiewicz, A. K., Rodman, S., Strong, D. R., Kahler, C. W., … Lejuez, C. W. (2010). Randomized controlled trial of behavioral activation smoking cessation treatment for smokers with elevated depressive symptoms. *Journal of Consulting and Clinical Psychology, 78*, 55–61. https://doi.org/10.1037/a0017939

*Magidson, J. F., Gorka, S. M., MacPherson, L., Hopko, D. R., Blanco, C., Lejuez, C. W., & Daughters, S. B. (2011). Examining the effect of the Life Enhancement Treatment for Substance Use (LETS ACT) on residential substance abuse treatment retention. *Addictive Behaviors, 36*, 615–623. https://doi.org/10.1016/j.addbeh.2011.01.016

*Magidson, J. F., Seitz-Brown, C. J., Safren, S. A., & Daughters, S. B. (2014). Implementing behavioral activation and life-steps for depression and HIV medication adherence in a community health center. *Cognitive and Behavioral Practice, 21*, 386–403. https://doi.org/10.1016/j.cbpra.2013.10.002

Marks, D. F. (2018). IAPT under the microscope. *Journal of Health Psychology, 23*, 1131–1135. https://doi.org/10.1177/1359105318781872

Marshall, D., Quinn, C., Child, S., Shenton, D., Pooler, J., Forber, S., & Byng, R. (2016). What IAPT services can learn from those who do not attend. *Journal of Mental Health, 25*, 410–415. https://doi.org/10.3109/09638237.2015.1101057

Martell, C. R., Addis, M. E., & Jacobson, N. S. (2001). *Depression in context: Strategies for guided action.* New York, NY: 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*. New York, NY: Guilford Press.

Mauskopf, J. A., Simon, G. E., Kalsekar, A., Nimsch, C., Dunayevich, E., & Cameron, A. (2009). Nonresponse, partial response, and failure to achieve remission: humanistic and cost burden in major depressive disorder. *Depression and Anxiety*, *26*, 83–97. https://doi.org/10.1002/da.20505

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

McCrone, P., Dhanasiri, S., Patel, A., Knapp, M., & Lawton-Smith, S. (2008). *Paying the price: The cost of mental health care in England to 2026*. London: King's Fund.

McDermut, W., Miller, I. W., & Brown, R. A. (2002). The efficacy of group psychotherapy for depression: A meta-analysis and review of the empirical research. *Clinical Psychology: Science and Practice*, *8*, 98–116. https://doi.org/10.1093/clipsy/8.1.98

McManus, S., Meltzer, H., Brugha, T., Bebbington, P., & Jenkins, R. (2009). *Adult psychiatric morbidity in England, 2007: Results of a household survey*. Leeds: The Health & Social Care Information Centre, Social Care Statistics. Retrieved from https://files.digital.nhs.uk/publicationimport/pub02xxx/pub02931/adul-psyc-morb-res-hou-sur-eng-2007-rep.pdf

Melartin, T. K., Rytsälä, H. J., Leskelä, U. S., Lestelä-Mielonen, P. S., Sokero, T. P., & Isometsä, E. T. (2004). Severity and comorbidity predict episode duration and recurrence of DSM-IV major depressive disorder. *Journal of Clinical Psychiatry*, *65*, 810–819. https://doi.org/10.4088/JCP.v65n0612

Meltzer, H., Bebbington, P., Brugha, T., Farrell, M., Jenkins, R., & Lewis, G. (2003). The reluctance to seek treatment for neurotic disorders. *International Review of Psychiatry*, *15*, 123–128. https://doi.org/10.1080/0954026021000046038

Michelson, S. E., Lee, J. K., Orsillo, S. M., & Roemer, L. (2011). The role of values-consistent behavior in generalized anxiety disorder. *Depression and Anxiety*, *28*, 358–366. https://doi.org/10.1002/da.20793

Mohr, D. C., Beutler, L. E., Engle, D., Shoham-Salomon, V., Bergan, J., Kaszniak, A. W., & Yost, E. B. (1990). Identification of patients at risk for nonresponse and negative outcome in psychotherapy. *Journal of Consulting and Clinical Psychology*, *58*, 622–628. https://doi.org/10.1037/0022-006X.58.5.622

Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods*, *22*, 6–27. https://doi.org/10.1037/met0000086

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

Moriana, J. A., Gálvez-Lara, M., & Corpas, J. (2017). Psychological treatments for mental disorders in adults: A review of the evidence of leading international organizations. *Clinical Psychology Review*, *54*, 29–43. https://doi.org/10.1016/j.cpr.2017.03.008

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*, 364–386. https://doi.org/10.1177/1094428106291059

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105–125. https://doi.org/10.1037/1082-989X.7.1.105

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

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry*, *180*, 461–464. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11983645

Murray, C. J. L., & Lopez, A. D. (1996). *The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020*. World Health Organisation. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/41864/0965546608_eng.pdf?sequence=1

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 [NICE]. (2009). *Depression: The treatment and management of depression in adults (update)*. (Clinical Guideline 90). London. Available at: www.nice.org.uk/CG90.

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

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

NHS England. (2016). *The five year forward view for mental health*. London: NHS England.

*Norton, M. J. (2010). *Evaluation of a brief behavioral activation therapy for depression (BATD) group protocol in an inpatient geriatric psychiatry facility*. (Unpublished doctoral dissertation). The University of Alabama, Tuscaloosa.

O'Connor, S. R., Tully, M. A., Ryan, B., Bradley, J. M., Baxter, G. D., & McDonough, S. M. (2015). Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: A comparison study. *BMC Research Notes*, *8*, 224–230. https://doi.org/10.1186/s13104-015-1181-1

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. https://doi.org/10.1037/a0029630

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

Page, A. C., & Hooke, G. R. (2009). Best practices: Increased attendance in inpatient group psychotherapy improves patient outcomes. *Psychiatric Services*, *60*, 426–428. https://doi.org/10.1176/appi.ps.60.4.426

Pagoto, S., Bodenlos, J. S., Schneider, K. L., Olendzki, B., Spates, C. R., & Ma, Y. (2008). Initial investigation of behavioral activation therapy for co-morbid major depressive disorder and obesity. *Psychotherapy*, *45*, 410–415. https://doi.org/10.1037/a0013313

Papakostas, G. I., Petersen, T., Mahal, Y., Mischoulon, D., Nierenberg, A. A., & Fava, M. (2004). Quality of life assessments in major depressive disorder: A review of the literature. *General Hospital Psychiatry*, *26*, 13–17. https://doi.org/10.1016/j.genhosppsych.2003.07.004

Parker, Z. J., & Waller, G. (2015). Factors related to psychotherapists' self-assessment when treating anxiety and other disorders. *Behaviour Research and Therapy*, *66*, 1–7. https://doi.org/10.1016/j.brat.2014.12.010

Parks–Stamm, E. J., Gollwitzer, P. M., & Oettingen, G. (2007). Action control by implementation intentions: Effective cue detection and efficient response Initiation. *Social Cognition*, *25*, 248–266. https://doi.org/10.1521/soco.2007.25.2.248

Parry, G. D., Crawford, M. J., & Duggan, C. (2016). Iatrogenic harm from psychological therapies – time to move on. *British Journal of Psychiatry*, *208*, 210–212. https://doi.org/10.1192/bjp.bp.115.163618

Pass, L., Hodgson, E., Whitney, H., & Reynolds, S. (2018). Brief behavioral activation treatment for depressed adolescents delivered by nonspecialist clinicians: A case illustration. *Cognitive and Behavioral Practice*, *25*, 208–224. https://doi.org/10.1016/j.cbpra.2017.05.003

Pass, L., Lejuez, C. W., & Reynolds, S. (2018). Brief behavioural activation (brief BA) for adolescent depression: A pilot study. *Behavioural and Cognitive Psychotherapy*, *46*, 182–194. https://doi.org/10.1017/S1352465817000443

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

Petersen, J. J., Paulitsch, M. A., Hartig, J., Mergenthal, K., Gerlach, F. M., & Gensichen, J. (2014). Factor structure and measurement invariance of the Patient Health Questionnaire-9 for female and male primary care patients with major depression in Germany. *Journal of Affective Disorders*, *170C*, 138–142. https://doi.org/10.1016/j.jad.2014.08.053

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

Plumb, J. C., & Hayes, S. C. (2008). *An examination of personal values in depressed and non-depressed individuals*. University of Nevada.

Plumb, J. C., Stewart, I., Dahl, J., & Lundgren, T. (2009). In search of meaning: Values in modern clinical behavior analysis. *The Behavior Analyst*, *32*, 85–103. https://doi.org/10.1007/BF03392177

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

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

Pybis, J., Saxon, D., Hill, A., & Barkham, M. (2017). The comparative effectiveness and efficiency of cognitive behaviour therapy and generic counselling in the treatment of depression: Evidence from the 2nd UK National Audit of psychological therapies. *BMC Psychiatry*, *17*. https://doi.org/10.1186/s12888-017-1370-7

Ramnerö, J., Folke, F., & Kanter, J. W. (2016). A learning theory account of depression. *Scandinavian Journal of Psychology*, *57*, 73–82. https://doi.org/10.1111/sjop.12233

Rees, C. S., McEvoy, P., & Nathan, P. R. (2005). Relationship between homework completion and outcome in cognitive behaviour therapy. *Cognitive Behaviour Therapy*, *34*, 242–247. https://doi.org/10.1080/16506070510011548

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

*Rehm, L. P., Fuchs, C. Z., Roth, D. M., Kornblith, S. J., & Romano, J. M. (1979). A comparison fo self-control and assertion skills treatments of depression. *Behavior Therapy*, *10*, 429–442.

*Rehm, L. P., Kaslow, N. J., & Rabin, A. S. (1987). Cognitive and behavioral targets in a self-control therapy program for depression. *Journal of Consulting and Clinical Psychology*, *55*, 60–67. https://doi.org/10.1037/0022-006X.55.1.60

*Rehm, L. P., Kornblith, S. J., O'Hara, M. W., Lamparski, D. M., Romano, J. M., &

Volkin, J. I. (1981). An evaluation of major components in a self-control therapy program for depression. *Behavior Modification*, *5*, 459–489. https://doi.org/10.1177/014544558154002

Reis, B. F., & Brown, L. G. (2006). Preventing therapy dropout in the real world: The clinical utility of videotape preparation and client estimate of treatment duration. *Professional Psychology: Research and Practice*, *37*, 311–316. https://doi.org/10.1037/0735-7028.37.3.311

Reuter, L., Munder, T., Altmann, U., Hartmann, A., Strauss, B., & Scheidt, C. E. (2016). Pretreatment and process predictors of nonresponse at different stages of inpatient psychotherapy. *Psychotherapy Research*, *26*, 410–424. https://doi.org/10.1080/10503307.2015.1030471

Reynolds, E. K., MacPherson, L., Tull, M. T., Baruch, D. E., & Lejuez, C. W. (2011). Integration of the brief behavioral activation treatment for depression (BATD) into a college orientation program: Depression and alcohol outcomes. *Journal of Counseling Psychology*, *58*, 555–564. https://doi.org/10.1037/a0024634

Reynolds, S. (2000). Evidence based practice and psychotherapy research. *Journal of Mental Health*, *9*, 257–266. https://doi.org/10.1080/jmh.9.3.257.266

Richards, D. A., & 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. https://doi.org/10.1016/j.jad.2011.03.024

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

6736(16)31140-0

Robinson, L., Delgadillo, J., & Kellett, S. (2019). The dose-response effect in routinely delivered psychological therapies: A systematic review. *Psychotherapy Research*, 1–28. https://doi.org/10.1080/10503307.2019.1566676

Rogers, C. R. (1964). Toward a modern approach to values: The valuing process in the mature person. *Journal of Abnormal and Social Psychology*, *68*, 160–167. https://doi.org/10.1037/h0046419

Roth, A. D., & Pilling, S. (2007). *The competences required to deliver effective cognitive and behavioural therapy for people with depression and with anxiety disorders*. London, UK: Department of Health.

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

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., … Maurizio Fava, M. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report. *The American Journal of Psychiatry*, *163*, 1905–1917.

Ryba, M. M., Lejuez, C. W., & Hopko, D. R. (2014). Behavioral activation for depressed breast cancer patients: The impact of therapeutic compliance and quantity of activities completed on symptom reduction. *Journal of Consulting and Clinical Psychology*, *82*, 325–335. https://doi.org/10.1037/a0035363

Salkovskis, P. (1995). Demonstrating specific effects in cognitive and behavioural therapy. In M. Aveline & D. Shapiro (Eds.), *Research foundations for psychotherapy practice* (pp. 191–228). Chichester, UK: Wiley.

Salmoiraghi, A., & Sambhi, R. (2010). Early termination of cognitive-behavioural interventions: Literature review. *The Psychiatrist*, *34*, 529–532. https://doi.org/10.1192/pb.bp.110.030775

Santos, M. M., Rae, J. R., Nagy, G. A., Manbeck, K. E., Hurtado, G. D., West, P., … Kanter, J. W. (2017). A client-level session-by-session evaluation of behavioral activation's mechanism of action. *Journal of Behavior Therapy and Experimental Psychiatry*, *54*, 93–100. https://doi.org/10.1016/J.JBTEP.2016.07.003

Saxon, D., Firth, N., & Barkham, M. (2017). The relationship between therapist effects and therapy delivery factors: Therapy modality, dosage, and non-completion. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*, 705–715. https://doi.org/10.1007/s10488-016-0750-5

Scamardo, M., Bobele, M., & Biever, J. L. (2004). A new perspective on client dropouts. *Journal of Systemic Therapies*, *23*, 27–38. https://doi.org/10.1521/jsyt.23.2.27.36639

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

Schindler, A., Hiller, W., & Witthöft, M. (2013). What predicts outcome, response, and drop-out in CBT of depressive adults? A naturalistic study. *Behavioural and Cognitive Psychotherapy*, *41*, 365–370. https://doi.org/10.1017/S1352465812001063

Schlagert, H. S., & Hiller, W. (2017). The predictive value of early response in patients with depressive disorders. *Psychotherapy Research*, *27*, 488–500.

https://doi.org/10.1080/10503307.2015.1119329

Schneider, K. L., Panza, E., Handschin, B., Ma, Y., Busch, A. M., Waring, M. E., … Pagoto, S. L. (2016). Feasibility of pairing behavioral activation with exercise for women with type 2 diabetes and depression: The Get It Study pilot randomized controlled trial. *Behavior Therapy*, *47*, 198–212. https://doi.org/10.1016/j.beth.2015.10.005

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, *25*, 1–65. https://doi.org/10.1016/S0065-2601(08)60281-6

Scott, M. J. (2018). Improving Access to Psychological Therapies (IAPT) - The need for radical reform. *Journal of Health Psychology*, *23*, 1136–1147. https://doi.org/10.1177/1359105318755264

Seedat, S., Scott, K. M., Angermeyer, M. C., Berglund, P., Bromet, E. J., Brugha, T. S., … Kessler, R. C. (2009). Cross-national associations between gender and mental disorders in the World Health Organization World Mental Health Surveys. *Archives of General Psychiatry*, *66*, 785–795. https://doi.org/10.1001/archgenpsychiatry.2009.36

Seligman, M. E. P. (1973). Fall into helplessness. *Psychology Today*, *7*, 43–48.

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist*, *50*, 965–974. https://doi.org/10.1037/0003-066X.50.12.965

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

Sheeran, P. (2002). Intention — behavior relations : A conceptual and empirical review. *European Review of Social Psychology*, *12*, 1–36. https://doi.org/10.1080/14792772143000003

Sheeran, P., Aubrey, R., & Kellett, S. (2007). Increasing attendance for psychotherapy: Implementation intentions and the self-regulation of attendance-related negative affect. *Journal of Consulting and Clinical Psychology*, *75*, 853–863. https://doi.org/10.1037/0022-006X.75.6.853

Sheeran, P., & Orbell, S. (2000). Using implementation intentions to increase attendance for cervical cancer screening. *Health Psychology*, *19*, 283–289. https://doi.org/10.1037/0278-6133.19.3.283

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

Shidhaye, R., Lund, C., & Chisholm, D. (2015). Closing the treatment gap for mental, neurological and substance use disorders by strengthening existing health care platforms: Strategies for delivery and integration of evidence-based interventions. *International Journal of Mental Health Systems*, *9*, 40. https://doi.org/10.1186/s13033-015-0031-9

Shrier, I., Boivin, J. F., Steele, R. J., Platt, R. W., Furlan, A., Kakuma, R., … Rossignol, M. (2007). Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *American Journal of Epidemiology*, *166*, 1203–1209. https://doi.org/10.1093/aje/kwm189

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

Smout, M., Davies, M., Burns, N., & Christie, A. (2013). Development of the Valuing Questionnaire (VQ). *Journal of Contextual Behavioral Science*, *3*, 164–172. https://doi.org/10.1016/j.jcbs.2014.06.001

*Soleimani, M., Mohammadkhani, P., Dolatshahi, B., Alizadeh, H., Overmann, K. A., & Coolidge, F. L. (2015). A comparative study of group behavioral activation and cognitive therapy in reducing subsyndromal anxiety and depressive symptoms. *Iranian Journal of Psychiatry*, *10*, 71–78.

Spielmans, G. I., Pasek, L. F., & McFall, J. P. (2007). What are the active ingredients in cognitive and behavioral psychotherapy for anxious and depressed children? A meta-analytic review. *Clinical Psychology Review*, *27*, 642–654. https://doi.org/10.1016/j.cpr.2006.06.001

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

Stangier, U., Ukrow, U., Schermelleh-Engel, K., Grabe, M., & Lauterbach, W. (2007). Intrapersonal conflict in goals and values of patients with unipolar depression. *Psychotherapy and Psychosomatics*, *76*, 162–170. https://doi.org/10.1159/000099843

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

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

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

Swift, J. K., & Callahan, J. L. (2008). A delay discounting measure of great expectations and the effectiveness of psychotherapy. *Professional Psychology: Research and Practice*, *39*, 581–588. https://doi.org/10.1037/0735-7028.39.6.581

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., Whipple, J. L., & Kominiak, N. (2012). Practice recommendations for reducing premature termination in therapy. *Professional Psychology: Research and Practice*, *43*, 379–387. https://doi.org/10.1037/a0028291

Tadić, A., Helmreich, I., Mergl, R., Hautzinger, M., Kohnen, R., Henkel, V., & Hegerl, U. (2010). Early improvement is a predictor of treatment outcome in patients with mild major, minor or subsyndromal depression. *Journal of Affective Disorders*, *120*, 86–93. https://doi.org/10.1016/j.jad.2009.04.014

Ten Have, M., De Graaf, R., Ormel, J., Vilagut, G., Kovess, V., & Alonso, J. (2010). Are attitudes towards mental health help-seeking associated with service use? Results from the European study of epidemiology of mental disorders. *Social Psychiatry and Psychiatric Epidemiology*, *45*, 153–163. https://doi.org/10.1007/s00127-009-0050-4

Thase, M. E., Greenhouse, J. B., Frank, E., Reynolds, C. F., Pilkonis, P. A., Hurley, K., … Kupfer, D. J. (1997). Treatment of major depression with psychotherapy or psychotherapy-pharmacotherapy combinations. *Archives of General Psychiatry*,

54, 1009–1015.

Thase, M. E., Simons, A. D., & Reynolds, C. F. (1993). Psychobiological correlates of poor response to cognitive behavior therapy: Potential indications for antidepressant pharmacotherapy. *Psychopharmacology Bulletin, 29*, 293–301.

The English Index of Multiple Deprivation (IMD). (2015). Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/465791/English_Indices_of_Deprivation_2015_-_Statistical_Release.pdf

Therapy deficit. (2012, September 26). *Nature Editorial*, *489*, 473–474. https://doi.org/10.1038/489473b

Thibodeau, M. A., Quilty, L. C., De Fruyt, F., De Bolle, M., Rouillon, F., & Bagby, R. M. (2015). Latent classes of nonresponders, rapid responders, and gradual responders in depressed outpatients receiving antidepressant medication and psychotherapy. *Depression and Anxiety*, *32*, 213–220. https://doi.org/10.1002/da.22293

*Thompson, L. W., Gallagher, D., Nies, G., & Epstein, D. (1983). Evaluation of the effectiveness of professionals and nonprofessionals as instructors of "Coping with Depression" classes for elders. *The Gerontologist*, *23*, 390–396. https://doi.org/10.1093/geront/23.4.390

*Thompson, Larry W., & Gallagher, D. (1983). A psychoeducational approach for treatment of depression in elders. *Psychotherapy in Private Practice*, *1*, 25–28. https://doi.org/10.1300/J294v01n03_07

Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, *21*, 1559–1573.

https://doi.org/10.1002/sim.1187

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

Tryon, G. S. (1999). Counseling dropout relative to client attractiveness, disturbance, and expected number of sessions. *Counselling Psychology Quarterly*, *12*, 285–291. https://doi.org/10.1080/09515079908254098

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

University College London. (2018). CBT Competency Framework. Retrieved July 29, 2019, from https://www.ucl.ac.uk/clinical-psychology/competency-maps/cbt-map.html

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

*Van Den Hout, J. H. C., Arntz, A., & Kunkels, F. H. J. (1995). Efficacy of a self-control therapy program in a psychiatric day-treatment center. *Acta Psychiatrica Scandinavica*, *92*, 25–29. https://doi.org/10.1111/j.1600-0447.1995.tb09538.x

Van, H. L., Dekker, J., Peen, J., Van Aalst, G., & Schoevers, R. A. (2008). Identifying patients at risk of complete nonresponse in the outpatient treatment of depression.

Psychotherapy and Psychosomatics, *77*, 358–364. https://doi.org/10.1159/000151389

Varley, R., Webb, T. L., & Sheeran, P. (2011). Making self-help more helpful: A randomized controlled trial of the impact of augmenting self-help materials with implementation intentions on promoting the effective self-management of anxiety symptoms. *Journal of Consulting and Clinical Psychology*, *79*, 123–128. https://doi.org/10.1037/a0021889

Veale, D. (2008). Behavioural activation for depression. *Advances in Psychiatric Treatment*, *14*, 29–36. https://doi.org/10.1192/apt.bp.107.004051

Vittengl, J. R., Jarrett, R. B., Weitz, E., Hollon, S. D., Twisk, J., Cristea, I., … Cuijpers, P. (2016). Divergent outcomes in cognitive-behavioral therapy and pharmacotherapy for adult depression. *American Journal of Psychiatry*, *173*, 481–490. https://doi.org/10.1176/appi.ajp.2015.15040492

Von Korff, M., & Goldberg, D. (2001). Improving outcomes in depression. *British Medical Journal*, *323*, 948–949. https://doi.org/10.1136/bmj.324.7339.737

Von Korff, M., Ormel, J., Katon, W., & Lin, E. H. B. (1992). Disability and depression among high utilizers of health care: A longitudinal analysis. *Archives of General Psychiatry*, *49*, 91–100. https://doi.org/10.1001/archpsyc.1992.01820020011002

Vowles, K. E., & McCracken, L. M. (2008). Acceptance and values-based action in chronic pain: A study of treatment effectiveness and process. *Journal of Consulting and Clinical Psychology*, *76*, 397–407. https://doi.org/10.1037/0022-006X.76.3.397

Walker, E. R., & Druss, B. G. (2015). Rate and predictors of persistent major depressive disorder in a nationally representative sample. *Community Mental Health Journal*,

51, 701–707. https://doi.org/10.1007/s10597-014-9793-9

Waller, G. (2009). Evidence-based treatment and therapist drift. *Behaviour Research and Therapy*, *47*, 119–127. https://doi.org/10.1016/j.brat.2008.10.018

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

Wang, P. S., Lane, M., Olfson, M., Pincus, H. A., Wells, K. B., & Kessler, R. C. (2005). Twelve-month use of mental health services in the United States: Results from the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 629–640. https://doi.org/10.1001/archpsyc.62.6.629

Webb, C. A., Trivedi, M. H., Cohen, Z. D., Dillon, D. G., Fournier, J. C., Goer, F., … Pizzagalli, D. A. (2018). Personalized prediction of antidepressant v. placebo response: evidence from the EMBARC study. *Psychological Medicine*, 1–10. https://doi.org/10.1017/S0033291718001708

Webb, T. L., Ononaiye, M. S. P., Sheeran, P., Reidy, J. G., & Lavda, A. (2010). Using implementation intentions to overcome the effects of social anxiety on attention and appraisals of performance. *Personality and Social Psychology Bulletin*, *36*, 612–627. https://doi.org/10.1177/0146167210367785

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin*, *132*, 249–268. https://doi.org/10.1037/0033-2909.132.2.249

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

Weck, F., Grikscheit, F., Jakob, M., Höfling, V., & Stangier, U. (2015). Treatment failure in cognitive-behavioural therapy: Therapeutic alliance as a precondition for an adherent and competent implementation of techniques. *British Journal of Clinical Psychology, 54*, 91–108. https://doi.org/10.1111/bjc.12063

Weitz, E. S., Hollon, S. D., Twisk, J., Van Straten, A., Huibers, M. J. H., David, D., … Cuijpers, P. (2015). Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs pharmacotherapy: An individual patient data meta-analysis. *JAMA - Journal of the American Medical Association Psychiatry*, *72*, 1102–1109. https://doi.org/10.1001/jamapsychiatry.2015.1516

Wells, K. B., Sherbourne, C., Schoenbaum, M., Duan, N., Meredith, L., Unützer, J., … Rubenstein, L. V. (2000). Impact of disseminating quality improvement programs for depression in managed primary care: A randomized controlled trial. *Journal of the American Medical Association*, *283*, 212–220. https://doi.org/10.1001/jama.283.2.212

Wersebe, H., Lieb, R., Meyer, A. H., Hoyer, J., Wittchen, H. U., & Gloster, A. T. (2017). Changes of valued behaviors and functioning during an Acceptance and Commitment Therapy Intervention. *Journal of Contextual Behavioral Science*, *6*, 63–70. https://doi.org/10.1016/j.jcbs.2016.11.005

*Wesson, M., Whybrow, D., Gould, M., & Greenberg, N. (2014). An initial evaluation of the clinical and fitness for work outcomes of a military group behavioural activation programme. *Behavioural and Cognitive Psychotherapy, 42*, 243–247. https://doi.org/10.1017/S135246581300043X

Westbrook, D., & Kirk, J. (2005). The clinical effectiveness of cognitive behaviour

therapy: Outcome for a large sample of adults treated in routine practice. *Behaviour Research and Therapy*, *43*, 1243–1261. https://doi.org/10.1016/j.brat.2004.09.006

Whiteford, H. A., Harris, M. G., McKeon, G., Baxter, A., Pennell, C., Barendregt, J. J., & Wang, J. (2013). Estimating remission from untreated major depression: a systematic review and meta-analysis. *Psychological Medicine*, *43*, 1569–1585. https://doi.org/10.1017/S0033291712001717

Whiteford, Harvey A., Ferrari, A. J., Degenhardt, L., Feigin, V., & Vos, T. (2015). The global burden of mental, neurological and substance use disorders: An analysis from the global burden of disease study 2010. *PLoS ONE*, *10*, 1–14. https://doi.org/10.1371/journal.pone.0116820

Whooley, M. A., & Wong, J. M. (2013). Depression and cardiovascular disorders. *Annual Review of Clinical Psychology*, *9*, 327–354. https://doi.org/10.1146/annurev-clinpsy-050212-185526

Wilson, D. B. (2010). Meta-analysis macros for SAS, SPSS, and Stata [Software]. Retrieved, April, 2017. Available from http://mason.gmu.edu/~dwilsonb/ma.html. Retrieved from http://mason.gmu.edu/~dwilsonb/ma.html

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

Wilson, K. G., & DuFrene, T. (2009). *Mindfulness for two: An acceptance and commitment therapy approach to mindfulness in psychotherapy*. Oakland, CA: New Harbinger.

Wilson, K. G., & Murrell, A. R. (2004). *Values work in acceptance and commitment*

therapy. (S. C. Hayes, V. M. Follette, & M. M. Linehan, Eds.). New York, NY: Guilford Press.

Wilson, K. G., Sandoz, E. K., Kitchens, J., & Roberts, M. (2010). The valued living questionnaire: Defining and measuring valued action within a behavioral framework. *The Psychological Record*, *60*, 249–272.

Wojnarowski, C., Firth, N., Finegan, M., & Delgadillo, J. (2019). Predictors of depression relapse and recurrence after Cognitive Behavioural Therapy: A systematic review and meta-analysis Investigators. *Behavioural and Cognitive Psychotherapy*, In press.

World Federation for Mental Health. (2012). *Depression: A global crisis*. Vancouver: World Federation for Mental Health. Retrieved from https://www.who.int/mental_health/management/depression/wfmh_paper_depression_wmhd_2012.pdf

World Health Organisation. (2004). *The World Health Report 2004: Changing History*. Geneva: World Health Organisation.

World Health Organisation. (2017). *Depression and other common mental disorders: Global health estimates*. Geneva: World Health Organisation. Retrieved from http://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf?sequence=1

*Wright, T. P. (2003). The effectiveness of behavioral activation group therapy: Treating comorbid depression on a specialized inpatient posttraumatic stress disorder unit for combat veterans. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *64*, 436. https://doi.org/10.1038/s41467-017-00982-x

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

Zemestani, M., Davoodi, I., Honarmand, M. M., Zargar, Y., & Ottaviani, C. (2016). Comparative effects of group metacognitive therapy versus behavioural activation in moderately depressed students. *Journal of Mental Health*, *25*, 479–485. https://doi.org/10.3109/09638237.2015.1057326

Zimmerman, M., Ellison, W., Young, D., Chelminski, I., & Dalrymple, K. (2015). How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Comprehensive Psychiatry*, *56*, 29–34. https://doi.org/10.1016/j.comppsych.2014.09.007

APPENDICES

Appendix A: Example meta-analysis search strategy (PsycINFO database)

No.	Search Term	Results
1	Behavior Therapy/	13011
2	Group Psychotherapy/	18127
3	Behavioural adj activation.ti,ab	209
4	Behavioral adj activation.ti,ab	1283
5	activity schedul*.tw	239
6	Pleasant event*.tw	215
7	Pleasurable event*.tw	24
8	Rewarding event*.tw	58
9	Pleasant activit*.tw	151
10	Pleasurable activit*.tw	154
11	Rewarding activit*.tw	78
12	Behavior Therap*.ti,ab	10540
13	Behavioral Therap*.ti,ab	12112
14	Behaviour therap*.ti,ab	2239
15	Behavioural therap*.ti,ab	2784
16	Behavioral intervention*.ti,ab	7123
17	Behavioural intervention*.ti,ab	936
18	1 to 17 (combined with OR)	58438
19	Exp "Depression (emotion)"	23078
20	Exp Major Depression	108964
21	Depression.ti,ab	197549
22	Depressive.ti,ab	84048
23	Depressed.ti,ab	42813
24	Mood disorder*.ti,ab	12691
25	Depressive disorder*.ti,ab	24374
26	19 to 25 (combined with OR)	254971
27	psychotherapeutic outcomes/	4526
28	treatment effectiveness evaluation/	20480
29	clinical trials/	9924
30	Efficac*.ti,ab	117037
31	Effectiv*.ti,ab	339434
32	27 to 31 (combined with OR)	440064
33	18 AND 26 AND 32	4633
34	Limit to humans	4532
35	Limit to adults (18+)	2502
	Total	**2502**

Appendix B: Quality ratings for studies included in the meta-analysis (using the modified Downs & Black scale)

Study First Author	Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	Total
Fuchs	1977	1	1	1	1	2	1	0	0	0	1	0	0	0	0	0	1	1	0	0	1	1	0	1	0	0	1	0	**14**
Shaw	1977	1	1	1	1	2	1	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	0	1	0	0	0	0	**14**
Beysner	1978	1	1	1	1	2	1	1	0	1	0	1	1	0	0	0	1	1	0	1	1	1	0	1	0	0	1	0	**18**
Barrera	1979	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	1	0	0	0	0	**8***
Cantanese	1979	1	0	1	1	1	0	1	0	1	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	**10***
Rehm	1979	1	1	1	1	1	1	0	0	1	0	1	0	0	0	0	1	1	0	0	1	1	1	1	0	0	0	0	**14**
Comas-Diaz	1981	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	1	1	0	1	1	1	0	1	0	0	0	0	**13**
Gallagher	1981	1	1	1	1	2	1	0	0	1	0	0	0	1	0	0	1	1	0	0	1	1	0	1	0	0	0	0	**14**
Rehm	1981	1	1	1	1	1	1	1	0	1	0	1	1	0	0	1	1	0	0	0	1	1	0	1	0	0	0	0	**15**
Kornblith	1983	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	**17***
Thompson; Thompson	1983a 1983b	0	1	0	1	1	1	1	0	1	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	**9**
Rehm	1987	1	1	1	0	1	1	1	0	1	0	1	1	0	0	1	1	1	1	1	1	1	0	0	0	0	1	0	**17**
Lovett	1988	0	1	0	1	1	1	1	0	0	1	0	0	0	0	0	1	0	1	0	1	1	0	1	0	0	0	0	**11***
Brand	1992	0	1	1	1	2	1	1	1	1	0	0	1	1	0	1	1	0	1	0	1	1	0	0	0	0	0	0	**16***
van dan Hout	1995	1	1	1	1	2	0	1	0	0	0	1	1	1	0	0	1	1	0	0	1	1	0	1	0	0	0	0	**15**
Gallagher-Thompson	2000	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	0	1	0	1	1	0	1	0	0	1	0	**18***
Wright	2003	1	1	1	1	2	1	1	0	1	1	0	0	1	0	0	1	1	1	1	1	1	0	0	0	0	1	0	**18**
Porter	2004	1	1	1	1	1	0	0	0	0	1	0	0	1	0	0	1	1	1	0	1	1	0	0	0	0	0	0	**12***
Daughters	2008	1	1	1	1	2	1	1	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	0	**22***
Houghton	2008	0	0	1	1	1	1	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	**11**
Norton	2010	1	1	1	1	1	1	1	0	1	1	0	0	1	0	1	1	0	1	0	1	1	0	0	0	1	1	1	**18**
Magidson	2011	1	1	1	1	2	0	0	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	1	0	0	**20**
Magidson	2014	1	1	0	1	1	1	1	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	**10**
Wesson	2014	0	1	1	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	**9***
Fereidooni	2015	0	1	1	1	1	1	1	0	0	1	0	0	1	0	0	1	1	0	0	1	1	1	1	0	0	0	0	**14***

Soleimani	2015	1	1	1	1	2	1	1	0	1	0	0	0	0	0	0	1	0	0	1	1	1	0	1	0	0	0	1	**15**
Kellett	2017	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	0	1	0	1	0	0	0	0	1	0	0	**16***
Study not included in the quantitative analysis																														
Zemestani	2016	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	1	1	1	1	1	1	0	1	0	1	1	0	**20***

*Indicates double rated studies

Reporting:

Q1. Is the hypothesis/aim/objective of the study clearly described?

Q2. Are the main outcomes to be measured clearly described in the Introduction or Methods section?

Q3. Are the characteristics of the patients included in the study clearly described?

Q4. Are the interventions of interest clearly described?

Q5. Are the distributions of principal confounders in each group of subjects to be compared clearly described?

Q6.Are the main findings of the study clearly described?

Q7. Does the study provide estimates of the random variability in the data for the main outcomes?

Q8. Have all important adverse events that may be a consequence of the intervention been reported?

Q9. Have the characteristics of patients lost to follow-up been described?

Q10.Have actual probability values been reported (e.g. 0.035 rather than <0.05) for the main outcomes except where the probability value is less than 0.001?

External validity:

Q11. Were the subjects asked to participate in the study representative of the entire population from which they were recruited?

Q12. Were those subjects who were prepared to participate representative of the entire population from which they were recruited?

Q13. Were the staff, places, and facilities where the patients were treated, representative of the treatment the majority of patients receive?

Internal validity – bias:

Q14. Was an attempt made to blind study subjects to the intervention they have received?

Q15. Was an attempt made to blind those measuring the main outcomes of the intervention?

Q16. If any of the results of the study were based on "data dredging", was this made clear?

Q17. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?

Q18. Were the statistical tests used to assess the main outcomes appropriate?

Q19. Was compliance with the intervention/s reliable?

Q20. Were the main outcome measures used accurate (valid and reliable)?

Internal validity – confounding:

Q21. Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?

Q22. Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?

Q23. Were study subjects randomised to intervention groups?

Q24. Was the randomised intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?

Q25. Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?

Q26. Were losses of patients to follow-up taken into account in the analysis?

Power: Q27. Was a power calculation performed and implemented to ensure the study had a sample size with sufficient power?

Appendix C: Evidence of ethical approval for three empirical studies in Chapter 4, 5 & 6 (Study protocol submitted to research ethics committee, letter confirming ethical approval, amendment documentation and approval)

Research Protocol

Funded by the Howard Morton Trust

Effect of treatment augmentations embedded in behavioural activation group therapy on reducing drop-out and stasis rates in depression

Mel Simmonds-Buckley

PhD Student Researcher

Department of Psychology, University of Sheffield

Supervisors:

Dr Stephen Kellett,

Clinical Psychology Unit, University of Sheffield,
Centre for Psychological Therapies Research, Sheffield Health and Social Care NHS Foundation Trust

Professor Glenn Waller,

Clinical Psychology Unit, University of Sheffield

NHS Collaboration Advisor:

Jennie Hague,

Clinical Psychology Unit, University of Sheffield

Improving Access to Psychological Therapies (IAPT), Sheffield Health and Social Care NHS Foundation Trust

Background

Depression, characterised by low mood and diminished pleasure in activities, can have a marked impact on individuals social and work functioning, affecting overall quality of life (Papakostas et al., 2004). It is one of the most prevalent mental health disorders (Ferrari et al, 2013) and represents both a considerable rate of suffering as well as a burden to the economy in terms of providing healthcare and lost employment. People suffering with depression experience thoughts centred around beliefs about their own inadequacies, heightening feelings of helplessness. Learned helplessness plays a key role in depression (Seligman, 1992) through the perception that depressive symptoms are unchangeable due to a previously learned lack of control acquired from prior inescapable aversive situations. Such feelings of uncontrollability interfere with the ability and motivation to learn new coping skills leading to stymied or limited attempts at change. Passive and avoidant behaviour causes negative thoughts and feelings to deteriorate further and therefore functions as a maintaining factor in depression.

Evidenced-based psychotherapy for depression has shown vast development and an established array of empirically validated treatments are now available. One of the most effective more recent approaches is behavioural activation (BA), an established treatment for depression based on principles of behaviour theory (Ferster, 1973). The function of behaviour and the role of reinforcement, both positive and negative, in maintaining depression is a key component of BA (Martell, Addis and Jacobson, 2001). Limited passive and avoidant behaviours associated with depression reduces an individual's exposure to meaningful and pleasurable events, resulting in less positive reinforcement of non-depressed behaviours. While avoidant coping behaviours allow relief from negative stimuli in the short-term, they act as negative reinforcement for depressive behaviours in the long-term. BA adopts a positive pragmatic approach, viewing depression as a consequence of such responses and contexts to negative thoughts and feelings. BA is grounded in activity scheduling through the use of between-session work to teach patients to plan alternative behaviour and re-engage with their life (Martell et al., 2001). Increased activity can tackle avoidance and increase engagement with valued living, resulting in increased exposure to positive reinforcement of non-depressed valued behaviours and in turn reduce symptoms of depression (Lewinsohn & Libet, 1972).

BA is clinically effective and has been consistently shown to reduce symptoms of depression, demonstrating equal efficaciousness as cognitive behavioural therapy (CBT) and anti-depressant medication treatments which continues to be evident at follow-up (Cuijpers, van Straten, & Warmerdam, 2007; Dimidjian et al., 2006; Dimidjian & Davies, 2009; Ekers et al., 2014; Jacobson et al., 1996; Mazzucchelli, Kane, & Rees, 2009). Furthermore, BA has been shown to be effective at treating all severities of depression (Dimidjian et al., 2006), suitable in a range of community and in-patient treatment settings (Veale, 2008), utilised for diverse populations (Mazzucchelli et al., 2009) and also effective delivered in a group format (Houghton, Curran & Saxon, 2008; Kellett, Bliss, Simmonds-Buckley, & Waller, 2016). Consequently, BA is National Institute of Health and Clinical Excellence (NICE, 2009) guideline recommended treatment for depression and Improving Access to Psychological Therapies (IAPT) services deliver the BA protocol nationwide in both low intensity and high intensity forms (Clark, 2011).

However, in spite of the delivery of evidenced-based practice, a considerable number of patients accessing treatment do not benefit from the intervention delivered. This emphasises that many more improvements could be made to interventions to improve outcomes rates and hence drive down stasis outcomes 'Stasis' can be defined as a patient who has shown neither positive nor negative clinical or reliable change in their outcome measures following treatment. Previous literature has identified the diverse definition of when a treatment fails to work (Hopko, Magidson & Lejuez, 2011; Lambert, 2011) including patients prematurely dropping-out, showing little response, deteriorating or relapsing. As a result of this lack of universally accepted criterion, stasis outcomes have been relatively overlooked as

deterioration outcomes have received more attention (Dimidjian & Hollon, 2010; Lambert, 2011). Similar precedents are evident in clinical practice, for example the IAPT initiative targets services to achieve 50% recovery rates, which inadvertently reinforces the idea of the 'forgotten 50%' who do not benefit from treatment and continue to psychologically suffer. The rate of depression stasis evident in overall IAPT psychotherapy outcomes, as well as BA specific research suggests up to 50% of patients potentially do not benefit from treatment (Chan & Adams, 2014; Firth, Barkham, Kellett, & Saxon, 2015; Green, Barkham, Kellett, & Saxon, 2014; Kellett et al., in press).

Little is known about the reasons why evidenced-based practice has little impact on some people and it is therefore imperative that stasis does not continue to be overlooked. Continued high rates of depression stasis outcomes will only further add to the economic burden and prolong patient suffering and hopelessness. Cost analyses have shown the indirect costs of depression exceed the costs of providing depression-associated healthcare (Thomas & Morris, 2003), therefore it makes sense to focus resources on depression stasis reduction to moderate the burden. Investigation of this subset of patients and why they are not responding can enable the optimisation of intervention effectiveness.

The relative simplicity of behavioural interventions means that BA is particularly well suited to the modification and addition to treatment protocols to potentially reduce stasis outcomes. Hopko et al (2011) identified the need to investigate non-response in behaviour therapy in its own right to progress the treatment field. The easily disseminated and simple, parsimonious principles BA is based upon provides a good opportunity for this investigation enabling minimal disruption without affecting the integrity of the intervention. Hopko et al (2011) proposed potential reasons behind patients' lack of response to BA, highlighting the importance of treatment compliance. The majority of the fundamental work to produce change in BA occurs within the between-session activity scheduling work. Therefore, patient's engagement and compliance with these activities is crucial to achieving their goals and experiencing a positive outcome (Addis & Jacobson, 2000; Burns & Spangler, 2000; Kazantzis, Deane, & Ronan, 2006).

It is a common phenomenon for individuals to set goals but have difficulty implementing them and the procrastination and avoidance of depression only adds to this difficulty. Implementation intentions are specific plans about how, when and where goals will be acted upon, formed using an if-then format in order to effectively implement actions (Gollwitzer, 1999). There is strong evidence demonstrating they can help achieve goal attainment (Gollwitzer & Sheeran, 2006) and implementation intentions have empirical support with people with mental health disorders. Toli, Webb & Hardy's (2016) meta-analysis demonstrated the beneficial effect of if-then planning in facilitating goal achievement across various mental health problems. Toli et al (2016) suggested clinical implications of this evidence could manifest through the integration of such techniques into existing treatments, specifically in the context of between-session work such as activity scheduling. Prestwich & Kellar (2010) have elaborated further on different forms and moderators of implementation intentions, identifying collaborative implementation intentions planning as a technique that holds promise. As suggested by Toli et al (2016), collaborative if-then planning would lend itself well incorporated into the BA approach as this intervention needs the patient to change their behaviour in order to change. Based on these findings, it is hypothesised that incorporating implementation intentions to provide an improved collaborative framework for setting and completing between-session tasks in group BA will increase patients' engagement and compliance with these tasks, enabling increased goal attainment and thus result in better outcomes.

The dose-effect evidence base demonstrates that attendance is another key aspect associated with improved outcome (Howard, Kopta, Krause, & Orlinsky, 1986; Cahill et al, 2003). Dropping out of treatment before receiving an appropriate dose of therapy places patients at risk of experiencing a stasis outcome and presents another avenue within which BA could be enhanced. Patient therapy expectations predict attendance (Hansen, Hoogduin, Schaap, & de Haan, 1992). Underestimating the necessary dose of therapy required to experience any significant improvement is commonplace among patients (Swift & Callahan, 2008). However, the effect of outcome expectation on drop-out rates is relatively under-investigated and the evidence inconclusive. Delgadillo, Moreea, Murphy, Ali and Swift (2015) found that

orientation leaflets to influence expectations posted to patients prior to attending low intensity guided self-help interventions had no eventual effect on attendance. Swift and Callahan (2011) on the other hand, demonstrated patients who received dose-response information remained in therapy significantly longer. Both these studies employed psychoeducation prior to patients attending for treatment. It hypothesised that psychoeducation aligning patient expectations with the reality of the BA intervention dose-response effect will increase engagement and attendance and result in reduced rates of drop-out.

Despite the ongoing research attention on evidenced-based psychotherapy, mechanisms by which treatments produce change in depression symptoms are yet to be established definitively. If treatments are to be refined, it is important to firstly understand what is crucial to produce change and secondly identify mechanisms which are then not activated in incidences of stasis, preventing change. To identify psychotherapy mechanisms of change, studies investigate the effect of theoretically identified potential mediating variables on depression outcome. Commonly proposed psychotherapy outcome mediators have included the therapeutic relationship, negative thoughts and activation (Gaynor & Harris, 2008; Kazdin, 2009). Given that the majority of work to produce change in BA takes place in the between-session activities (Addis & Jacobson, 2000), theoretically it seems logical to investigate potential mechanisms that bring about change within this process. In depression there tends to be an inconsistency between behaviour and values. People with depression struggle to live in accordance with what they value in life and therefore do not experience the associated positive reinforcement, which impacts on their mood (Martell, Dimidjian & Herman-Dunn, 2010). Increasing engagement with valued living is therefore a potential mechanism of BA through which changes in depression symptoms are produced. Examining the extent to which patients change how they engage in valued living during BA and how this affects their treatment outcome provides opportunity to investigate whether this differs in responder and stasis outcomes. Such comparison of mediating effects could identify if this mechanism contributes to preventing responsivity in stasis patients. It is therefore hypothesised that valued living will be a mediator of outcome for responder patients but not stasis patients.

In summary, this project proposes to firstly investigate rates of stasis and associated predictors following the existing behavioural activation group (BAG) therapy delivered in an IAPT service, secondly test an implementation intention and psychoeducation modified BAG intervention on depression stasis and drop-out rates and finally investigate the role of valued living as a mediator of depression outcome.

Objectives

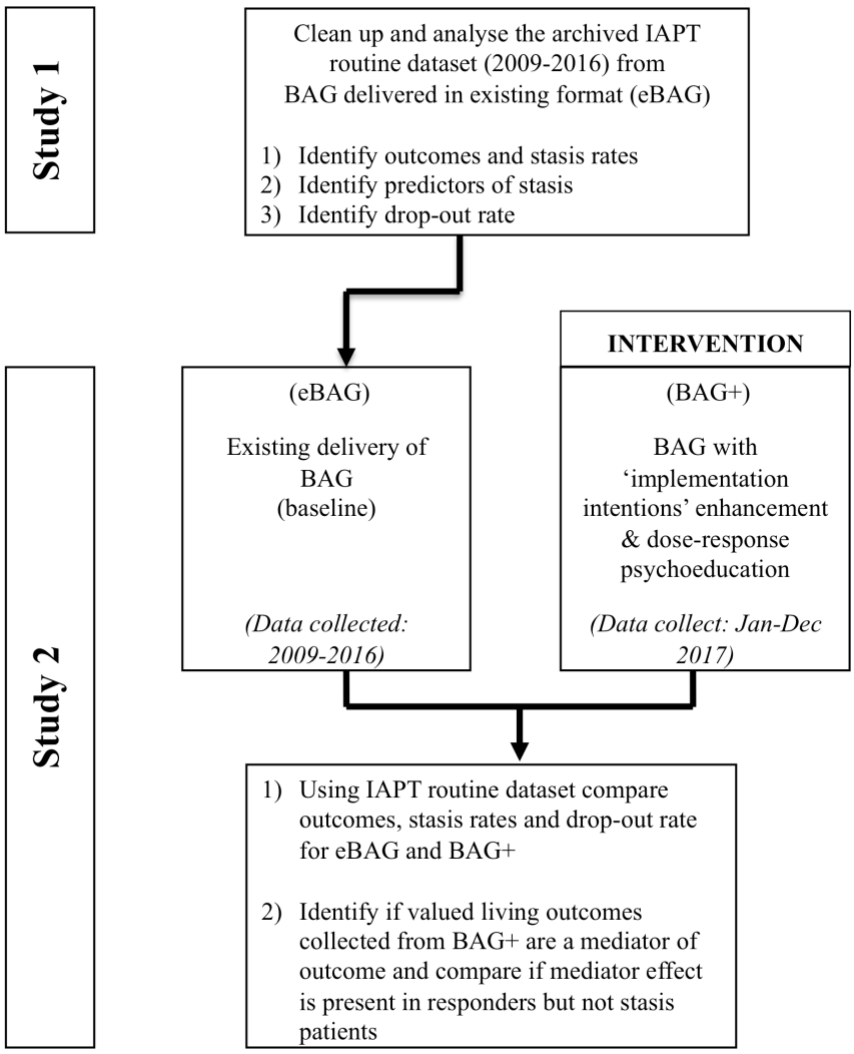
The objectives of this research are to (a) understand why some patients do not respond to depression treatment, (b) determine whether depression stasis rates can be reduced through enhancing an existing evidence-based treatment (BAG), (c) test whether implementation intentions increase the clinical effectiveness of BAG, (d) whether psychoeducation about dose-effect relations reduce dropout and (e) investigate a potential mechanisms of change in BAG.

Aims

1) Identify prevalence of stasis outcomes, associated predictors and drop-out rates in the existing BAG intervention
2) Reduce stasis outcomes and drop-out for BAG by implementing two augmentations into BAG (through embedding implementation intention and psychoeducation enhancements into the existing BAG treatment)
3) Compare depression recovery rates, stasis outcomes and drop-out for the augmented BAG intervention (BAG+) with the existing BAG intervention (eBAG)
4) Investigate the role of patients' engagement with valued living as a mediator of outcome by clustering outcomes and comparing effects in responder and stasis clusters

Plan of Investigation

Study Design Flowchart



Research Design

The project will consist of two connected but separate studies in a single-centre study based around the delivery of the BAG therapy delivered in the Sheffield IAPT service (part of Sheffield Health and Social Care NHS Foundation Trust).

- **Study one** will be a retrospective observational analysis of stasis outcome prevalence and predictors using the archived routine outcome measures. This data has been collected in routine practice from patients who have received BAG since in 2009.
- **Study two** will be an quasi-experimental design using two augmentations embedded into the existing BAG intervention to compare the patient depression outcomes (and subsequently stasis rates) and drop-out following the conclusion of treatment. One augmentation will be

theoretically informed and incorporate an implementation intentions enhanced aspect into the process of collaboratively agreeing activation 'homework' tasks and activities at the end of BAG sessions. The second augmentation will be informed by clinical practice evidence regarding psychoeducation being provided to BAG attendees about dose-response predictors of BAG outcomes to target drop-out rates. The drop-out rate, stasis rates and depression outcomes following the intervention augmentations (BAG+) will be compared with the baseline archived drop-out and stasis rates from the existing BAG intervention (eBAG) identified in study one.

STUDY ONE

A retrospective observational analysis of quantitative outcomes for patients who have been treated for depression by the existing BAG intervention. Retrospective design enables data collected in the past to identify stasis prevalence for an existing intervention without the need to assemble a study, recruit participants and data collect.

Aim

1) Identify prevalence of stasis outcomes and drop-out rates in the existing BAG intervention and explore associated predictors of stasis and drop-out.

Hypothesis

a) That 50% of patients who have received BAG treatment for depression will have a stasis outcome.
b) Initial severity of depression, co-morbidity and poor attendance will predict stasis.

Study Methodology

Participants

Study one will involve accessing anonymised routine outcome data from patients seen in routine NHS practice who accessed the Sheffield IAPT service for treatment for depression between 2009 and 2016 and received the existing BAG intervention. The sample size will be based on the amount of archived data available in order to get a complete overview of outcomes for BAG. As approximately six BAG groups have been/will be delivered each year from 2009-2016 with an approximate average of eight patients per group, the dataset is estimated to consist of more than N=350 patients. It is predicted a dataset of this estimated size will be sufficient to achieve statistical power. Post hoc power analysis using G*Power software will be conducted when the effect size has been established to ensure an appropriate sample size.

Inclusion criteria:

Patients who accessed the Sheffield IAPT service with depression as the primary presenting problem;

- Received the existing BAG intervention delivered between 2009-2016
- Completed a course of treatment defined as attending one or more BAG treatment sessions
- Patients with co-morbid anxiety symptoms can be included as long as depression is the primary diagnosis
- Aged 18 or over

Exclusion criteria:

- Primary diagnosis that is not depression

- Patients who have not attended at least one session of BAG
- Patients who receive BAG intervention after December 2016
- Aged under 18

Recruitment procedure and consent

The study will analyse data from all patients who have received the existing BAG treatment from 2009 until 2016. Patients give consent to the service for their data to be used for service evaluation purposes. All anonymised data will be extracted in an anonymised format by the data manager in the Sheffield NHS service. This will ensure that it does not contain any patient identifiable information (name, address, NHS number), before being handled by the student researcher. The data that will be extracted for the study will be existing routine demographic, service usage and outcome data collected routinely in IAPT services as part of the performance management of the service. The IAPT outcome measures are collected at every contact session as part of the IAPT Data Standard (IAPT, 2011). The Information Standards Board for Health and Social Care (ISB), the National Information Governance Board (NIGB) and the Review of Central Returns (RoCR) have approved the use of the IAPT Data Standard. The NIGB conditions specify that IAPT services do not need to obtain consent from patients to use the data for secondary purposes (IAPT, 2011). Patients are informed their treatment outcomes may also be used and shared in secondary analyses of treatment delivery and response, but that all data is anonymised and summarized so it is impossible for any individual patients to be identified (IAPT, 2011). Patients are informed that they have the right to request their data not be used in analyses. The student researcher will clean up the data, address missing data and remove any patients whose outcomes violate the inclusion criteria.

Materials and Measures

The extracted data of interest will be outcome scores on the IAPT minimum dataset, which will have been completed at every BAG session patient's attended and the patient attendance/drop-out information. The IAPT minimum dataset consists of the Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer & Williams, 2001), Generalised Anxiety Disorder Assessment (GAD-7; Spitzer, Kroenke, Williams & Lowe, 2006) and the Work and Social Adjustment Scale (WSAS; Mundt, Marks, Shear & Greist, 2002).

PHQ-9

The primary outcome will be patients' depression scores as measured by the PHQ-9 (appendix 1). The PHQ-9 is a nine item self-report questionnaire scored from 0-27 and is designed to detect depression within primary care settings. The PHQ-9 has sensitivity and specificity scores of 92% and 80% respectively at the ≥10 clinical cut off point (Gilbody, Richards, Brearly & Hewitt, 2007). Patients scores are classified into depression severity according the following thresholds; 0-4 = no depression, 5-9 = mild, 10-14 = moderate, 15-19 = moderately severe and 20-27 = severe. A score of 10 and above is classed as clinically significant symptoms of depression. Patients' PHQ-9 scores pre and post BAG treatment will be used to calculate the rate of recovery and stasis outcomes.

GAD-7

Anxiety scores will be collected as a secondary outcome, measured by the GAD-7 (appendix 2), which is a seven item self-report questionnaire scored from 0-21 designed to detect generalized anxiety disorder. The GAD-7 has sensitivity and specificity scores of 92% and 76% respectively at the ≥8 clinical cut off point (Swinson, 2006). Patients scores are classified into anxiety severity according the following thresholds; 0-4 = no anxiety, 5-10 = mild, 11-15 = moderate and 15-21 = severe. A score of 8 and above is classed as clinically significant symptoms of anxiety. GAD-7 scores pre and post BAG treatment will be used to calculate the effect of BAG on co-morbid anxiety symptoms and their role in contributing to depression stasis outcomes.

WSAS

Functional impairment as a result of mental health problems will also be collected as a secondary outcome, measured by the WSAS (appendix 3). The WSAS is a five item self-report measure of the impact the patients' symptoms are having on their work, home life, leisure activities and social relationships.

Data on patients' attendance and drop-out, BAG group variables (e.g. size, facilitators) and anonymised patient demographics (e.g. gender, age) and previous episodes of IAPT care will also be extracted from the IAPT Data Standard, as secondary outcomes to explore predictors of stasis.

Study Procedure

IAPT therapists will have inputted each patient's outcome measures/variables into the NHS secure electronic management system following each BAG session. The anonymised data for all BAG patients will be extracted from the NHS computer system by the data manager and given to the student researcher as a downloadable file. The download will be converted into Microsoft Excel and the anonymised data cleaned by the student researcher to address missing data and errors.

Data Analysis

Cleaning the data

Missing data will be analysed using the Intention-to-treat (ITT) principle. The final available measure will be used as the post score or if there is only one score available it will be assumed there was no change. Patients who do not score above the clinical cut-off for depression (score of ≥10 on PHQ-9) prior to commencing BAG will not be included in the analysis to avoid a floor effect when calculating stasis outcomes.

Statistical analysis

The dataset will be quantitatively analysed using SPSSv22, applying the following analysis;

1) Pre and post-treatment means and standard deviations will be calculated for each of the outcome measures and analysed using t-tests and treatment effect sizes (pre-treatment score minus post-treatment score divided by the pre-treatment standard deviation).
2) Reliable and clinical significant change criteria (Jacobson & Truax, 1991) will be applied to individual cases to calculate recovery and stasis prevalence rates for existing BAG. Reliable change is observed when the change in patients' scores exceeds the measurement error of the measure. For the IAPT routine measures (IAPT, 2014), *reliable improvement* is a decrease in scores of ≥6 on the PHQ-9 or ≥4 on the GAD-7; *reliable deterioration* is an increase in scores of ≥6 on the PHQ-9 or ≥4 on the GAD-7. Clinical change is observed when a patient's score moves from above the clinical cut-off on the measure pre-treatment (≥10 on the PHQ-9; ≥8 on the GAD-7) to below the clinical cut-off post-treatment (<10 on the PHQ-9; <8 on the GAD-7). Reliable and clinically significant recovery is recorded when a patient shows both reliable improvement and clinical change. The research team will define a *stasis outcome* as cases where neither reliable improvement nor reliable deterioration occurs. Patients who exhibit clinical change without experiencing reliable improvement will also come under the stasis definition.
3) Drop-out rates will be analysed by calculating the number of sessions attended by patients and the percentage of patients who dropped out of BAG.
4) Predictors of stasis outcomes using patient variables collected by the IAPT service (e.g. initial depression severity, drop-out, number of sessions attended etc.) will be calculated using regression and ANOVA.

Data Storage and Confidentiality

The electronic download file will be transferred using an encrypted data stick and stored on a password-protected computer, only accessible by the student researcher until the project is completed. All the data will be anonymised with no patient identifiable information included. Following completion, the data will be securely stored for future reference.

STUDY TWO

A quasi-experimental study will be used to test the effect of two embedded BAG treatment augmentations on treatment outcomes and to identify a potential outcome mediator for patients with a stasis outcome. A non-randomised design enables data to be collected which reflects routine clinical practice and address stasis outcomes as they occur in real-world services. A matched pairs design will be implemented in the analysis to allow comparison of the data with the retrospective data from study one (i.e. the baseline data).

Aim

1) To test the effect of implementation intentions on reducing the treatment stasis outcome rate from BAG
2) To test the effect of psychoeducation on reducing dropout from BAG
3) Compare the depression recovery rates, stasis outcomes and drop-out rates of the augmented BAG intervention (BAG+) with the baseline existing BAG intervention (eBAG).
4) Identify if valued living is a mediator of outcome and test whether a mediating effect is present in responder patients, but not stasis patients.

Hypothesis

a) There will be a significant reduction in depression stasis outcomes following the enhanced BAG intervention (BAG+) compared to the baseline existing BAG intervention (eBAG).
b) There will be a significant reduction in dropout rate following the enhanced BAG intervention (BAG+) compared to the baseline existing BAG intervention (eBAG).
c) A mediator effect of valued living on outcome will be found for responding patients following BAG+ but the effect will not be present for patients with a stasis outcome.

Study Methodology

Participants

Study two will recruit from patients who access the Sheffield IAPT service for treatment for depression and are offered the BAG intervention from January 2017 until December 2017. Six BAG groups are scheduled to run in a year starting every 2 months. Using a previous service evaluation of six groups delivered in a year, it is predicted that more than N=50 patients will attend BAG. An analysis of sample size using G*Power, indicated that a sample of 34 patients in the BAG+ group matched with 34 patients from the eBAG group (from study 1) would be needed to detect a small to medium effect size (d=0.3) with .80 power using a repeated measures, between factors ANOVA at p=0.05. As the intervention will be group-based, the effect of nesting of participants within BAG groups will need to be considered. This sample size was calculated based on the assumption of a low (<0.05) intra-class correlation coefficient (ICC) in terms of the correlation between depression outcomes for patients within each BAG group, as the data to calculate the ICC is not available yet. When the data is analysed, if the ICC is >0.05 (i.e. considered moderate) the sample of patients matched between the eBAG and BAG+ groups will be adjusted to take into account the ICC (i.e. the amount of dependency between outcomes for patients within the same BAG group) and the results reported accordingly.

Inclusion criteria:

Patients who access the IAPT service with depression as the primary presenting problem

- Patients with co-morbid anxiety symptoms can be included as long as depression is the primary diagnosis
- Are referred to and choose the BAG treatment option
- Able to attend the BAG+ intervention
- Aged 18 or over

Exclusion criteria:

- Primary diagnosis that is not depression
- Patients who do not choose BAG as a treatment option.
- Aged under 18

Recruitment procedure and consent

Participants will be recruited from patients attending the Sheffield IAPT service who are screened by an IAPT therapist and referred to the BAG treatment intervention from January 2017. Patients who sign up for the BAG group are sent a BAG information pack a week before the group start date. The patient information sheet (appendix 4) will be included in the BAG information pack to inform patients about the research before the first session. The research will be introduced to the patients at the first BAG session by the Cognitive Behavioural Psychotherapists facilitating the group. Patients will be given a demographic information sheet (appendix 5) and a consent form (appendix 6) to complete if they consent to their routine outcome data being used in the research.

Materials and Measures

The outcome measures collected will be the IAPT minimum dataset (PHQ-9, GAD-7 and WSAS) previously described in study 1, which will be completed by patients at every BAG session as part of IAPT routine outcome monitoring. Information about patient attendance will be collected to calculate drop-out and information about any current depression medication patients are also taking will be recorded for reference purposes. An additional measure, the Valued Living Questionnaire (VLQ; Wilson & Groom, 2002) which patients complete as part of BAG session 4 to assess valued living will also be administered to patients pre BAG and post BAG for the purpose of this research.

VLQ

VLQ scores will be collected as a secondary outcome. The VLQ is a 20-item clinical tool (appendix 7) designed to measure the extent that patients engage with the valued areas of their lives (valued living). It consists of Importance and Consistency subscales, both of which are rated on a 10-point Likert scale for 10 domains of valued living; family, marriage/couples/intimate relationships, parenting, friendships, work, education, recreation, spirituality, citizenship and physical self-care. The first part of the VLQ involves the patient rating how important they personally deem each of these 10 domains to be in terms of their life. The second part involves patients rating how consistently they have engaged with each valued domain in the past week. The VLQ quantifies how the patient's actual behaviours/engagement with activities matches up with the valued areas of their life to produce a weighted composite valued living score. The composite score is calculated by multiplying each domain's Importance and Consistency scores and the averaging the cross products to produce a score ranging from 1-100. Higher composite scores signify increased valued living. Preliminary results have shown the VLQ composite has adequate internal consistency (α= .65 - .74), good test-retest reliability (α= .75) and adequate construct validity (Wilson, Sandoz, Kitchens & Roberts, 2010).

Study Procedure

BAG Intervention

Patients who have been referred to BAG will receive an enhanced version of the existing BAG intervention. BAG is delivered as a step 3 high intensity intervention in the IAPT stepped care model. BAG consists of eight two-hour sessions delivered on a weekly basis for 8 weeks. BAG is delivered in a city centre non-clinical location. The BAG intervention is based on the 10-session group protocol developed by Houghton et al (2008) and has been adapted for use in primary care. Facilitators use a treatment manual to guide treatment and patients are given a workbook.

BAG+ Augmentations

The augmentations will consist of two strands; 1) implementation intentions to directly target reducing stasis and 2) psychoeducation to target reducing drop-out.

1. Implementation Intentions

The first augmentation will be a top-down theoretically informed 'implementation intentions' enhancement to target reducing the stasis outcome rate. The BAG facilitators will introduce and explain the idea of implementation intentions during the 'homework' section of the first session. An 'Achieving Your Goals' information sheet (see appendix 8) will be added to the patient workbook to explain the process instructions. Facilitators will set their own implementation intentions to model the approach to patients. At every subsequent session when homework is set at the end of each session, the facilitators will prompt patients to set homework in pairs using if-then planning (implementation intentions) to help them implement their goal rather than just commit to it. An implementation intentions worksheet (see appendix 9) for each session will be added to the patient workbook for patients to complete when setting the between-session activities. The worksheet includes an example and has space for patients to set their goals and write exactly how they intend to implement the actions required to achieve them. The 'if' section will indicate patients to identify a potential obstacle or a good opportunity to act on the intention and the 'then' section will direct patients to choose a suitable response/action to the identified opportunity oriented towards completing their desired goal. Patients will be asked to silently repeat their intention to themselves three times and then repeat it once out loud to their homework partner to verbally commit to the homework.

2. Dose-Response Psychoeducation

The second augmentation will be the bottom-up dose-response psychoeducation enhancement aimed at reducing the dropout rate. The psychoeducation will also be added to the treatment content in the first session (on the agenda) and the BAG facilitators will verbally reiterate the information when introducing the treatment as part of the first session. A psychoeducation information sheet (see appendix 10) will be added to session one in the patient workbook outlining the research evidence about the dose-response and outcome effect for BAG informing patients that;

1) The research findings show that attendance at least 4 BAG sessions is required for patients to show recovery.
2) BAG is effective at treating patients regardless of the severity of their depression.
3) BAG is also effective at reducing co-existing anxiety symptoms.

Patients will also be asked to read through the psychoeducation information sheet as part of their homework from session one.

BAG Session Outlines

Each BAG session is based on a different topic relevant to the principles of BA to encourage increased participation in rewarding personally meaningful activities. Patients are given between session work to complete, which is fed back and reviewed at each session. The augmented facilitators treatment manual and patient workbook with the above enhancements highlighted in red are available in appendix 11 and appendix 12 respectively. The session content is outlined below;

Session 1: Learn your patterns and start to change them

Session 2: Getting out of TRAPs and back on TRAC

Session 3: Taking action: a problem solving approach

Session 4: Values: the guide to who we are (VLQ)

Session 5: Developing responses to thinking, worry and rumination

Session 6: Making changes one step at a time

Session 7: Freeing yourself from mood dependence

Session 8: Building the relationships you want/tying it all together

BAG Facilitators

The Sheffield IAPT service has a pool of trained and experienced BAG facilitators who are all CBT therapists accredited by the British Association for Behavioural and Cognitive Psychotherapies (BABCP). Each course of BAG sessions will be facilitated by two CBT therapists.

Treatment Integrity

The content and structure of the existing BAG session outlines will not be altered. The treatment protocol has been delivered in the Sheffield IAPT service since 2009 and uses a manualised approach to ensure fidelity to the protocol. The augmentations have been developed by the research team in collaboration with the BAG facilitator lead to ensure the material is suitable for patients and the intervention being delivered. All BAG facilitators are experienced CBT therapists and have received additional in-house training specifically in delivering BAG. The research team will conduct a training workshop for BAG facilitators on the updated treatment manual to introduce the augmentations and how to deliver them in practice. In addition, the student researcher will attend the quarterly BAG meetings in the IAPT service to manage any research issues. Treatment adherence to the protocol will be assessed using a BAG adherence checklist (appendix 13) adapted from the adherence check used in a previous BA trial (Ekers, Richards, McMillan, Bland & Gilbody, 2011). Adherence will be checked and compared using self-report and an expert rater; 1) after each session the BAG facilitators will complete the session integrity measure to check self-report adherence and 2) the BAG facilitator lead will observe and rate one session from each course of BAG to provide an expert adherence check. To allow comparison, a version of the adherence checklist for the existing BAG treatment has also been developed and will be used for therapists to self-report and rate the adherence for sessions of the existing BAG intervention delivered until the end of 2016.

Data Collection

Data collection for the study will run for a year from January 2017 until December 2017 incorporating six BAG treatment groups.

<u>IAPT minimum dataset</u>

The IAPT minimum dataset will be administered to patients by the BAG facilitators at the start of every BAG session, as is routine practice in IAPT services. After each session the BAG facilitators will take the patient outcome scores back to the IAPT central office at St George's Hospital and input them into the NHS secure electronic management system. The student researcher will collect the completed measures for the patients who have given consent from the IAPT central office after each session. At the end of the data collection period, additional data about group characteristics (group size, gender ratio) and consenting patients (attendance, previous episodes of IAPT care) will be extracted by the data manager in the Sheffield IAPT service and a download will be given to the student researcher.

VLQ

BAG facilitators will administer the VLQ to patients at the first session (pre), the fourth session (as part of the values content) and at the last session (post) BAG treatment. As there is no procedure for inputting VLQ scores into the electronic management system, the student researcher will collect the completed VLQ's from the IAPT central office at St George's Hospital after the first, fourth and last session of each course of 8-session BAG.

After the student researcher has collected all the relevant completed consent forms, demographic information sheets, IAPT minimum dataset measures and VLQ measures for the patients who have given consent from the IAPT central office after each session, the data will be inputted into Microsoft Excel, cleaned up and anonymised by allocating participant identification numbers.

Data Analysis

Cleaning the data

Missing data will be analysed using the ITT principle. The final available measure will be used as the post score or if there is only one score available (with regards to the VLQ administered pre and post) it will be assumed there was no change. Patients who do not score above the clinical cut-off for depression (score of ≥10 on PHQ-9) prior to commencing BAG will not be included in the analysis to avoid a floor effect when calculating stasis outcomes.

Statistical analysis

The dataset will be quantitatively analysed using SPSSv22, applying the following analysis;

1) Pre and post-treatment means and standard deviations will be calculated for each of the routine measures and statistically analysed using t-tests and treatment effect sizes (pre-treatment score minus post-treatment score divided by the post-treatment standard deviation).
2) Reliable and clinical significant change criteria (Jacobson & Truax, 1991) will be applied to individual cases to calculate recovery and stasis prevalence rates for existing BAG (as outlined in study 1).
3) Drop-out rates will be analysed by calculating the number of sessions attended by patients and the percentage of patients who dropped out of BAG.
4) Recovery rates, stasis outcomes and drop-out rates from the BAG+ sample will be compared with a matched sample of eBAG data from study 1 using logistic regression while controlling for baseline co-variables. Comparable samples will be matched using propensity score matching (PSM) which accounts for variables that are known to predict outcomes for depression treatment within IAPT settings. Predictors will be identified from the existing literature (Delgadillo, Moreea & Lutz, 2016; Firth et al., 2015).
5) Mediation analysis will be performed on the relationship between the effect of BAG (independent variable) on PHQ-9 outcome scores (dependent variable) using the VLQ scores as a mediator variable. The depression outcomes will be clustered and the analysis will be performed separately on the responder patient outcomes and the stasis patient outcomes to see if there is evidence of a mediating effect for responders, which is not present for stasis patients.

Data Storage and Confidentiality

The electronic download file will be transferred using an encrypted data stick and stored on a password-protected computer, only accessible by the student researcher. The completed measures and consent forms will be stored in a locked filing cabinet on University premises which can only be accessed by the student researcher. Identification numbers will be used to input the data and ensure anonymity.

Ethical Considerations

The project will be conducted in collaboration with the Sheffield IAPT service (Sheffield Health and Social Care NHS Foundation Trust) and ethical approval will be sought from the NHS ethics committee. All patients involved in the study will continue to receive a routine clinical service delivered by the IAPT service, which will not be compromised by the research. The study has been designed to complement the current service delivery with minimum burden and disruption to NHS staff and service users. Associated ethical considerations are outlined below;

1. ***Routine Practice***

The IAPT service will continue to operate routine practice with regards assessing, referring and treating patients. The study has been designed to collect routine sessional monitoring data as the primary outcome in order to address the research question causing minimal disruption. The manualised BAG treatment protocol is an evidenced-based treatment that has been delivered in the service since 2009 and the underlying structure and content will remain the same. The only alteration to routine practice will be the two treatment augmentation additions, which have been developed in collaboration with the BAG facilitator lead. The only additional measure to be administered will be the VLQ pre and post BAG, which is already a component of the existing BAG delivered in session 4 (Values: the guide to who we are). Patients who attend the BAG group will be asked for informed consent and they can choose not to have their outcome data included in the study if they wish.

2. ***Risk and Patient Distress***

Any patients who experience any risk issues or distress during BAG sessions will be addressed by the BAG facilitators according to the IAPT service risk protocol. This process will remain unaffected by the study.

3. ***Confidentiality***

The electronic download will be transferred using an encrypted data stick and will be stored in a password-protected folder on a University computer which only the student researcher has access to. After the data has been uploaded, it will be removed from the data stick. All the archived data in study 1 will be anonymised so patient names will not be used at all. Identification numbers will be used for inputting and analysing the dataset from the study 2 BAG+ intervention. A separate file identifying the ID numbers will be stored separately. All hard copies of self-report measures and consent forms will be kept in a secure locked filing cabinet until the research project is completed. Following completion, the data will be securely stored for future reference.

4. ***Service Collaboration***

The BAG facilitator lead in the Sheffield service, Jennie Hague is an advisor to the research project and provides a link with the IAPT service. The studies have been developed in collaboration to ensure suitability. The student researcher will work closely with her to coordinate and implement the augmentations and collect the data. BAG facilitator meetings are scheduled quarterly and the student researcher will be on hand throughout the data collection to answer any questions or address any issues from the BAG facilitators arising from the research. The service will receive feedback on the project and findings will be reported back to aid service development and clinical practice.

5. ***Supervision***

The student researcher will have ongoing supervision to discuss any issues throughout the duration of the research project.

Resources and Cost

The student researcher is funded by the Howard Morton Trust, a scholarship awarded for research into depression. As the research is based around routine practice, it is not anticipated that there will be any additional costs. Should any costs arise, a £500 per year Research Training Support Grant is available as part of the scholarship funding.

Timetable for Research

Submission for ethical review:	June 2016
Proposed study 1 start date:	September 2016
Proposed study 2 start date:	January 2017
Data collection start date:	January 2017
Data collection end date:	December 2017
Study 2 data analysis begins:	January 2018

Benefits and Significance

This study has benefits for NHS patients, therapists and service providers. The research project could have a substantial and wide-reaching impact on improving outcomes for people suffering with depression, particularly those who currently find it difficult to benefit from the existing evidenced-based practice. It also offers an opportunity to promote and stimulate interest in group therapy as a more cost effective treatment for depression, as well as evaluate and update the existing BAG treatment and protocol. Sheffield IAPT service will receive an updated treatment manual, service reports to use with commissioners and involvement with developing BA masterclasses. The findings will contribute to the evidence base and be shared with IAPT services in order to improve the patient experience, support therapists, inform service policy and help shape the treatment of depression.

References

Addis, M. E., & Jacobson, N. S. (2000). A closer look at the treatment rationale and homework compliance in cognitive-behavioral therapy for depression. *Cognitive Therapy and Research, 24,* 313-326.

Burns, D. D., & Spangler, D. L. (2000). Does psychotherapy homework lead to improvements in depression in cognitive-behavioral therapy or does improvement lead to increased homework compliance? *Journal of Consulting and Clinical Psychology, 68,* 48-66.

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

Chan, S. W. Y., & 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*(6), 747-759. doi:10.1017/s1352465813000544

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry, 23*(4), 318-327. doi:10.3109/09540261.2011.606803

Cuijpers, P., van Straten, A., & Warmerdam, L. (2007). Behavioral activation treatments of depression: A meta-analysis. *Clinical Psychology Review, 27*(3), 318-326. doi:10.1016/j.cpr.2006.11.001

Delgadillo, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profilling and risk stratification. *Behaviour Research and Therapy, 79,* 15-22. doi: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 psychologcial therapy? Double-blind randomized controlled trial. *Journal of Clinical Psychology, 71*(12), 1139-1152. doi:10.1002/jclp.22228

Dimidjian, S., & Davis, K. J. (2009). Newer variations of cognitive behavioral therapy: Behavioral activation and mindfulness-based cognitive therapy. *Current Psychiatry Reports, 11*(6), 453-458.

Dimidjian, S., & Hollon, S. D. (2010). How would we know if psychotherapy were harmful? *American Psychologist, 65*, 21–33.

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., . . . Gollan, J. K. (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*(4), 658.

Ekers, D., Richards, D., McMillan, D., Bland, J. M., & Gilbody, S. (2011). Behavioural activation delivered by the non-specialist: phase II randomised controlled trial. *The British Journal of Psychiatry, 198*(1), 66-72. doi:0.1192/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*(6), 11. doi:10.1371/journal.pone.0100100

Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman G., Murray, C. J. L., … Whiteford, H. A. (2013). Burden of depressive disorders by country, sex, age, and year: Findings from the global burden of disease study 2010. *Plos Medicine, 10*(11), 1-12. doi:10.1371/journal.pmed.1001547

Ferster, C. (1973). A functional analysis of depression. *American psychologist, 28*(10), 857.

Firth, N., Barkham, M., Kellett, S., & 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

Gaynor, S. I., & Harris, A. (2008). Single-participant assessment of treatment mediators: Strategy description and examples from a behavioral activation intervention for depressed adolescents. *Behavior Modification, 32*(3), 372-402. doi:10.1177/0145445507309028

Gilbody, S., Richards, D., Brealey, S., & 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*(11), 1596-1602.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54*, 493–503.

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology, 38*, 69–119.

Green, H., Barkham, M., Kellett, S., & 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

Hansen, A. M. D., Hoogduin, C. A. L., Schaap, C., & de Haan, E. (1992). Do drop-outs differ from successfully treated obsessive-complusives? *Behaviour Research and Therapy, 30*(5), 547-550. doi:10.1016/0005-7967(92)90040-N

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: In Session, 67*(11), 11106-1116. doi: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*(2), 235-239. doi:10.1017/s1352465808004207

Howard, K. I., Kopta, S. M., Krause, M. S.., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist, 41*, 159-164.

Improving Access to Psychological Therapies (IAPT). (2014). *Measuring improvement and recovery*. Available at http://www.iapt.nhs.uk/data/.

Improving Access to Psychological Therapies (IAPT). (2011). *The IAPT data handbook.* Available at http://www.iapt.nhs.uk/silo/files/iapt-data-handbook-v2.pdf

Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., . . . Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology, 64*(2), 295-304. doi:10.1037//0022-006x.64.2.295

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*(1), 12-19. doi:10.1037//0022-006x.59.1.12

Kazantzis, N., Deane, F. P., & Ronan, K. R. (2006). Homework assignments in cognitive and behavioral therapy: A meta-analysis. *Clinical Psychology: Science and Practice, 7*, 189-202.

Kazdin, A. E. (2009). Understanding how and why psychotherapy leads to change. *Psychotherapy Research, 19*(4-5), 418-428. doi:10.1080/10503300802448899

Kellett, S., Bliss, P., Simmonds-Buckley, M., & Waller, G. (2016). Clinical effectiveness of behavioural activation delivered in a group format: A pilot study. Manuscript submitted for publication.

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

Lambert, M. J. (2011). What have we learned about treatment failure in empirically supported treatments? Some suggestions for practice. *Cognitive and Behavioural Practice, 18*, 413-420.

Lewinsohn, P. M., & Libet, J. (1972). Pleasant events, activity schedules, and depressions. *Journal of Abnormal Psychology, 79*(3), 291-295. doi:10.1037/h0033207

Martell, C. R., Addis, M. E., & Jacobson, N. S. (2001). *Depression in context: Strategies for guided action*. New York, NY: WW Norton & Co.

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

Mazzucchelli, T., Kane, R., & Rees, C. (2009). Behavioral activation treatments for depression in adults: A meta-analysis and review. *Clinical Psychology: Science and Practice, 16*(4), 383-411.

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. 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: Author. Available at www.nice.org.uk/CG90.

Papakostas, G. I., Peterson, T., Mahal, Y., Mischoulon, D., Nierenberg, A. A., & Fava, M. (2004). Quality of life assessments in major depressive disorder: A review of the literature. *General Hospital Psychiatry, 26*(1), 13-17. doi:10.1016/j.genhosppsych.2003.07.004

Prestwich, A., & Kellar, I. (2010). How can implementation intentions as a behaviour change intervention be improved? *European Review of Applied Psychology, 64*(1), 35-41. doi:10.1016/j.erap.2010.03.003

Seligman, M. E. P. (1992). *Helplessness: on depression, development, and death.* (2nd ed.). New York, NY: W. H. Freeman & Co Ltd.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder - The GAD-7. *Archives of Internal Medicine, 166*(10), 1092-1097. doi:10.1001/archinte.166.10.1092

Swift, J. K. & Callahan, J. L. (2008). A delay discounting measure of great expectations and the effectiveness of psychotherapy. *Professional Psychology: Research and Practice, 39*(6), 581-588. doi:10.1037/0735-7028.39.6.581

Swift, J. K. & Callahan, J. L. (2011). Decreasing treatment dropout by addressing expectations for treatment length. *Psychotherapy Research, 21*(2), 193-200. doi:10.1080/10503307.2010.541294

Swinson, R. P. (2006). The GAD-7 scale was accurate for diagnosing generalised anxiety disorder. *Evidence-Based Medicine, 11*(6), 184-184. doi:10.1136/ebm.11.6.184

Thomas, C. M., & Morris, S. (2003). Cost of depression among adults in England in 2000. *British Journal of Psychiatry, 183*(6), 514-519. doi:10.1192/00-000

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*(1), 69-90. doi:10.1111/bjc.12086

Veale, D. (2008). Behavioural activation for depression. *Advances in Psychiatric Treatment, 14*, 29-36. doi: 10.1192/apt.bp.107.004051

Wilson, K., & Groom, J. (2002). *The valued living questionnaire.* Avaliable from K. Wilson.

Wilson, K. G., Sandoz, E. K., Kitchens, J., & Roberts, M. (2010). The valued living questionnaire: Defining and measuring valued action within a behavioural framework. *The Psychological Record, 60*, 249-272.

Health Research Authority letter confirming research governance and ethical approval for studies conducted in Chapter 4, 5 and 6

Miss Melanie Simmonds-Buckley
PhD Research Student
University of Sheffield
Department of Psychology
University of Sheffield
Western Bank
S10 2TP

Email: hra.approval@nhs.net

11 November 2016

Dear Miss Simmonds-Buckley

Letter of HRA Approval

Study title:	**Effect of treatment augmentations embedded in behavioural activation group therapy on reducing drop-out and stasis rates in depression**
IRAS project ID:	**202197**
REC reference:	**16/YH/0324**
Sponsor	**University of Sheffield**

I am pleased to confirm that **HRA Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications noted in this letter.

Participation of NHS Organisations in England

The sponsor should now provide a copy of this letter to all participating NHS organisations in England.

Appendix B provides important information for sponsors and participating NHS organisations in England for arranging and confirming capacity and capability. **Please read *Appendix B* carefully**, in particular the following sections:

- *Participating NHS organisations in England* – this clarifies the types of participating organisations in the study and whether or not all organisations will be undertaking the same activities
- *Confirmation of capacity and capability* - this confirms whether or not each type of participating NHS organisation in England is expected to give formal confirmation of capacity and capability. Where formal confirmation is not expected, the section also provides details on the time limit given to participating organisations to opt out of the study, or request additional time, before their participation is assumed.
- *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* - this provides detail on the form of agreement to be used in the study to confirm capacity and capability, where applicable.

Further information on funding, HR processes, and compliance with HRA criteria and standards is also provided.

It is critical that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details and further information about working with the research management function for each organisation can be accessed from www.hra.nhs.uk/hra-approval.

Appendices

The HRA Approval letter contains the following appendices:

- A – List of documents reviewed during HRA assessment
- B – Summary of HRA assessment

After HRA Approval

The document *"After Ethical Review – guidance for sponsors and investigators",* issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

In addition to the guidance in the above, please note the following:

- HRA Approval applies for the duration of your REC favourable opinion, unless otherwise notified in writing by the HRA.
- Substantial amendments should be submitted directly to the Research Ethics Committee, as detailed in the *After Ethical Review* document. Non-substantial amendments should be submitted for review by the HRA using the form provided on the HRA website, and emailed to hra.amendments@nhs.net.
- The HRA will categorise amendments (substantial and non-substantial) and issue confirmation of continued HRA Approval. Further details can be found on the HRA website.

Scope

HRA Approval provides an approval for research involving patients or staff in NHS organisations in England.

If your study involves NHS organisations in other countries in the UK, please contact the relevant national coordinating functions for support and advice. Further information can be found at http://www.hra.nhs.uk/resources/applying-for-reviews/nhs-hsc-rd-review/.

If there are participating non-NHS organisations, local agreement should be obtained in accordance with the procedures of the local participating non-NHS organisation.

IRAS project ID	202197

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please email the HRA at hra.approval@nhs.net. Additionally, one of our staff would be happy to call and discuss your experience of HRA Approval.

HRA Training

We are pleased to welcome researchers and research management staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

Your IRAS project ID is **202197**. Please quote this on all correspondence.

Yours sincerely

Alison Thorpe
Senior Assessor

Email: hra.approval@nhs.net

Copy to: *Ms Deborah McClean, University of Sheffield, Sponsor Contact*
Mr Daniel Last, Sheffield Health & Social Care NHS Foundation Trust, Lead NHS R&D Contact

NIHR CRN Portfolio Applications Team

Notification of minor amendment to research protocol form

(relates to study conducted in Chapter 4)

Partner Organisations:
Health Research Authority, England NIHR Clinical Research Network, England
NHS Research Scotland NISCHR Permissions Co-ordinating Unit, Wales
HSC Research & Development, Public Health Agency, Northern Ireland

Notification of Non-Substantial/Minor Amendments(s) for NHS Studies

This template **must only** be used to notify NHS/HSC R&D office(s) of amendments, which are **NOT** categorised as Substantial Amendments.
If you need to notify a Substantial Amendment to your study then you MUST use the appropriate Substantial Amendment form in IRAS.

Instructions for using this template

- For guidance on amendments refer to http://www.hra.nhs.uk/research-community/during-your-research-project/amendments/
- This template should be completed by the CI and optionally authorised by Sponsor, if required by sponsor guidelines.
- This form should be submitted according to the instructions provided for NHS/HSC R&D at http://www.hra.nhs.uk/research-community/during-your-research-project/amendments/which-review-bodies-need-to-approve-or-be-notified-of-which-types-of-amendments/ . If you do not submit your notification in accordance with these instructions then processing of your submission may be significantly delayed.

1. Study Information

Full title of study:	Effect of treatment augmentations embedded in behavioural activation group therapy on reducing drop-out and stasis rates in depression
IRAS Project ID:	202197
Sponsor Amendment Notification number:	1
Sponsor Amendment Notification date:	23/02/17
Details of Chief Investigator:	
Name [first name and surname]	Melanie Simmonds-Buckley
Address:	Department of Psychology University of Sheffield Floor E, Cathedral Court 1 Vicar Lane
Postcode:	S1 2LT
Contact telephone number:	07986298350
Email address:	mksimmonds-buckley1@sheffield.ac.uk
Details of Lead Sponsor:	
Name:	University of Sheffield (Deborah McClean)
Contact email address:	D.McClean@sheffield.ac.uk
Details of Lead Nation:	
Name of lead nation *delete as appropriate*	**England** / Northern Ireland / Scotland / Wales
If England led is the study going through CSP? *delete as appropriate*	Yes / **No**
Name of lead R&D office:	Sheffield Health & Social Care NHS Foundation Trust (Contact: Daniel Last)

Partner Organisations:
Health Research Authority, England NIHR Clinical Research Network, England
NHS Research Scotland NISCHR Permissions Co-ordinating Unit, Wales
HSC Research & Development, Public Health Agency, Northern Ireland

2. Summary of amendment(s)

This template **must only** be used to notify NHS/HSC R&D office(s) of amendments, which are **NOT** categorised as Substantial Amendments.
If you need to notify a Substantial Amendment to your study then you MUST use the appropriate Substantial Amendment form in IRAS.

No.	Brief description of amendment *(please enter each separate amendment in a new row)*	Amendment applies to *(delete/ list as appropriate)*		List relevant supporting document(s), including version numbers *(please ensure all referenced supporting documents are submitted with this form)*		R&D category of amendment *(category A, B, C) For office use only*
		Nation	**Sites**	**Document**	**Version**	
1	Request for an additional subset of anonymised secondary data to be included in the secondary data download requested from the IAPT service. The additional data will not alter the methodology or design of the study. Amendment refers to a request for the secondary data to include the outcomes for all patients who have received a Behavioural Activation intervention (in all formats) for depression delivered in the service within the previously specified timeframe (2009-2016).	England	All sites	N/A	N/A	
2						
3						
4						
5						

[Add further rows as required]

Partner Organisations:
Health Research Authority, England NIHR Clinical Research Network, England
NHS Research Scotland NISCHR Permissions Co-ordinating Unit, Wales
HSC Research & Development, Public Health Agency, Northern Ireland

3. Declaration(s)

Declaration by Chief Investigator

☐ I confirm that the information in this form is accurate to the best of my knowledge and I take full responsibility for it.

☐ I consider that it would be reasonable for the proposed amendment(s) to be implemented.

Signature of Chief Investigator: …  …….

Print name: ……Melanie Simmonds-Buckley……

Date: …24th February 2017……

Optional Declaration by the Sponsor's Representative (as per Sponsor Guidelines)

The sponsor of an approved study is responsible for all amendments made during its conduct.

The person authorising the declaration should be authorised to do so. There is no requirement for a particular level of seniority; the sponsor's rules on delegated authority should be adhered to.

☐ I confirm the sponsor's support for the amendment(s) in this notification.

Signature of sponsor's representative: ……

Print name:……

Post: ……

Organisation:……

Date:……

The University Of Sheffield.

Melanie K Simmonds-Buckley <mksimmonds-buckley1@sheffield.ac.uk>

RE: IRAS 202197. Amendment assessment outcome 2017_04_03

3 messages

AMENDMENTS, Hra (HEALTH RESEARCH AUTHORITY) <hra.amendments@nhs.net> 3 April 2017 at 12:06
To: Melanie K Simmonds-Buckley <mksimmonds-buckley1@sheffield.ac.uk>
Cc: "D.McClean@sheffield.ac.uk" <D.McClean@sheffield.ac.uk>, "daniel.last@shsc.nhs.uk" <daniel.last@shsc.nhs.uk>

.

.

.

hra.amendments@nhs.net

.

.

.

Kind Regards

Steph Macpherson

HRA Senior Assessor

.

Health Research Authority

HRA, Ground Floor, Skipton House, 80 London Road, London, SE1 6LH
E: hra.approval@nhs.net

T: 02079722505

www.hra.nhs.uk

Appendix D: Outcome measures used in the thesis studies

Patient Health Questionnaire-9 (PHQ-9)

(Kroenke et al., 2001)

(used in Chapter 4, 5 & 6)

	Over the last 2 weeks, how often have you been bothered by any of the following problems?	Not at all	Several days	More than half the days	Nearly every day
1	Little interest or pleasure in doing things	0	1	2	3
2	Feeling down, depressed, or hopeless	0	1	2	3
3	Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4	Feeling tired or having little energy	0	1	2	3
5	Poor appetite or overeating	0	1	2	3
6	Feeling bad about yourself — or that you are a failure or have let yourself or your family down	0	1	2	3
7	Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8	Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3
9	Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3

PHQ9 total score ☐

(Data item 37 in the IAPT Data Standard)

The PHQ-9 is also available in the following languages:
Hindi, Punjabi, Bengali, Gujarati, Urdu

Visit http://www.iapt.nhs.uk/services/measuring-outcomes/ to download alternate language versions.

Generalised Anxiety Disorder-7 (GAD-7)

(Spitzer et al., 2006)

(used in Chapter 4 & 5)

Over the last 2 weeks, how often have you been bothered by any of the following problems?	Not at all	Several days	More than half the days	Nearly every day
1 Feeling nervous, anxious or on edge	0	1	2	3
2 Not being able to stop or control worrying	0	1	2	3
3 Worrying too much about different things	0	1	2	3
4 Trouble relaxing	0	1	2	3
5 Being so restless that it is hard to sit still	0	1	2	3
6 Becoming easily annoyed or irritable	0	1	2	3
7 Feeling afraid as if something awful might happen	0	1	2	3
	GAD7 total score (Data item 38 in the IAPT Data Standard)			☐

The GAD7 is also available in the following languages:
Hindi, Punjabi, Arabic, Bengali, Gujarati, Urdu

Visit http://www.iapt.nhs.uk/services/measuring-outcomes/ to download alternate language versions.

Work and Social Adjustment Scale (WSAS)

(Mundt et al., 2002)

(used in Chapter 4 & 5)

People's problems sometimes affect their ability to do certain day-to-day tasks in their lives. To rate your problems look at each section and determine on the scale provided how much your problem impairs your ability to carry out the activity.

1. **WORK** - if you are retired or choose not to have a job for reasons unrelated to your problem, please tick N/A (not applicable)

0	1	2	3	4	5	6	7	8	N/A
Not at all		Slightly		Definitely		Markedly	Very severely, I cannot work		☐

2. **HOME MANAGEMENT** – Cleaning, tidying, shopping, cooking, looking after home/children, paying bills etc

0	1	2	3	4	5	6	7	8
Not at all		Slightly		Definitely		Markedly	Very severely	

3. **SOCIAL LEISURE ACTIVITIES -** With other people, e.g. parties, pubs, outings, entertaining etc.

0	1	2	3	4	5	6	7	8
Not at all		Slightly		Definitely		Markedly	Very severely	

4. **PRIVATE LEISURE ACTIVITIES –** Done alone, e.g. reading, gardening, sewing, hobbies, walking etc.

0	1	2	3	4	5	6	7	8
Not at all		Slightly		Definitely		Markedly	Very severely	

5. **FAMILY AND RELATIONSHIPS –** Form and maintain close relationships with others including the people that I live with

0	1	2	3	4	5	6	7	8
Not at all		Slightly		Definitely		Markedly	Very severely	

W&SAS total score

(Data item 39 in the IAPT Data Standard)

Valued Living Questionnaire (VLQ)

(Wilson et al., 2010)

(used in Chapter 6)

Valued Living Questionnaire
Self-Care Assessment Part 1

Below are areas of life that are valued by some people. This questionnaire will help clarify your own quality-of-life in each of these areas. One aspect of quality-of-life involves the importance you put on different areas of living. Rate the importance of each area (by circling a number) on a scale of 1-10. A "1" means that area is *not at all important*. A "10" means that area is *very important*. Not everyone will value all of these areas, or value all areas the same. Rate each area according to **your own personal sense of importance**.

Area:	**not at all important**									**extremely important**
1) Family (other than marriage or parenting)	1	2	3	4	5	6	7	8	9	10
2) Marriage/couples/ intimate relationships	1	2	3	4	5	6	7	8	9	10
3) Parenting	1	2	3	4	5	6	7	8	9	10
4) Friends/social life	1	2	3	4	5	6	7	8	9	10
5) Work	1	2	3	4	5	6	7	8	9	10
6) Education/training	1	2	3	4	5	6	7	8	9	10
7) Recreation/fun	1	2	3	4	5	6	7	8	9	10
8) Spirituality/meaning & purpose in life	1	2	3	4	5	6	7	8	9	10
9) Citizenship/ Community Life	1	2	3	4	5	6	7	8	9	10
10) Physical self-care (nutrition, exercise/ movement, rest/sleep)	1	2	3	4	5	6	7	8	9	10

Reflection:
How do you feel about this? Are there any areas that surprised you?

Valued Living Questionnaire
Self-Care Assessment Part 2

In this section, please give a rating of how **consistent** your actions have been with each of your values. Please note that this is **not** asking about your ideal in each area, **nor** what others think of you. Everyone does better in some areas than in others. People also do better at some times than at others. **Please just indicate how you think you have been doing during the past week.** Rate each area (by circling a number) on a scale of 1-10. A "1" means that your actions have been *completely inconsistent with your value*. A "10" means that your actions have been *completely consistent with your value*.

During the past week…

Area:	**not at all consistent with my value**									**completely consistent with my value**
1) Family (other than marriage or parenting)	1	2	3	4	5	6	7	8	9	10
2) Marriage/couples/ intimate relationships	1	2	3	4	5	6	7	8	9	10
3) Parenting	1	2	3	4	5	6	7	8	9	10
4) Friends/social life	1	2	3	4	5	6	7	8	9	10
5) Work	1	2	3	4	5	6	7	8	9	10
6) Education/training	1	2	3	4	5	6	7	8	9	10
7) Recreation/fun	1	2	3	4	5	6	7	8	9	10
8) Spirituality/meaning & purpose in life	1	2	3	4	5	6	7	8	9	10
9) Citizenship/ Community Life	1	2	3	4	5	6	7	8	9	10
10) Physical self-care (nutrition, exercise/ movement, rest/sleep)	1	2	3	4	5	6	7	8	9	10

Total: _____

Add up the total circled numbers for Part 2, where 10 is the minimum and 100 is the maximum. The higher the number the more likely you are to experience happiness in your life.

Behavioural Activation for Depression

Effectiveness of Behavioural Activation Group Therapy

What you need to KNOW about attending

Evidence shows that behavioural activation groups (BAG) are effective at reducing symptoms of depression.

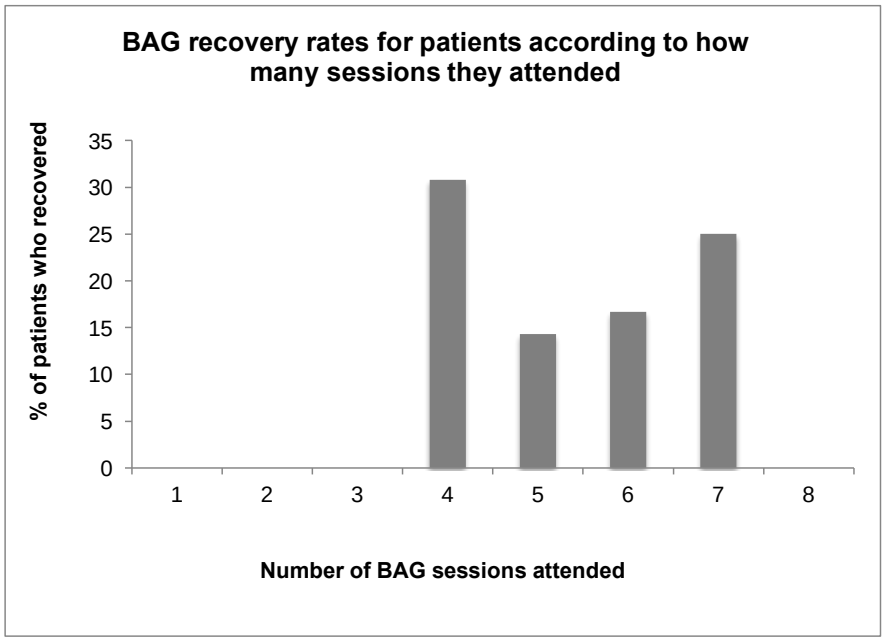
- Research has looked at which situations and circumstances BAG can be effective so you KNOW how you can benefit.
- Research has also looked at what predicts a good outcome following treatment and identified what you can DO which will be more likely to result in a better outcome.

Below are some findings that can help you understand how to get the best out of the group;

1. You need to receive the necessary amount of therapy to experience any significant improvement.

Think of BAG therapy as like a dose of antibiotics – you need to complete the full course of treatment to recover. If you walk away from the group (even if you are feeling better), you are at risk of not getting the full benefit and leaving your depression unaddressed.

The findings show that patients only reach recovery after attending at least 4 sessions of BAG (i.e. the more sessions you attend the more likely you are to see improvement).

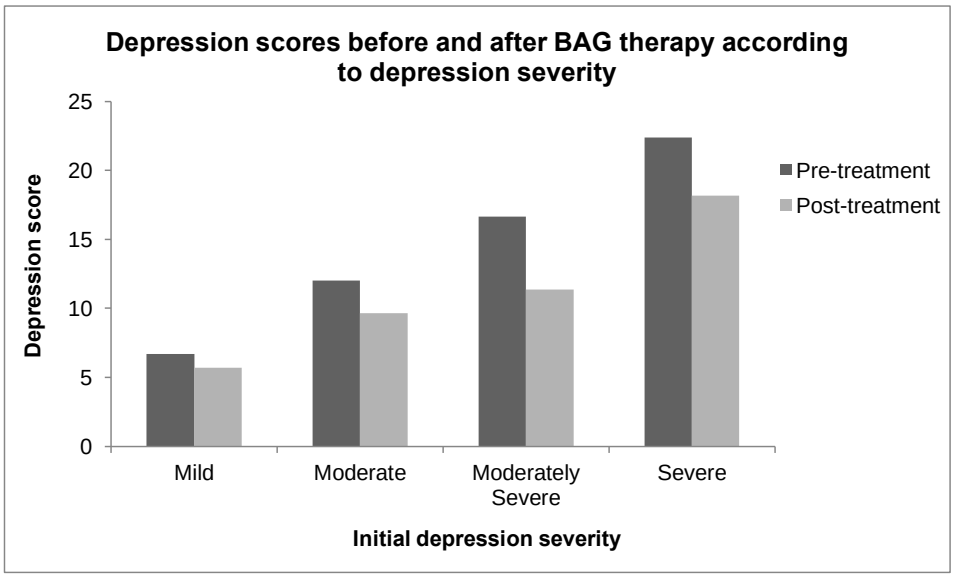


So remember you may not see change immediately, if you still feel depressed stick with it!

Version 2.0: 25th May 2016

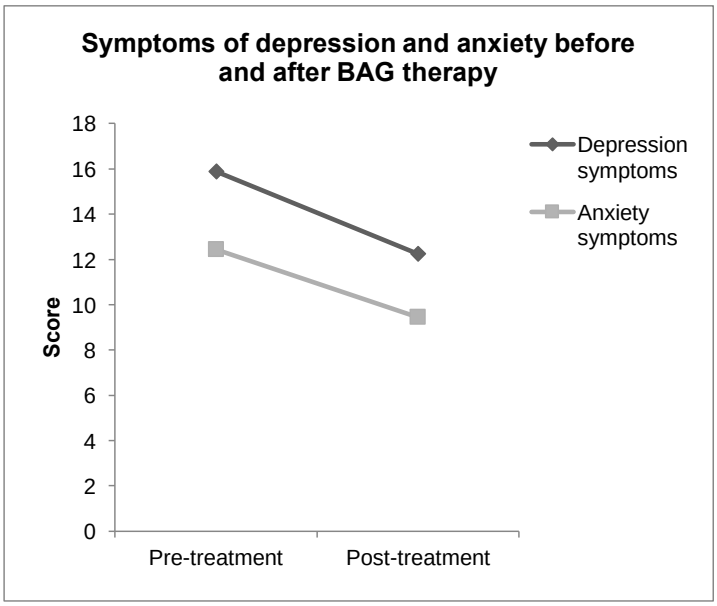
2. BAG is effective at treating all severities of depression

BAG therapy can have a beneficial effect on all severities of depression. The findings showed BAG can reduce depression symptoms regardless of how severe your depression is.



So even if you are sceptical about your depression getting better, stick with BAG!

3. BAG is also effective at reducing co-existing anxiety symptoms



When you are depressed, it is common to also experience feelings of anxiousness and worry as well.

Although BAG is designed to treat low mood, the findings have shown that in addition to reducing depression symptoms, BAG can also reduce co-existing anxiety symptoms.

So you might find you get more benefit from BAG therapy than you expect.

Appendix F: Implementation intentions materials ('Achieving your goals' and 'If-then plans' worksheet) used to augment BAG+ in Chapter 5

Behavioural Activation for Depression

Achieving Your Goals

Why do people struggle to change their behaviour when they are feeling down?

People often set goals to try and change their behaviour, but then find it difficult to put them into action. This is true generally, but really true when we are feeling down in the dumps. This can happen for two reasons, either 1) people have difficulty getting started and initiating the behaviour or 2) people get started and then encounter obstacles or barriers, which push them off track and prevent them maintaining the behaviour.

Common barriers that prevent people achieving their goal when feeling down include;

- Forgetting to do the new behaviour (e.g. make a good plan, but then it slips your mind when you really need it)
- Missing chances to action the new behaviour (e.g. don't notice an opportunity)
- Having interfering second thoughts at crucial moments (e.g. doubting yourself / not backing yourself)
- Getting distracted by tempting alternatives (e.g. I'll watch this, then I'll do it)
- Falling back into unhelpful habits (e.g. putting things off)
- Allowing negative moods to justify not putting a new behaviour into action (e.g. I'll do this when I feel a bit better)

Helping you to achieve your goals

Evidence has shown that specific 'if-then' planning really helps. 'If-then' plans are statements that identify the barriers might get in the way of achieving a goal and a pre-planned response for how to deal with them if they happen. So, simple planning in advance results in people being more likely to follow through and achieve what they want. This planning technique of clearly stating how you are going to achieve a goal is known as forming 'implementation intentions'.

- The **'IF'** part of the plan outlines the most likely barrier that might get in the way of a desired goal and when it might happen.
- The **'THEN'** part of the plan outlines what the planned response to overcome that barrier will be.

E.g. '*If I am feeling anxious about attending the group, then I will accept that feeling, but remind myself that is understandable and turn up anyway.*'

Using this specific planning technique will really help you to put your behavioural activation homework into practice and will help to shift your depression.

How will 'if-then' planning help you to put your plans into action?

Forming plans to act to help reduce your low mood works because;

a) Pre-planning in advance removes distracting choices from everyday decisions we have to make
b) You learn to override what has in the past got in the way
c) Knowing your plan in advance will make it happen more automatically. This removes the possibility of overthinking which can derail us from taking immediate action.

Using 'If-then' planning for each week's homework

Use the 6 steps below to make a specific homework plan;

1. Choose your homework goal for this week
2. Make a specific plan for how you will put your homework goal into action (what, where, when, who with)
3. Thank about the potential barriers that have stopped you in the past or are most likely to get in the way
4. Write an 'If' statement outlining what barrier might stop you acting and when it might happen (e.g. a situation or mood)
5. Write a 'Then' statement with a planned response to deal with that barrier (e.g. thinking, doing, acknowledging or accepting something).
6. Repeat your 'if-then' statement silently to yourself 3 times and then to your homework partner out loud once

E.g. ***Homework plan:*** *On weekday mornings set an alarm and get out of bed at 7.30am*

If-then plan: *If when my alarm goes off I feel too tired and want to stay in bed, then I will remind myself how low it will make me feel later and immediately get up and go make a cup of tea.*

See the 'Making My Plans Really Happen' worksheet for a template for setting your behavioural activation homework each week. If needed, there is space to write multiple goals or alternatively think off two barriers which might stop you achieving your goal.

Making My Plans Really Happen - Worksheet

Stage	Content	Session 1 Examples
1) Homework goal:	.....	Complete the mood diary
Specific Plan: What / When / Where	....	Keep it in my bag and fill it in everyday immediately after I finish an activity
Barrier: Internal or External	If	If I think what is the point in doing this
Response: Thinking, doing, acknowledging or accepting something	Then	Then I will remind myself that this is important for overcoming my depression and will fill it in anyway
2) Homework goal:	.....	Get up in the mornings
Specific Plan: What / When / Where	....	On weekday mornings set an alarm and get up at 7.30am
Barrier: Internal or External	If	If when my alarm goes off I feel too tired
Response: Thinking, doing, acknowledging or accepting something	Then	Then I will remind myself how low it will make me feel later and immediately go make a cup of tea.

Work in pairs to identify your goal and 'if-then' plans for this week's homework. Repeat the 'if-then' statement silently to yourself 3 times and then say it out loud to your partner.

Appendix G: Behavioural activation in groups (BAG) adherence checklist (for session 1) used in Chapter 5

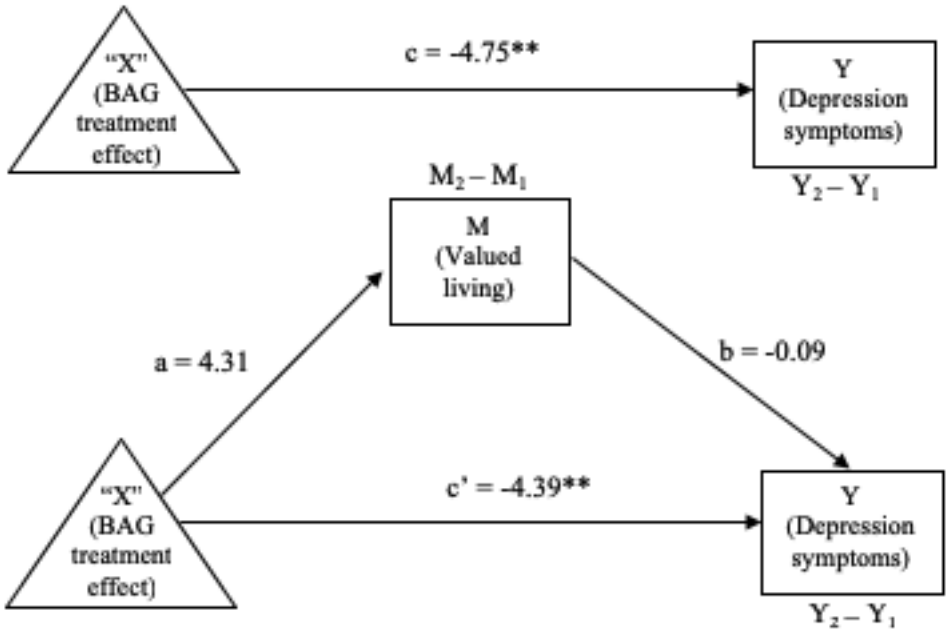
BEHAVIOURAL ACTIVATION GROUP (BAG) THERAPY: TREATMENT ADHERENCE CHECKLIST Rating Form							
Session No: 1	**Reviewer:**				**Date:**		
BAG Domain	**Key Features**	**N/A**	**No evidence**	**Little evidence**	**Some evidence**	**Sufficient evidence**	**Very clear evidence**
General Adherence							
Behavioural Rationale	Simple depression formulation**						
	Focussing on behaviours						
	Overcoming mood dependence						
	Identifying & targeting avoidant behaviours linked to depression						
	Self-monitoring of mood-activity linkages						
	Emphasis on reconnecting or continued connection to positively reinforcing activities (including TRAC)*						
	Troubleshooting barriers to behavioural activation*						
Between-Session Work	Review homework from previous session*						
	Identification & development of meaningful short-term goals connected to session content						
	Behavioural activation planning as homework based upon session content						
	Use of implementation intentions						
*Feature not expected to be present in session 1 **May not feature in later sessions		(If feature is not expected to be present in the session, rate as N/A)					
Session Specific Adherence							
Session 1	Introduction to depression & activity-mood monitoring						
Session 2	Explore role of values & engaging in valued activity (VLQ)						
Session 3	Use of TRAP & TRAC approach for avoidance						
Session 4	Use of problem-solving approach to take action						
Session 5	Function of thinking & monitoring rumination (2-minute rule)						
Session 6	Use of RCA, mindfulness & self-soothing to deal with rumination						
Session 7	Dealing with physical symptoms & setting short-term goals						
Session 8	Building relationships in context of barriers to change (ACTION)						
Overall would you rate the session as behavioural activation?		No			Yes		

Appendix H: Exploratory mediation analyses from Chapter 6

Valued living as a mediator of depression outcome during BAG: Exploratory Analyses

To test the hypothesis (4) that change in *valued living* would mediate change in depression outcomes during BAG, a repeated-measures mediation model was employed presented in Figure H.1 (Montoya & Hayes, 2017). Separate analyses were performed for total depression, affective and somatic symptom outcomes. The independent variable (X) was the BAG treatment effect, pre-post change in *valued living* measured by the VLQ Valued Living score was the mediating variable (M), and pre-post change in depression symptoms measured by the PHQ-9 was the outcome variable (Y). As the data were repeated-measures, a value for 'X' was not represented in the data. 'X' was therefore represented by the change scores in the mediator (*valued living*) and the outcome variable (depression symptoms).

Figure H.1 displays the mediation model for *valued living* change on overall depression symptom change. There was a significant negative direct effect of BAG treatment on depression symptoms (c' path). BAG significantly reduced depression symptoms by 4.39 units on the PHQ-9. Although BAG treatment increased patients *valued living* by 4.31 units on the VLQ (a path) and the direct effect of *valued living* change on depression change was negative (b path), neither effect was significant. The resulting indirect effect (a*b path) showed the increase in *valued living* in turn decreased depression symptoms by 0.36 units on the PHQ-9. However, the bootstrapped CIs crossed zero. Therefore, contrary to the hypothesis, change in depression was not significantly mediated through change in *valued living*.



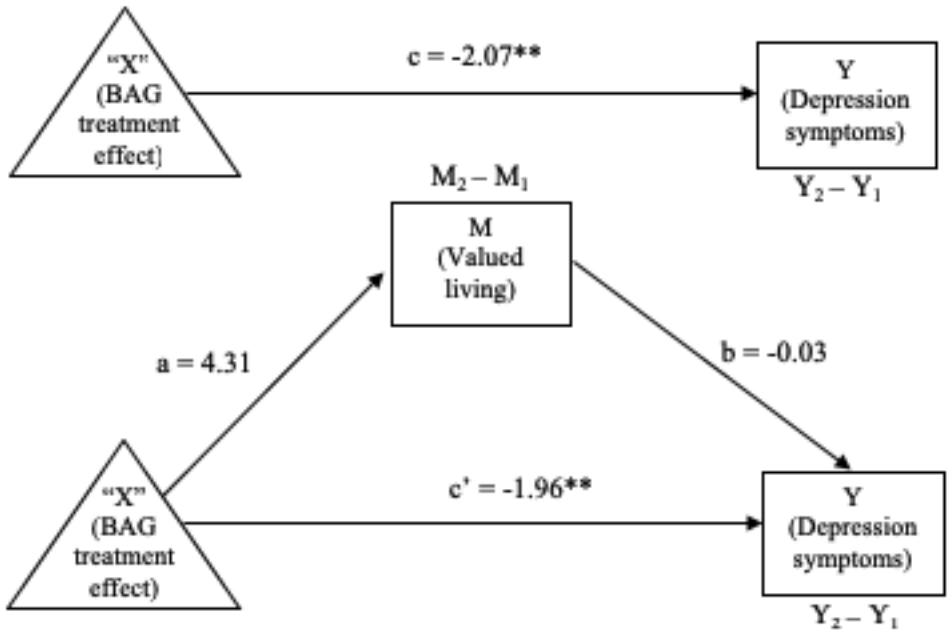
a*b = -0.36 (bootstrapped 95% CI -0.05 to 2.40)

Path a = direct effect of BAG treatment on change in valued living
Path b = direct effect of change in valued living on change in depression symptoms
Path c = total effect of BAG treatment on change in depression symptoms
Path c' = direct effect of BAG treatment on change in depression symptoms
Path a*b = indirect effect of BAG treatment on change in depression symptoms mediated through change in valued living

Figure H.1. Repeated-measures Mediation Model for the Effect of Change in Valued Living on Depression Symptoms During BAG

Figures H.2 and H.3 display the mediation models for *valued living* change on affective and somatic depression symptom change. Again, there was a significant negative direct effect of BAG treatment on depression symptoms (c' path) in both models. That is, BAG significantly reduced affective and somatic depression symptoms by 1.96 and 2.42 units on the PHQ-9 respectively. The effect of BAG on *valued living* remained constant (a paths; increasing by 4.31 units) and both direct effects of *valued living* change on affective and somatic depression change were negative (b paths), but neither effect was significant. The resulting indirect effects (a*b path) showed the increase in *valued living* in turn decreased affective and somatic depression symptoms by 0.11 and 0.26 units on the PHQ-9 respectively. However, the bootstrapped CIs

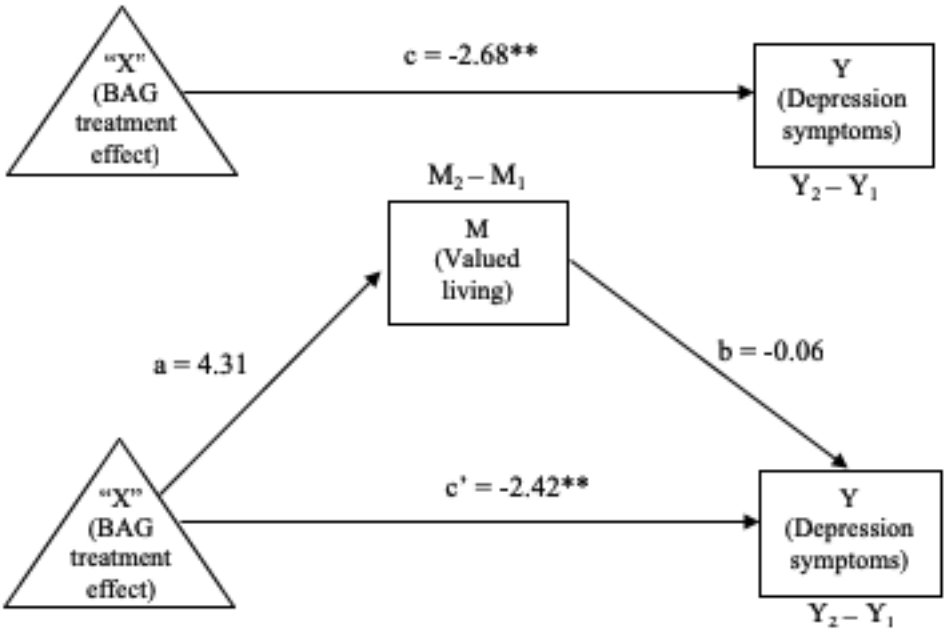
crossed zero. Therefore, contrary to the hypothesis, change in affective and somatic depression symptoms were not significantly mediated through change in *valued living*.



a*b = -0.11 (bootstrapped 95% CI -0.08 to 1.09)

Path a = direct effect of BAG treatment on change in valued living
Path b = direct effect of change in valued living on change in depression symptoms
Path c = total effect of BAG treatment on change in depression symptoms
Path c' = direct effect of BAG treatment on change in depression symptoms
Path a*b = indirect effect of BAG treatment on change in depression symptoms mediated through change in valued living

Figure H.2. Repeated-measures Mediation Model for the Effect of Change in Valued Living on Affective Symptom Clusters During BAG



a*b = -0.26 (bootstrapped 95% CI -0.04 to 1.44)

Path a = direct effect of BAG treatment on change in valued living
Path b = direct effect of change in valued living on change in depression symptoms
Path c = total effect of BAG treatment on change in depression symptoms
Path c' = direct effect of BAG treatment on change in depression symptoms
Path a*b = indirect effect of BAG treatment on change in depression symptoms mediated through change in valued living

Figure H.3. Repeated-measures Mediation Model for the Effect of Change in Valued Living on Somatic Symptom Clusters During BAG