Contents lists available at ScienceDirect

Behaviour Research and Therapy

How well do Cognitive Behavioural Therapy and Behavioural Activation for depression repair anhedonia? A secondary analysis of the COBRA randomized controlled trial

Batool Alsayednasser [a], Emily Widnall [b], Heather O'Mahen [a], Kim Wright [a], Fiona Warren [c], Asha Ladwa [a], Gabriela Kattan Khazanov [d], Sarah Byford [e], Willem Kuyken [f], Ed Watkins [a], David Ekers [g,k], Nigel Reed [a], Emily Fletcher [c], Dean McMillan [h], Paul Farrand [i], David Richards [c,j], Barnaby D. Dunn [a,*]

[a] Mood Disorders Centre, University of Exeter, UK
[b] Bristol Medical School, University of Bristol, UK
[c] College of Medicine and Health, University of Exeter, UK
[d] Corporal Michael J. Crescenz VA Medical Center, USA
[e] Health Service & Population Research Department, Kings College London, UK
[f] Department of Psychiatry, University of Oxford, UK
[g] Department of Health Science, University of York, UK
[h] Department of Health Sciences and Hull York Medical School, University of York, UK
[i] CEDAR, University of Exeter, UK
[j] Department of Health and Caring Sciences, Western Norway University of Applied Sciences, Norway
[k] Tees Esk and Wear Valleys NHS Foundation Trust, UK

ARTICLE INFO

ABSTRACT

A secondary analysis of the COBRA randomized controlled trial was conducted to examine how well Cognitive Behavioural Therapy (CBT) and Behavioural Activation (BA) repair anhedonia. Patients with current major depressive disorder ($N = 440$) were randomized to receive BA or CBT, and anhedonia and depression outcomes were measured after acute treatment (six months) and at two further follow up intervals (12 and 18 months). Anhedonia was assessed using the Snaith Hamilton Pleasure Scale (SHAPS; a measure of consummatory pleasure). Both CBT and BA led to significant improvements in anhedonia during acute treatment, with no significant difference between treatments. Participants remained above healthy population averages of anhedonia at six months, and there was no further significant improvement in anhedonia at 12-month or 18-month follow up. Greater baseline anhedonia severity predicted reduced repair of depression symptoms and fewer depression-free days across the follow-up period in both the BA and CBT arms. The extent of anhedonia repair was less marked than the extent of depression repair across both treatment arms. These findings demonstrate that CBT and BA are similarly and only partially effective in treating anhedonia. Therefore, both therapies should be further refined or novel treatments should be developed in order better to treat anhedonia.

Depression is a distressing, chronically recurrent and relapsing condition that is highly prevalent and a major contributor to disability (Murray & Lopez, 1997; World Health Organization, 2017). Current psychotherapies for depression help many, but are not optimally effective. For example, meta-analyses suggests that response (showing at least a 50% reduction in symptoms during treatment) and remission (scoring below clinical cut-offs on depression scales) criteria are met by less than half of participants who undergo depression psychotherapies, with no clear differences across therapy approaches (Cuijpers, Karyotaki, et al., 2014; Cuijpers, Karyotakie, Ciharova, Miguel & Noma, 2021). Of those who no longer meet diagnostic criteria for depression after psychotherapy, more than 50% will relapse within two years (Steinert, Hofmann, Kruse, & Leichsenring, 2014; Vittengl, Clark, Dunn, & Jarrett, 2007). There is a pressing need to refine existing interventions

or develop new treatments to improve rates of treatment response and prevent subsequent relapse.

The two cardinal affective symptoms of depression are depressed mood (elevated sadness and other negative emotions) and anhedonia (a loss of interest or pleasure in activities that were previously recognized as rewarding). This is reflected in the fact that at least one of these two symptoms needs to be present to meet criteria for a diagnosis of current major depressive disorder (see DSM-V, American Psychiatric Association, 2013). These two symptoms reflect an imbalance in two somewhat separate neurobiological dimensions: the negative valence system (NVS) and the positive valence system (PVS; Medeiros, Rush, Jha et al., 2020; Paulus et al., 2017; Watson, Wiese, Vaidya, & Tellegen, 1999). The NVS regulates withdrawal from punishing stimuli and leads to negative affect, including emotions such as anger, resentment, sadness, anxiety, and fear. The PVS regulates approach to rewarding stimuli and generates positive affect, including emotions like happiness, excitement, elation, and enthusiasm. Anhedonia can be conceptualized as impaired activation of the PVS, leading to reduced positive affective reactivity when anticipating, experiencing or remembering rewarding stimuli. In addition to these hedonic ('liking') components, anhedonia also is characterized by reduced motivation and effort to work for rewards ('wanting') and reduced capacity to learn from rewarding outcomes ('learning'; Dunn, 2019). Anhedonia forms one part of broader deficits in wellbeing in depression (i.e., reduced capacity to experience pleasure, meaning and social connection; cf., Keyes, 2002).

It has long been acknowledged that depressed mood is central to the onset and maintenance of depression, but it is now increasingly apparent that anhedonia is also pivotal. Some degree of anhedonia is present in a nearly all depressed cases and severe anhedonia occurs in approximately a third of cases (Pelizza & Ferrari, 2009). Anhedonia predicts greater risk of depression onset (Bennik, Nederhof, Ormel, & Oldehinkel, 2014; Pine, Cohen, Cohen, & Brook, 1999; Wilcox & Anthony, 2004) and a greater chance of a chronic, relapsing course (Spijker, Bijl, de Graaf, & Nolen, 2001). While clinicians tend to focus on depressed mood as key to recovery from depression, patients are clear that repair of anhedonia and building of positive mood is as (and possibly more) important to their recovery (Zimmerman et al., 2006; Demyttenaere et al., 2015). Therefore, to treat depression effectively it is likely necessary to reduce both depressed mood and anhedonia.

Existing treatments, however, focus explicitly on reducing depressed mood and to some extent neglect repairing anhedonia (Dunn, 2012, 2019; Dunn & Roberts, 2016). For example, while early stages of Cognitive Behavioral Therapy (CBT) involve encouraging clients to engage with pleasurable activities, the main focus of treatment is on changing patterns of negative thinking that drive negative emotions and maintain a negative view of the self, world and future (Beck, Rush, Shaw, & Emery, 1979; Moore & Garland, 2004). It therefore seems plausible that CBT and related psychotherapies will be relatively ineffective at repairing anhedonia.

Consistent with this prediction, a recent secondary analysis of two randomized controlled trials comparing CBT to anti-depressant medication found that both treatments normalized negative affect (related to depressed mood) to general population average levels, whereas positive affect (related to low anhedonia) remained at least one standard deviation below the general population average at the end of treatment. Change in positive affect and negative affect independently predicted change in depression symptom severity in a linear regression analysis, showing that they have dissociable relationships with treatment outcome (Dunn, German, Khazanov et al., 2020).

At face value, this suggests that current treatments do not adequately repair anhedonia and that new treatments need to be developed to target it better. However, there are a number of limitations with the Dunn et al. (2020) secondary analysis that suggest that this conclusion is premature. First, the analyses were post-hoc and exploratory, so require replication. Second, the analyses focused on changes in positive affect. While this is associated with anhedonia, it is not entirely the same construct. There is

a subtle but important difference between free-floating 'background' positive affect and changes in positive affect in response to rewarding stimuli (positive affect reactivity; the core hedonic component of anhedonia). It is important to see if findings replicate using validated, widely-used measures of positive affect reactivity. For example, a recent review (Rizvi, Pizzagalli, Sproule, & Kennedy, 2016) concluded that the Snaith-Hamilton Pleasure Scale (SHAPS; Snaith, Hamilton, Morley et al., 1995) – a measure of consummatory pleasure that individuals would experience when engaging with rewarding stimuli – arguably remains the closet to a current "gold standard" tool for assessing anhedonia. This was on the basis that the SHAPS assesses several domains of (consummatory) reward responsiveness, has been psychometrically validated in depressed populations, and is not culturally biased (Rizvi et al., 2016). Third, Dunn et al. (2020) looked only at outcomes immediately post treatment, and long-term anhedonia outcomes have not been examined. On the one hand, it is conceivable that anhedonia simply takes longer to repair and will normalize in due course once depressed mood has recovered and individuals are regularly engaging with potentially rewarding and valued activities in their environment (a 'sleeper' effect; Fluckiger & Del Re, 2017). On the other hand, it may be that any residual deficits in anhedonia persist or even worsen in the period of time after acute therapy ends. To differentiate between these possibilities, follow-up data in the time after treatment are required.

Perhaps the key limitation is that the Dunn et al. (2020) analyses focused on trials of CBT and it is conceivable that other types of evidence-based psychotherapy for depression may be more effective at repairing anhedonia. For example, Behavioural Activation (BA) represents one component of CBT, but is also a standalone therapy in its own right that focuses solely on reactivating clients to engage with previously or potentially rewarding/valued activities (Dimidjian, Barrera, Martell, Munoz, & Lewinsohn, 2011). While the explicit reward focus of BA at face value suggests it may be better than CBT at repairing anhedonia, this remains an unresolved empirical question. There is general evidence across populations, study designs, and therapy variants (Mazzucchelli, Kane, & Rees, 2010) that BA can to some extent improve the broader construct of wellbeing. However, no studies to date have evaluated how well BA repairs anhedonia in diagnosed depressed populations using randomized controlled trial designs.

Another clinically important issue to explore is whether anhedonia severity at baseline predicts response to psychotherapy for depression. This is important for the purposes of selecting which psychotherapy individuals are allocated to and predicting their likely prognosis to inform case management. We are aware of only a handful of studies that have examined this issue. McMakin, Olinio, Porta et al. (2012) carried out a secondary analysis of a trial comparing a medication switch alone to a medication switch combined with CBT in treatment-resistant depressed youth. Greater anhedonia symptoms at baseline predicted a longer time to remission and a reduced number of depression-free days in both arms, and these associations held even when covarying for other symptom dimensions of depression at baseline. Khazanov et al. (2020) conducted a secondary analysis of a trial comparing medication alone to medication combined with CBT for the treatment of chronic and severe depression. This analysis found that reduced levels of positive affect at baseline predicted a longer time to remission and sustained recovery across treatment arms. Furthermore, baseline levels of positive affect also predicted differential response to treatment. In particular, those with lower levels of positive affect benefitted more from combined drug and cognitive therapy relative to drug therapy alone. In contrast, those with higher levels of positive affect responded equally well to each treatment. These results held whether or not baseline depression severity was adjusted for in the models.

As with the Dunn et al. (2020) analyses, the McMakin et al. (2012) and Khazanov et al. (2020) moderation analyses have methodological limitations that mean a number of questions about the extent to which anhedonia predicts treatment outcomes following psychotherapy for depression remain unresolved. First, neither trial included in the

analyses had a psychotherapy alone condition. It is unclear to what extent anhedonia moderated drug versus therapy response in the combined condition. Second, both analyses focused solely on outcomes at the end of acute treatment and it is unknown whether anhedonia moderates longer-term outcomes in the months after treatment has finished. Finally, both of these moderation analyses focus solely on CBT as the psychotherapy and it is uncertain if a similar picture would emerge for other therapy types like BA.

Also of interest is the degree to which anhedonia is satisfactorily repaired relative to symptoms of depression. Previous research examining outcomes of routine delivery of CBT in UK Improving Access to Psychological Therapy (IAPT) Services has shown that while CBT does improve wellbeing, the extent of wellbeing repair was less relative to alleviation of depression and anxiety symptoms (Widnall, Price, Trompetter, & Dunn, 2020). On the basis of these findings, it seems likely that CBT will be less effective at repairing anhedonia relative to symptoms of depression, but this possibility has yet to be formally examined. Moreover, no work to date has examined the capacity of BA to repair anhedonia relative to depression.

To explore this series of unresolved issues further, the present study conducted a secondary analysis of the COBRA non-inferiority randomized controlled trial that compared the clinical- and cost-effectiveness of CBT versus BA in the treatment of major depressive disorder (Richards, Ekers, McMillan et al., 2016). This trial found that CBT and BA were equivalently clinically effective at repairing depression immediately after acute treatment and over longer-term follow up. However, BA was more cost-effective than CBT because it was delivered by treatment providers that were lower cost and required less training (Richards et al., 2016). COBRA included a measure of anhedonia (the SHAPS) as a secondary outcome measure. The present analyses interrogate the SHAPS data to: 1) evaluate the efficacy of CBT versus BA to treat anhedonia symptoms over the short and long term; 2) investigate whether anhedonia severity at baseline moderated depression treatment outcomes in each arm; and 3) explore whether repair of anhedonia was of a smaller magnitude compared to repair of depression in each arm.

1. Method

1.1. Participants and trial design

Participants for the COBRA trial were identified by searching electronic case records of primary care general practices in three UK sites (Devon, Durham and Leeds), eventually resulting in 440 individuals being recruited. Participants were adults, aged 18 and older, who met criteria for a current major depressive disorder episode according to a standard clinical interview (Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition [SCID]; Spitzer, Williams, Gibbon, & First, 1992). Participants were excluded if they were currently receiving other psychological therapy, were dependent on alcohol or drugs, were suicidal or had made a suicide attempt in the previous two months, had cognitive impairment, had a diagnosis of bipolar disorder or a psychotic disorder, or were experiencing psychotic symptoms (see Rhodes et al., 2014 for full trial protocol paper).

Participants were randomly allocated in a 1:1 ratio to either BA ($N =$ 221) or CBT ($N = 219$) after baseline assessment, stratified based on baseline depression symptom severity on the Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer, & Williams, 2001; scoring <19 versus ≥19), antidepressant use (currently using anti-depressants or not) and recruitment site (Devon, Durham or Leeds). CBT consisted of up to 20 individual (approximately weekly) 1-h sessions delivered by one of 12 high-intensity psychological therapists (accredited CBT therapists), following a Beckian CBT treatment protocol based on Beck's original writings (Beck et al., 1979) and updated to include elements targeting complex and treatment resistant depression (Moore & Garland, 2004). BA consisted of up to 20 individual (approximately weekly) 1-h sessions

delivered by one of 10 low-intensity psychological therapists (junior mental health workers) following a BA treatment protocol based on Martell (Martell, Dimidjian, Herman-Dunn, Lewinsohn, & DeRubeis, 2010). The primary emphasis in the protocol was on helping individuals to reconnect to personally meaningfully activities. The protocol was updated to include material on addressing rumination, anxiety, and communication within the principles of BA.

Therapists were trained and supervised by experienced therapists in the appropriate modality. A random selection of tapes in each arm were coded for competence by external expert therapists to ensure treatment was delivered to protocol. Therapists in both arms met acceptable competency standards. Participants attended on average 12.5 CBT sessions and 11.5 BA sessions, with 72% of clients in the CBT arm and 67% of clients in the BA arm receiving a minimum adequate dose of therapy (eight or more sessions attended). The researchers conducting follow-up assessments were blinded to group allocation.

1.2. Measures

Participants completed a battery of clinical and cost-effectiveness measures at baseline, and six months, 12 months, and 18 months post-randomization. The focus of the present secondary analyses is on anhedonia outcomes (and how this relates to depression outcomes), so only these measures are described further here.

To measure anhedonia severity, participants completed the SHAPS (Snaith et al., 1995). This is a 14-item self-report questionnaire asking participants to judge retrospectively if they would have been able to enjoy engaging with a range of potentially rewarding activities over the past few days, with items selected to cover four life domains (interests/pastimes, social interaction, sensory experience, and food/drink). Participants rate to what extent they agree with a series of statements describing their capacity to experience pleasure in each domain (for example, enjoy being with family or close friends, enjoy a favorite meal), on a 4-point Likert scale ranging from definitely agree to definitely disagree. The SHAPS can be conceptualized as a measure of the hedonic ('liking') component of anhedonia, focusing specifically on the consummatory (and not the anticipatory or recall) phase.

In the original scale development papers (Snaith et al., 1995), the two "agree" responses were coded as "0" and the two "disagree" responses were coded as "1" for each item and then the items were summed to indicate a total scale score (ranging from 0 to 14, with higher scores indicating greater anhedonia). More recent papers have used a continuous scoring method to increase sensitivity to change (1 for definitely agree, 2 for agree, 3 for disagree, and 4 for definitely disagree; Franken, Rassin, & Muris, 2007), producing scores ranging from 14 (not at all anhedonic) to 56 (severely anhedonic).The present study adopts this continuous scoring approach.

The internal reliability of the continuously scored SHAPS has been found to be adequate in both non-clinical ($\alpha = 0.91$) and clinical ($\alpha = 0.94$) samples (Franken et al., 2007). The internal reliability for the continuously scored SHAPS was also adequate in the current sample ($\alpha = 0.82$). Normative data on the continuously scored SHAPS are available in adult populations from a recent meta-analysis (Trøstheim, Eikemo, Meir et al., 2020), with a mean score of 20.2 (SD of 2.1) in healthy participants (41 samples with 3405 participants). There was negligible impact of age and gender on these scores (Trøstheim et al., 2020). There is no currently accepted clinical cut-off on the continuous scoring of the scale.[1] A proxy estimate was derived for the current study, specifying that caseness would be indicated by individuals scoring more than 1.96

---

[1] We chose not to report the original categorical coding of the SHAPS, as dichotomising the response options reduces scale sensitivity to change. While a recovery criterion (scoring <2) has been proposed for the categorical coding of the SHAPS (<2; Snaith et al., 1995) this was largely arbitrary and based on visual analysis of the distribution of scores in a small initial validation study.

SDs above the general population average reported in the Trøstheim et al. (2020) meta-analysis. Using this method, a SHAPS score of 25 or greater indicates caseness.

To measure depression severity, participants completed the Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001). This is a nine-item self-report measure indexing depression symptom severity over the past two weeks. It measures the frequency of nine core DSM-IV criteria for depression (for example, low mood, sleep difficulties), experiences of which are each rated on a scale ranging from "0" (not at all) to "3" (nearly every day). Scores for each item are summed, generating a score from 0 (not at all depressed) to 27 (severely depressed). The internal reliability of the PHQ-9 has been found to be excellent (Cronbach's α of 0.89; Kroenke et al., 2001) and was comparable in the current trial (α = 0.77). Comparison normative data on the PHQ-9 are available, with a mean score of 2.9 (SD of 3.5) found in a German general population sample of 5018 adults (Kocalevent, Hinz, & Brähler, 2013). A score of 10 or greater is used to indicate clinically significant depression and a score of less than 10 is used to indicate remission (Kroenke et al., 2001).

As an additional way to index long-term depression outcomes to use in moderation analyses, the number of depression-free days individuals experienced over the follow-up period were computed (cf., Vannoy, Arean, & Unützer, 2010). To index depression-free days, at each follow-up assessment participants were asked to estimate the number of days over that six-month period they had been free of depression. The values for six-month, 12-month, and 18-month assessments were added together to index cumulative depression-free days.

1.3. Analysis plan

Analyses were conducted on an intent-to-treat basis, using all available data, and were run deploying the Statistical Package for Social Science Version 26 (SPSS; IBM Corp, 2019). Alpha was set at 0.05, two tailed statistical tests were used throughout, and no corrections were made for multiple comparisons.

To examine if rates of missing SHAPS and PHQ-9 data varied as a function of treatment arm, a series of chi-squared tests were run. To examine if baseline levels of anhedonia varied between arms, an analysis of covariance (ANCOVA) was conducted on continuous SHAPS scores, with group as a between-subjects variable and the trial stratification variables entered as covariates.

1.3.1. Impact of CBT versus BA on anhedonia

As the data were in a multi-level longitudinal structure (time points within individuals), an analysis approach that could take into account this nested structure was selected to examine if the extent of anhedonia repair differed between treatment arms. Generalized Estimating Equations (GEEs) were deployed[2], fitting a normal distribution model with parameter estimation based on the hybrid method (using 100 maximum iterations and maximum step-halving of 5). As these analyses are robust to missing data, no imputation procedure was used to simulate missing values. An unstructured correlation matrix was specified and beta co-efficients are reported. The dependent variable was SHAPS severity. Key predictor variables were group (CBT or BA, coded as 0 and 1 respectively), time (baseline, six months, twelve months, and eighteen months, treated as a continuous variable and coded as 0,1,2 and 3 respectively), and the interaction between group and time. The trial stratification variables (site, medium vs severe baseline depression severity, and antidepressant medication status; all coded as categorical variables) were

entered as covariates. The equation for this initial model is shown in the supplementary materials.

1.3.2. Does baseline anhedonia moderate depression symptom repair?

To determine the extent to which depression and anhedonia were dissociable constructs, a correlation was run between baseline SHAPS anhedonia and PHQ-9 depression. To examine the possibility of a main and an interactive moderating effect (cf., Kraemer, Wilson, Fairburn, & Agras, 2002) of baseline anhedonia severity on depression repair, two GEE analyses were run. GEE analyses again fitted a normal distribution model, with parameter estimation via the hybrid method. As the moderation models would not resolve when using an unstructured correlation matrix, instead a first order auto-regressive structure (AR1) was specified. PHQ-9 severity was the dependent variable and the trial stratification variables (site, moderate versus severe depression severity at intake, and medication status) were entered as categorical covariates in both cases. In the first analysis (focusing on a main moderating effect), the key predictor variables were time (intake, six months, twelve months and eighteen months), baseline anhedonia severity and their interaction. The second analysis (focusing on an interactive moderating effect) additionally included group and its two- and three-way interactions with the other predictors.

As an additional way to examine the moderation hypothesis, whether intake anhedonia predicted the number of depression-free days participants experienced over the entire trial follow-up period was examined using a linear regression approach. The dependent variable was depression-free days. At step one of the model, baseline SHAPS severity, baseline PHQ-9 severity, and the trial stratification variables were entered as predictors. At step two of the model, the group term and its interaction with baseline SHAPS severity were additionally entered as predictors.

1.3.3. Comparison of extent of depression versus anhedonia repair

As the PHQ-9 and SHAPS have different scale ranges (0–27 versus 14-56), with different points indicating recovery on these scales, it is not meaningful to compare continuous PHQ-9 and SHAPS scores directly in a single analysis. Instead, we computed a range of binary variables that summarise these measures on a common metric. The proportion of clients meeting caseness, response, and reliable and clinically significant improvement was computed for the PHQ-9 (scores >9) and the SHAPS (>24). Response was defined as a 50% reduction in symptoms relative to baseline assessment (for the SHAPS only, first subtracting 14 from each entry point to 'zero' the scale). Reliable and clinically significant improvement was computed following the methods proposed by Jacobson and Truax (1991). Criterion b was used as the cut off for clinically significant improvement and an improvement of more than 1.96 times the standard error of measurement on the scale was used as the cut off for reliable improvement. Baseline internal consistency (Cronbach's α; collapsed across groups) on each scale in the current sample was entered as the estimate of scale reliability. Baseline scores on the PHQ-9 and SHAPS (collapsed across groups) were used as the clinical group reference values. Healthy comparison values for the SHAPS were derived from the meta-analysis of Trøstheim et al. (2020) and for the PHQ-9 from the general population values reported by Kocalevent et al. (2013).

These categorical dependent variables were analysed in a series of GEEs, each fitting a binary (logit) model. As before, the hybrid method was used to estimate parameters and an unstructured correlation matrix was specified. Exponentiated beta-coefficients (odds ratios) are reported. The first set of analyses excluded the group term to simplify interpretation. Key predictors were measure, time and the interaction between measure and time. Measure was coded as 0 for SHAPS and 1 for PHQ-9 in all analyses. Time had four levels in the caseness analyses (baseline, six months, twelve months, and eighteen months; coded continuously as 0,1,2 and 3 respectively) and three levels in the response analyses (six months, twelve months, and eighteen months; coded

---

[2] GEEs are reported rather than hierarchical linear modelling, as they can include both continuous and binary dependent variables, allow for non-normal distributions, and can more flexibly handle different covariance structures in the data. Moreover, when running linear models on the continuous outcomes, these models would not reliably converge and/or produced a non-positive Hessian matrix, meaning that the estimates were likely unreliable.

continuously as 1,2 and 3 respectively). In the second set of analyses, the group term (coded 0 for CBT and 1 for BA) and its interactions with measure and time were additionally included to explore if the pattern of response observed varied as a function of treatment condition. In all cases, the trial stratification variables were included as categorical covariates.

As an additional way to compare the magnitude of depression versus anhedonia repair that allowed us to make use of the data in continuous rather than categorical form, paired sample t-tests were used to examine change from baseline at six month, twelve month and eighteen month assessment on each scale, focusing on Hedges g mean (95% confidence interval) as a measure of effect size (cf., Widnall et al., 2020). These analyses were run on each group separately and when collapsing across groups. We examined whether the effect size confidence intervals overlapped for the two measures.[3]

1.4. Ethical arrangements

The original trial gained ethical approval from an NHS review board. Participants gave written, informed consent before taking part in the trial. Additional ethical approval was gained to conduct this secondary analysis (West of Scotland Research Ethics Committee, reference 18/WS/0188). As the data were shared in anonymised form and participants had already been informed secondary analyses might be conducted on the trial data after its completion, individual participant consent for this secondary analysis was not sought.

2. Results

2.1. Participant characteristics

Table 1 reports the clinical and demographic characteristics of participants in each arm. The sample were on average middle-aged, predominantly female, and of White British ethnic origin. Participants were a mix of moderately and severely depressed, with a majority of partic-

Table 1
Demographic and clinical characteristic of the sample.

	CBT (n = 219)	BA (n = 221)
Age	43.03 (14.05)	43.88 (14.09)
Female gender	148 (68%)	142 (64%)
White British ethnicity	197 (90%)	204 (92%)
PHQ-9 depression	17.43 (4.82)	17.71 (4.79)
GAD-7 anxiety	12.63 (5.12)	12.69 (5.09)
SHAPS anhedonia	31.92 (5.33)	32.18 (6.34)
Severely depressed (PHQ-9>19)	101 (46%)	103 (46%)
Taking antidepressants	175 (79%)	172 (78%)

Note: Data are mean (SD) values for continuous variables and number (%) for count variables.

---

[3] An alternative approach could have been to express scores in Z score values (relative to general population normative data) as used in Dunn et al. (2020) to compare positive affect and negative affect repair. This method has the advantage of placing different measures on the same scale. A score of zero using this method would align to the general population average, whereas a score of plus/minus one would indicate that score is one standard deviation above/below the general population average. However, this method is only appropriate where comparison data scores are approximately normally distributed, have no clear floor or celling effects, and where normative data on all scales are ideally taken from the same sample or at least a similar size sample (given standard deviation estimates are partly a function of sample size). Normative data for all three of the scales here are positively skewed, suffer from floor effects, and normative data come from very different samples of different sizes, so this method was not suitable for the current data.

ipants having comorbid anxiety symptoms and taking anti-depressant medication.

2.2. Data completeness

Table 2 reports the proportion of participants with outcome data available at each time point for each of the measures. Rates of data completeness were acceptable for the SHAPS (ranging from a maximum of 98% at baseline to a minimum of 66% at twelve months). There were no significant differences in levels of missing data between treatment arms at baseline, six months and eighteen months, χ^2s < 1. There was a non-significant trend for greater rates of missing data in the BA relative to the CBT arm at twelve months, χ^2 = 3.68, P = .06. Rates of data completeness were slightly higher for the PHQ-9 relative to the SHAPS (ranging from a maximum of 100% at baseline to a minimum of 79% at twelve months). Again, there was no significant difference in rates of PHQ-9 missing data between each arm at baseline, six month and eighteen month assessment, χ^2s < 2.66, Ps > .10. However, there were more missing PHQ-9 data in the BA arm relative to the CBT arm at twelve month assessment, χ^2 = 3.90, P = .048.

2.3. Impact of CBT versus BA on anhedonia

Fig. 1 plots SHAPS scores at each time point for the BA and CBT arms. Inspection of the baseline data revealed that SHAPS levels in both groups were elevated compared to healthy comparison samples and comparable to previous clinically depressed samples (Franken et al., 2007). ANCOVA found no significant difference in baseline SHAPS levels between the BA and CBT groups at baseline, F (1,417) = 0.022, P = .882, η_p^2 = 0.006.

GEE analysis found a significant main effect of time, B = −1.63 (SE = 0.19; 95% CI = −2.00 to 1.25), Wald χ^2 = 72.96, P < .001, no significant main effect of group, B = .89 (SE = 0.56; 95% CI = −0.22 to 2.00), Wald χ^2 = 2.49, P = .11, and no significant time by group interaction, B = −0.08 (SE = 0.29; 95% CI = -0.66 to 0.49), Wald χ^2 = 0.08, P = .78. To decompose the main effect of time, a series of pairwise (paired sample *t*-test) comparisons between each consecutive set of time points was run having first collapsed across group. SHAPS scores significantly reduced (in the direction of clinical improvement) from baseline to six month assessment (mean∆ = -5.70, SD = 6.90), t = 14.69, P < .001. There was no further significant reduction from six months to twelve months (mean∆ = −0.69, SD = 7.12), t = 1.61, P = .11, nor from twelve months to eighteen months (mean∆ = 0.06, SD = 6.49), t < 1.

Average SHAPS level at each of the six-, twelve- and eighteen- month assessments in each arm were still above general population mean levels and fell above the cut-off to indicate clinical symptoms (more than 1.96 SDs above the general population average; SHAPS scores>25). Collapsing across groups, the percentage of participants scoring in the clinical range was 58% at six months, 49% at twelve months, and 56% at eigtheen months (compared to 91% at baseline).

To help interpret the likely clinically meaningfulness of these findings, changes in SHAPS across time and difference in SHAPS between conditions at each time-point were expressed as a function of the intake

Table 2
Percentage of participants with complete data on each outcome measure broken down by assessment point and treatment arm.

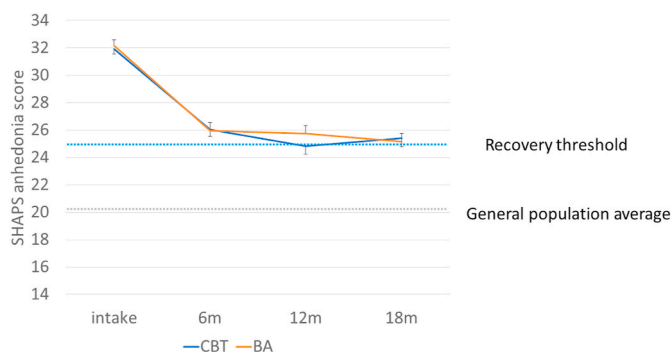| Assessment	Arm	SHAPS anhedonia	PHQ-9 depression
Baseline	CBT	98%	100%
	BA	98%	100%
Six months	CBT	74%	89%
	BA	72%	84%
Twelve months	CBT	74%	86%
	BA	66%	79%
Eighteen months	CBT	74%	82%
	BA	70%	80%

Fig. 1. SHAPS anhedonia score at each time point in BA and CBT arms.
Note: General population average taken from Trosteim et al. (2020); recovery threshold = scoring no more than 1.96 standard deviations above this general population average.

standard deviation on the SHAPS (i.e., standardized effect size differences). These can be interpreted according to Cohen's standard rules of thumb ($<.2$ = negligible; 0.2 to 0.49 = small; 0.50 to 0.79 = medium, ≥ 0.8 = large; Cohen, 1988). It has also been approximated that a value ≥ 0.24 equates to minimum clinically important difference for depression outcomes (Cuijpers, Turner, Koole, van Dijke, & Smit, 2014). Collapsing across groups, the standardized effect size difference was 0.97 for the change in SHAPS from intake to six months (a large and clinically relevant effect), 0.11 for the change in SHAPS from six months to twelve months (a negligible effect), and -0.01 for the change in SHAPS from twelve months to eighteen months (a negligible effect). The difference between CBT and BA at each assessment point was negligible (standardized effect size difference = 0.04 at intake, -0.01 at six months, 0.15 at twelve months, and -0.04 at eighteen months).

2.4. Does anhedonia at baseline moderate depression outcomes?

Table 3 reports PHQ-9 severity at each time point for each group and for the pooled sample. Greater SHAPS severity was significantly, though moderately (Cohen, 1988), associated with greater PHQ-9 severity at baseline, Pearson's r = 0.37, $P < .001$, indicating they are at least partially dissociable constructs.

The first GEE analyses tested for the possibility of a main moderating effect (excluding group terms from the analysis). In this model, baseline SHAPS predicted overall PHQ-9, B = 0.29 (SE = 0.03; 95% CI = 0.23 to 0.36), Wald χ^2 = 81.75, $P < .001$, such that those with greater SHAPS severity had greater PHQ-9 severity averaged across all time points. There was also a significant interaction between baseline SHAPS and time, B = -0.08 (SE = 0.02; 95% CI = -0.13 to -0.04), Wald χ^2 = 13.32, $P < .001$, indicating that the repair of the PHQ-9 over time was less marked in individuals with more severe baseline SHAPS symptoms. This provides evidence for a main moderating effect.

The analysis was then repeated a second time when additionally entering group and its two- and three-way interactions with time and baseline anhedonia to assess for the possibility of an interactive moderating effect. The three-way interaction was not significant, B = .01(SE = 0.01; 95% CI = -0.01 to 0.02), Wald χ^2 = 0.27, P = .61, which

means that the moderating effect of SHAPS did not vary as a function of CBT versus BA treatment (i.e., no significant interactive moderation effect).

Total depression-free days during follow up were on average 338 days (SD = 118; range 0–522) in the CBT arm, 323 days (SD = 128; 0 to 540) in the BA arm, and 331 days (SD = 123; range 0–540) when pooling across both arms. Regression analysis found a significant effect of baseline SHAPS (even when controlling for baseline PHQ-9) at step one of the model, B = -3.12 (SE = 1.45; 95% CI = -5.97 to -0.27), t = 2.15, P = .03, such that those with greater baseline SHAPS severity reported fewer depression-free days over the course of the follow up (a main moderating effect). The interaction between group and baseline SHAPS at step two of the model was significant, B = -5.92 (SE = 2.73, 95% CI = -11.30 to -0.55), t = 2.17, P = .03. This interaction was resolved by running the analysis separately for each group. This revealed that in the CBT group there was no significant association between baseline SHAPS and cumulative depression-free days, B = 1.73 (SE = 2.18; 95% CI = -2.59 to 6.06), $t < 1$. In the BA group, there were significantly fewer depression-free days in those who had greater baseline SHAPS severity, B = -6.87 (SE = 1.97; 95% CI = -10.77 to -2.98), t = 3.50, $P < .001$.

2.5. What is the magnitude of anhedonia versus depression repair?

Table 4 reports the proportion of clients in each group and the pooled sample meeting caseness, response, and reliable and clinically significant improvement criteria for the PHQ-9 and the SHAPS at each assessment point.

To simplify interpretation, the first set of GEE analyses collapsed across treatment arms and did not include any group terms. Caseness analysis found no main effect of measure, Exp(B) = 1.23 (95% CI = 0.86 to 0.1.76), Wald χ^2 = 1.24, P = .27, and a significant effect of time, Exp (B) = 0.52 (95% CI = 0.47 to 0.57), Wald χ^2 = 180.06, $P < .001$, which was qualified by a significant time by measure interaction, Exp(B) = 1.27 (95% CI = 1.06 to 1.51), Wald χ^2 = 6.81, $P < .01$. To resolve the significant interaction, rates of caseness of PHQ-9 versus SHAPS were compared at each time point separately using a series of binary logistic regressions. At baseline, there was no significant difference in caseness rates for SHAPS versus the PHQ-9, P = .12. At all other time points, caseness levels were significantly greater for the SHAPS than the PHQ-9, $Ps < .001$.

Response analysis found a main effect of measure, Exp(B) = 0.52 (95% CI = 0.37 to 0.75), Wald χ^2 = 12.97, $P < .001$, and time, Exp(B) = 1.22 (95% CI = 1.09 to 1.37), Wald χ^2 = 11.61, $P < .001$, but no significant time by measure interaction, Exp(B) = 0.93 (95% CI 0.80 to 1.08), Wald χ^2 = 0.98, P = .32. With regards to the main effect of measure, response rates were greater for the PHQ-9 than the SHAPS. To

Table 4
Proportion of clients in caseness, responding, and showing reliable and clinically significant improvement at each time point for SHAPS anhedonia and PHQ-9 depression.

	CBT		BA		Pooled	
	SHAPS	PHQ-9	SHAPS	PHQ-9	SHAPS	PHQ-9
Caseness						
Baseline	.92	.94	.90	.94	.91	.94
6m	.59	.43	.57	.48	.58	.45
12m	.46	.34	.53	.34	.49	.34
18m	.55	.41	.57	.34	.56	.37
Response						
Baseline to 6m	.30	.46	.32	.47	.31	.47
Baseline to 12m	.43	.59	.34	.59	.39	.60
Baseline to 18m	.40	.57	.36	.60	.38	.58
RCSC						
Baseline to 6m	.31	.48	.33	.47	.32	.48
Baseline to 12m	.46	.58	.34	.61	.40	.59
Baseline to 18m	.38	.56	.36	.53	.36	.58

Note: RCSC = reliable and clinically significant change (improvement).

Table 3
Depression severity (measured using the PHQ-9) at each time point in each arm and in the pooled sample.

Assessment	CBT	BA	Pooled
Baseline	17.43 (4.82)	17.71 (4.79)	17.57 (4.80)
Six months	9.74 (7.28)	9.77 (6.86)	9.75 (7.07)
Twelve months	8.37 (7.45)	8.37 (6.97)	8.37 (7.21)
Eighteen months	8.50 (7.21)	8.31 (7.09)	8.41 (7.14)

Note: data are mean (one standard deviation) values.

Table 5

Effect sizes for paired sample t-tests comparing baseline anhedonia and depression scores to each follow-up point in each treatment arm and the pooled sample.

		6m	12m	18m
CBT	SHAPS	.94 (.75–1.14)	1.05 (.85–1.27)	.92 (.75–1.18)
	PHQ-9	1.23 (1.06–1.42)	1.42 (1.24–1.62)	1.42 (1.24–1.63)
BA	SHAPS	.87 (.70–1.05)	.88 (.68–1.09)	.99 (.80–1.20)
	PHQ-9	1.37 (1.17–1.58)	1.54 (1.34–1.76)	1.54 (1.34–1.76)
Pooled	SHAPS	.90 (.78–1.04)	.97 (.82–1.13)	.98 (.83–1.13)
	PHQ-9	1.30 (1.17–1.44)	1.48 (1.34–1.63)	1.49 (1.35–1.64)

Note: Effect sizes are Hedges g (95% confidence interval).

resolve the main effect of time, comparisons were repeated on each separate consecutive pair of assessment points. Response rates improved from six months to twelve months, Exp(B) = 1.68 (95% CI = 1.36 to 2.09), Wald χ^2 = 22.10, P < .001, but there was no significant change from twelve months to eighteen months, Exp(B) = 0.95 (95% CI = 0.79 to 1.5), Wald χ^2 = 0.24, P = .63.

Reliable and clinically significant change analysis found a main effect of measure, Exp(B) = 0.60 (95% CI = 0.42 to 0.85), Wald χ^2 = 8.36, P < .01, a significant effect of time, Exp(B) = 1.18 (SE = 1.05; 95% CI = 1.05 to 1.32), Wald χ^2 = 7.67, P < .01, and no significant time by measure interaction, Exp(B) = 0.87 (95% CI = 0.75 to 1.01), Wald χ^2 = 3.75, P = .053. Rates of reliable and clinically significant improvement were greater for the PHQ-9 than for the SHAPS. To resolve the main effect of time, comparisons were again repeated on each separate consecutive pair of assessment points. RCSC rates improved from six months to twelve months, Exp(B) = 1.64 (95% CI = 1.32 to 2.04), Wald χ^2 = 20.10, P < .001, but there was no further change from twelve months to eighteen months, Exp(B) = 0.92 (95% CI = 0.76 to 1.11), Wald χ^2 = 0.77, P = .38.

All analyses were then repeated in a second set of GEE analyses when including group and its interactions with time and measure. In all cases, there were no significant main or interactive effects of group, *Ps* > .09.

The differences between SHAPS and PHQ-9 were at least 13% in the favour of the PHQ-9 for clinical caseness and at least 16% in favour of the PHQ-9 for response and RCSC analyses. While there are no established rules of thumb for what minimum clinically important difference is on categorical variables of this kind, this is nevertheless likely to be clinically meaningful.

Finally, a series of paired sample t-tests looked at change in the PHQ-9 and the SHAPS from baseline to each follow-up point in each arm separately and in the pooled sample. There were significant reductions in the SHAPS and the PHQ-9 from baseline to all other assessment points, whether looking at CBT alone, BA alone, or the pooled sample, *ts* > 9.10, *Ps* < .001. Table 5 reports the Hedges g effect size (and its 95% confidence interval) for each of these analyses. All effect sizes were large according to conventional rules of thumb (Hedges *gs* > 0.8; Cohen, 1988). Visual inspection revealed that the effect sizes were greater for the PHQ-9 than the SHAPS at each assessment point, irrespective of whether the focus was on each arm separately or the pooled sample. As the 95% confidence intervals for the SHAPS and the PHQ-9 effect sizes did not overlap at any time points, this suggests treatment was having a greater impact on depression relative to anhedonia.

3. Discussion

The present study conducted an analysis of the COBRA trial to assess: (1) the extent to which CBT and BA for depression repaired anhedonia over the short and long term; (2) whether anhedonia at baseline moderated depression treatment outcomes following CBT and BA; and (3) if the repair of anhedonic symptoms was of a similar magnitude to the repair of depression symptoms during treatment.

Both CBT and BA were shown to be similarly, but only partially, effective at repairing anhedonia, with no significant difference between

treatment arms. Anhedonia at baseline was significantly above general population averages. Although treatment improved anhedonia from baseline to six months, anhedonia levels remained significantly above general population averages. There was no further significant improvement in anhedonia symptoms during the follow-up phase. At each assessment point, a relatively high proportion of clients remained in clinical caseness (58% at six months 49% at twelve months, and 56% at eighteen months) and a relatively low proportion of clients met response criteria (31% at six months, 39% at twelve months, and 38% at eighteen months).

In moderation analyses, baseline anhedonia severity did to some extent predict how effective CBT and BA were at improving depression symptoms. Individuals with more severe anhedonia had greater depression symptoms averaged across all time points and also showed a smaller repair of depression symptom over time (a main moderation effect). This relationship held irrespective of whether individuals received BA or CBT (i.e., there was no evidence of an interactive moderation effect). When considering depression-free days across the entire follow-up period (a measure of sustained benefit from treatment), individuals with more severe baseline anhedonic symptoms had fewer depression-free days on average (a main moderation effect), even when controlling for baseline depression symptoms. There was also a significant interactive moderation effect, such that the relationship between anhedonia at baseline and depression repair varied as a function of treatment group. While anhedonia at baseline did not predict depression repair during CBT, those with more marked anhedonia showed less-optimal depression repair when receiving BA.

Across both treatment arms and at all follow-up time points, anhedonia was repaired to a lesser extent than depression, irrespective of whether the focus was on caseness, response or reliable and clinically significant improvement. Similarly, in continuous within-arm analyses, effect sizes were of greater magnitude for depression repair than anhedonia repair. A strength of these analyses is they are comparing anhedonia and depression on identical metrics of change. The one slight exception was the way in which caseness was defined, which used an externally defined criterion for depression (scoring >9 based on cross-validation of the PHQ-9 with diagnostic interviews) and a statistically defined criterion for anhedonia (scoring more than 1.96 standard deviations about the general population average). This could potentially have led to arbitrary differences in how liberal or conservative the caseness criteria were between measures. It is reassuring in this regard that if using a statistical criterion for the PHQ-9 based on the general population normative data, a broadly similar (but slightly more conservative) caseness cut-off would have been selected (scoring >10 rather than >9). This would have led to even clearer evidence of more marked depression repair relative to anhedonia repair on this metric.

These findings overcome a series of limitations of previous studies examining the extent of anhedonia repair during psychotherapy. For example, Dunn et al. (2020) relied on measures of positive affect change as a proxy for anhedonia, only looked at short term outcomes during acute treatment, and only included a single form of therapy (CBT). The present findings reach a similar conclusion that CBT fails to repair anhedonia adequately, but this time using a validated measure of consummatory anhedonia rather than a positive affect scale. Further, the current results show that anhedonia improves only during the acute treatment phase of CBT and there are no further improvements during follow up (providing no support for the notion of an anhedonia 'sleeper' effect; Fluckiger & Del Re, 2017). It now seems relatively incontrovertible that CBT is not currently optimized for repairing positive valence system disturbances in depression, on the basis that CBT fails to normalize positive affect levels (Dunn et al., 2020), fails to normalize wellbeing levels (Widnall et al., 2020), and in the present analyses fails to normalize anhedonia levels. Moreover, the extent of repair of the PVS is less marked than the extent of NVS repair across all of these studies. There was greater negative affect repair than positive affect repair in Dunn et al. (2020), greater depression and anxiety symptom repair than

wellbeing repair in Widnall et al. (2020), and greater depression symptom repair in the present study.

Perhaps most importantly, the current data show for the first time that an identical pattern of findings emerges when looking at an ostensibly reward-focused psychotherapy like BA. It therefore is also looking increasingly likely that BA is insufficient to repair positive valence system deficits, although given the present study is the first to test this issue directly it will now be important to see if this finding replicates in other samples.

The fact that BA was no better than CBT at repairing anhedonia is at first glance paradoxical, given that BA has a sole focus on reconnecting individuals to rewarding activity. One possibility is that anhedonia is not driven just by a behavioural disconnection from previously rewarding stimuli. In addition, there may be a range of psychological mechanisms that limit the amount of pleasure that is experienced during the times individuals do behaviourally engage with reward. For example, basic science studies have shown that the tendency to engage in dampening appraisals (e.g., 'this is too good to last') can reduce the amount of pleasure individuals experience and can even turn a potentially rewarding experience into something actively aversive (Burr, Javiad, Jell, Werner-Seidler, & Dunn, 2017). Moreover, a failure to engage experientially (pay attention to sensory experience as it unfolds moment to moment) during rewarding activities may limit the amount of pleasure that is experienced (Gadeikis et al., 2018). It may be that BA protocols need to be updated to additionally target these 'pleasure blocking' mechanisms if they wish to repair anhedonia fully (see Dunn, 2019; Forbes, 2020).

A range of novel positive clinical psychology interventions are now emerging that do more explicitly target up-regulation of the positive valence system, including positive CBT (Geschwind, Arntz, Bannink, & Peeters, 2019), Augmented Depression Therapy (Dunn et al., 2019), Positive Affect Treatment (PAT; Craske et al., 2019), Wellbeing Therapy (Fava, Rafanelli, Grandi, Conti, & Belluardo, 1998, 2004) and group based positive psychology protocols (Chaves, Lopez-Gomez, Hervas, & Vazquez, 2017). Many of these therapies do explicitly target these pleasure blocking mechanisms. These treatments are worthy of further empirical examination to establish if they are better able to repair anhedonia than current therapies like CBT and BA.

It is also important to consider if alternatives to psychotherapy might have potential for better repairing anhedonia. There is preliminary evidence that novel pharmacological approaches like ketamine infusions (for example, Rodrigues et al., 2020) or psilocybin (Carhart-Harris et al., 2021) can significantly improve anhedonia symptoms measured on scales like the SHAPS. A recent systematic review identified seventeen trials evaluating the impact of pharmacological interventions on anhedonia symptoms (Cao et al., 2019), concluding that most agents studied (including melatonergic agents, glutamatergic agents, monoaminergic agents, stimulants, and psychedelics) to some extent repaired anhedonia. There is also emerging evidence that stimulation approaches like transcranial magnetic stimulation (TMS) can improve anhedonia measured on the SHAPS (Fukuda et al., 2021). Inspection of the findings from across these studies suggests that while these pharmacological and stimulation treatments can improve anhedonia, individuals nevertheless typically remain with some residual anhedonic symptoms (and score above general population average levels) after treatment has finished. Therefore, alternative (non-psychotherapy) treatments for depression also appear not to be currently optimized in the treatment of anhedonia and further refinement is needed.

The moderation findings replicate and extend previous work showing that more marked baseline anhedonia severity is associated with poorer depression treatment outcomes following CBT in combination with anti-depressant medication (Khazanov, Ruscio, & Forbes, 2020; McMakin et al., 2012). The present results provide clearer evidence that anhedonia moderates response to psychological therapy, not just pharmacotherapy, as previous trials all had combined therapy and drug treatment arms with no therapy-only condition. Further, the present results extend previous findings by showing that the moderation is found across both BA and CBT and is observed both in terms of depression severity at acute treatment end and in depression-free days over an extended follow-up period. This pattern further supports the claim that anhedonia is a prognostically important component of depression (Bennik et al., 2014; Pine et al., 1999; Spijker et al., 2001; Wilcox & Anthony, 2004). As an interactive moderation effect of anhedonia was only found when considering depression-free days as the outcome (and not when looking at overall depression severity), the present data only provide a weak mandate to recommend personalized allocation to particular treatment based on baseline anhedonia presentation. However, a clearer pattern may have emerged if comparing more distinct treatments or looking at anhedonia in combination with other baseline features to predict treatment response (cf. Cohen & DeRubeis, 2018).

There are a number of limitations with the current work that need to be considered First, while the SHAPS is a validated and widely used measure of anhedonia, it focuses solely on the hedonic (consummatory) component of the construct. It is increasingly realized that anhedonia is multi-facetted, including anticipation and recall in addition to consumption and also extending to broader motivational and learning disturbances of the reward system. Future work should look at the impact of treatment on multi-dimensional measures of anhedonia, perhaps aligning to the components of the reward system outlined in the Research Domains Criteria framework (Insel, Cuthbert, Garvey et al., 2010; see Khazanov, Ruscio & Forbes et al., 2020). Second, the COBRA trial only compared two active psychological therapies (BA and CBT). It is now considered unethical to include a no-intervention control group for conditions like depression where we have effective therapies (Gold, Enck & Hasselmann et al., 2017), but this nevertheless reduces the sensitivity and interpretability of the moderation analyses. Third, as anhedonia was not measured at multiple intervals during the acute treatment phase, it was not possible to examine if anhedonia repair mediated subsequent depression improvement during acute BA or CBT treatment. Therefore, we were unable to interrogate in the current data whether repair in anhedonia was an active mechanism driving greater depression symptom relief during treatment. Future research could collect depression and anhedonia symptoms at regular intervals through treatment to test this possibility in a way that meets temporal precedence criteria for mediation analyses.

A fourth limitation is that there were slightly greater levels of missing data in the BA versus the CBT arm. While GEE models account for missing data, this nevertheless suggests a potential source of bias in the data. This would have been more of a concern if clear differences emerged between treatment arms, which was not the case in the any of the current analyses. Fifth, the finding that treatment repairs depression to a greater degree than it does anhedonia may simply be an artefact of differential sensitivity to change of the PHQ-9 versus the SHAPS measure. However, the reliable and clinically significant change analyses (which look at whether clients move closer to the general population distribution than the clinical population distribution) showed the same pattern of findings and are relatively immune to this problem. Sixth, the depression-free days analysis relies on participants being able retrospectively to recall depression status over an extended period of time. Future research should consider more regular screening of depression status over the follow-up period, for example fortnightly completion of a brief screening measure of depression like the Patient Health Questionniare-2 (PHQ-2; Kroenke, Spitzer, & Williams, 2003). Seventh, we made no corrections for multiple comparisons when running pairwise comparisons to resolve significant omnibus effects. However, all significant pairwise comparisons would have still been significant if using a Bonferonni correction.

Finally, while a provisional analysis plan was prepared prior to accessing the data (and was submitted and approved by an ethics board), a number of changes were made to this analysis plan once conducting the analysis and sharing findings with collaborators. In particular, a

post-hoc decision was made to move to analyzing the data using GEE models to capture the longitudinal nature of the data rather than conducting a series of separate ANCOVA analyses at each time point. An identical pattern of conclusions was reached for the primary hypotheses when carrying out the original, pre-specified ANCOVA analysis plan. Nevertheless, as the analyses reported here did change and were extended from those that were pre-specified, the conservative approach is to view these analyses as hypothesis generating and not hypothesis testing and to recommend that findings are replicated in future research.

In summary, the present results provide evidence that: 1) neither CBT or BA satisfactorily repairs anhedonia over the short or longer term; 2) that greater anhedonia severity at intake predicts poorer depression treatment outcomes; and 3) that the extent of anhedonia repair during treatment is less marked than the extent of depression repair. This highlights the need for treatment innovation to develop more effective ways to up-regulate the positive valence system (and therefore repair anhedonia) in depression.

CRediT authorship contribution statement

Batool Alsayednasser: Formal analysis, Methodology, Project administration, Writing - original draft. **Emily Widnall:** Formal analysis, Data curation, Writing - review & editing. **Heather O'Mahen:** Writing - Conceptualization, Funding acquisition, Writing review & editing. **Kim Wright:** Writing - Conceptualization, Funding acquisition, Writing review & editing. **Fiona Warren:** Formal analysis, Conceptualization, Funding acquisition, Writing - review & editing. **Asha Ladwa:** Formal analysis, Writing - review & editing. **Gabriela Kattan Khazanov:** Writing - review & editing. **Sarah Byford:** Conceptualization, Funding acquisition, Writing review & editing. **Willem Kuyken:** Conceptualization, Funding acquisition, Writing review & editing. **Ed Watkins:** Conceptualization, Funding acquisition, Writing review & editing. **David Ekers:** Conceptualization, Funding acquisition, Writing review & editing. **Nigel Reed:** Conceptualization, Funding acquisition, Writing review & editing. **Emily Fletcher:** Data curation, Writing review & editing. **Dean McMillan:** Conceptualization, Funding acquisition, Writing review & editing. **Paul Farrand:** Conceptualization, Funding acquisition, Writing review & editing. **David Richards:** Conceptualization, Funding acquisition, Data Curation, Project Administration, Writing review & editing. **Barnaby D. Dunn:** Formal analysis, Supervision, Conceptualization, Data curation, Methodology, Project administration, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript (TRUE).

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2022.104185.

References

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

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

Bennik, E. C., Nederhof, E., Ormel, J., & Oldehinkel, A. J. (2014). Anhedonia and depressed mood in adolescence: Course, stability, and reciprocal relation in the TRAILS study. *European Child & Adolescent Psychiatry, 23*, 579–586. https://doi.org/10.1007/s00787-013-0481-z

Burr, L.-A., Javiad, M., Jell, G., Werner-Seidler, A., & Dunn, B. D. (2017). Turning lemonade into lemons: Dampening appraisals reduce positive affect and increase negative affect during positive activity scheduling. *Behaviour Research and Therapy, 91*, 91–101. https://doi.org/10.1016/j.brat.2017.01.010

Cao, B., Zhu, J., Zuckerman, H., Rosenblat, J. D., Brietzke, E., Pan, Z., et al. (2019). Pharmacological interventions targeting anhedonia in patients with major depressive disorder: A systematic review. *Progress in Neuropsychopharmacology & Biological Psychiatry, 92*, 109–117. https://doi.org/10.1016/j.pnpbp.2019.01.002

Carhart-Harris, R., Giribaldi, B., Watts, R., Baker-Jones, M., Murphy-Biner, A., Murphy, R., et al. (2021). Trial of psilocybin versus escitalopram for depression. *New England Journal of Medicine, 384*, 1402–1411. https://doi.org/10.1056/nejmoa2032994

Chaves, C., Lopez-Gomez, I., Hervas, G., & Vazquez, C. (2017). A comparative study on the efficacy of a positive psychology intervention and a cognitive behavioral therapy for clinical depression. *Cognitive Therapy and Research, 41*, 417–433. https://doi.org/10.1007/s10608-016-9778-9

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment selection in depression. *Annual Review of Clinical Psychology, 14*, 209–236. https://doi.org/10.1146/annurev-clinpsy-050817-084746

Craske, M. G., Meuret, A. E., Ritz, T., Treanor, M., Dour, H., & Rosenfield, D. (2019). Positive affect treatment for depression and anxiety: A randomized clinical trial for a core feature of anhedonia. *Journal of Consulting and Clinical Psychology, 87*, 457–471. https://doi.org/10.1037/ccp0000396

Cuijpers, P., Karyotaki, E., Ciharova, M., Miguel, C., Noma, H., & Furukawa, T. A. (2021). The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. *Acta Psychiatrica Scandinavica, 144*, 288–299. https://doi.org/10.1111/acps.13335

Cuijpers, P., Karyotaki, E., Weitz, E., Andersson, G., Hollon, S. D., & van Straten, A. (2014). The effects of psychotherapies for major depression in adults on remission, recovery, and improvement: A meta-analysis. *Journal of Affective Disorders, 159*, 118–126. https://doi.org/10.1016/j.jad.2014.02.026

Cuijpers, P., Turner, E. H., Koole, S. L., van Dijke, A., & Smit, F. (2014). What is the threshold for a clinically relevant effect? The case of major depressive disorders. *Depression and Anxiety, 31*, 374–378. https://doi.org/10.1002/da.22249

Demyttenaere, K., Donneau, A. F., Albert, A., Ansseau, M., Constant, E., & Van Heeringen, K. (2015). What is important in being cured from depression? Discordance between physicians and patients (1). *Journal of Affective Disorders, 174*, 390–396. https://doi.org/10.1016/j.jad.2014.12.004

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

Dunn, B. D. (2012). Helping depressed clients reconnect to positive emotion experience: Current insights and future directions. *Clinical Psychology & Psychotherapy, 19*, 326–340. https://doi.org/10.1002/cpp.1799

Dunn, B. D. (2019). Augmenting cognitive behavioural therapy to build positive mood in depression. In J. Gruber (Ed.), *Oxford handbook of positive emotion and psychopathology* (pp. 539–560). Oxford, England: Oxford University Press.

Dunn, B. D., German, R. E., Khazanova, G., Xu, C., Hollen, S. D., & DeRubeis, R. J. (2020). Changes in positive and negative affect during pharmacological treatment and cognitive therapy for major depressive disorder: A secondary analysis of two randomized controlled trials. *Clinical Psychological Science, 8*, 36–51. https://doi.org/10.1177/2167702619863427

Dunn, B. D., & Roberts, H. (2016). Improving the capacity to treat depression using talking therapies. In A. M. Wood, & J. Johnson (Eds.), *The wiley handbook of positive clinical psychology* (pp. 183–204). Hoboken, NJ: John Wiley & Sons.

Dunn, B. D., Widnall, E., Reed, N., Owens, C., Campbell, J., & Kuyken, W. (2019). Bringing light into darkness: A multiple baseline mixed methods case series evaluation of augmented depression therapy (ADepT). *Behaviour Research and Therapy, 120*. https://doi.org/10.1016/j.brat.2019.103418. Article 103418.

Fava, G. A., Rafanelli, C., Grandi, S., Conti, S., & Belluardo, P. (1998). Prevention of recurrent depression with cognitive behavioral therapy: Preliminary findings. *Archives of General Psychiatry, 55*, 816–820. https://doi.org/10.1001/archpsyc.55.9.816

Fava, G. A., Ruini, C., Rafanelli, C., Finos, L., Conti, S., & Grandi, S. (2004). Six-Year outcome of cognitive behavior therapy for prevention of recurrent depression. *American Journal of Psychiatry, 161*(10), 1872–1876. https://doi.org/10.1176/appi.ajp.161.10.1872

Fluckiger, C., & Del Re, A. C. (2017). The sleeper effect between psychotherapy orientations: A strategic argument of sustainability of treatment effects at follow-up. *Epidemiology and Psychiatric Sciences, 26*, 442–444. https://doi.org/10.1017/S2045796016000780

Forbes, C. N. (2020). New directions in behavioural activation: Using findings from basic science and translational neuroscience to inform the exploration of potential mechanisms of change. *Clinical Psychology Review, 79*, Article 101860. https://doi.org/10.1016/j.cpr.2020.101860

Franken, H. A., Rassin, E., & Muris, P. (2007). The assessment of anhedonia in clinical and non-clinical populations: Further validation of the Snaith-Hamilton Pleasure Scale (SHAPS). *Journal of Affective Disorders, 99*, 83–89. https://doi.org/10.1016/j.jad.2006.08.020

Fukuda, A. M., Kang, J. W. D., Gobin, A. P., Tirrell, E., Kokdere, F., & Carpenter, L. L. (2021). Effects of transcranial magnetic stimulation on anhedonia in treatment resistant major depressive disorder. *Brain & Behaviour, 11*, e2329. https://doi.org/10.1002/brb3.2329

Gadeikis, D, Bos, N., Schweizer, S, Murphy, F., & Dunn, B. D. (2017). Engaging in an experiential processing mode increases positive emotional response during recall of pleasant autobiographical memories. *Behaviour Research and Therapy, 92*, 68–76. https://doi.org/10.1016/j.brat.2017.02.005le

Geschwind, N., Arntz, A., Bannink, F., & Peeters, F. (2019). Positive cognitive behavior therapy in the treatment of depression: A randomized order within-subject comparison with traditional cognitive behavior therapy. *Behaviour Research and Therapy, 116*, 119–130. https://doi.org/10.1016/j.brat.2019.03.005

Gold, S. M., Enck, P., Hasselmann, H., Friede, T., Hegerl, U., Mohr, D. C., et al. (2017). Control conditions for randomised trials of behavioural interventions in psychiatry: A decision framework. *The Lancet Psychiatry, 4*(9), 725–732. https://doi.org/10.1016/S2215-0366(17)30153-0

IBM Corp. Released. (2019). *IBM SPSS statistics for windows, version 26.0*. Armonk, NY: IBM Corp.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., et al. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry, 167*, 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

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

Keyes, C. L. M. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Social Behavior, 43*(2), 207–222. https://doi.org/10.2307/3090197

Khazanov, G. K., Ruscio, A. M., & Forbes, C. N. (2020). The positive valence systems scale: Development and validation. *Assessment, 27*, 1045–1069. https://doi.org/10.1177/1073191119869836

Khazanov, G. K., Xu, C., Dunn, B. D., Cohen, Z. D., DeRubeis, R. J., & Hollon, S. D. (2020). Distress and anhedonia as predictors of depression treatment outcome: A secondary analysis of a randomized clinical trial. *Behaviour Research and Therapy, 125*, Article 103507. https://doi.org/10.1016/j.brat.2019.103507

Kocalevent, R. D., Hinz, A., & Brähler, E. (2013). Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population. *General Hospital Psychiatry, 35*, 551–555. https://doi.org/10.1016/j.genhosppsych.2013.04.006

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, 59*, 877–883. https://doi.org/10.1001/archpsyc.59.10.877

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

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2003). The patient health questionnaire-2: Validity of a two-item depression screener. *Medical Care, 41*, 1284–1292. https://doi.org/10.1097/01.mlr.0000093487.78664.3c

Martell, C. R., Dimidjian, S., Herman-Dunn, R., Lewinsohn, P. M., & DeRubeis, R. J. (2010). *Behavioural activation for depression: A clinician's guide*. Guilford Press.

Mazzucchelli, T. G., Kane, R. T., & Rees, C. S. (2010). Behavioral activation interventions for well-being: A meta-analysis. *The Journal of Positive Psychology, 5*, 105–121. https://doi.org/10.1080/17439760903569154

McMakin, D. L., Olino, T. M., Porta, G., Dietz, L. J., Emslie, G., Clarke, G., et al. (2012). Anhedonia predicts poorer recovery among youth with selective serotonin reuptake inhibitor treatment–resistant depression. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*, 404–411. https://doi.org/10.1016/j.jaac.2012.01.011

Medeiros, G. C., Rush, A. J., Jha, M., Carmody, T., Furman, J. L., Czysz, A. H., et al. (2020). Positive and negative valence systems in major depression have distinct clinical features, responses to antidepressants, and relationships with immunomarkers. *Depression and Anxiety, 37*, 771–783. https://doi.org/10.1002/da.23006

Moore, R. G., & Garland, A. (2004). *Cognitive therapy for chronic and persistent depression*. John Wiley & Sons.

Murray, C. J., & Lopez, A. D. (1997). Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *The Lancet, 349*(9063), 1436–1442. https://doi.org/10.1016/S0140-6736(96)07495-8

Paulus, M. P., Stein, M. B., Craske, M. G., Bookheimer, S., Taylor, C. T., Simmons, A. N., … Fan, B. (2017). Latent variable analysis of positive and negative valence processing focused on symptom and behavioural units of analysis in mood and anxiety disorders. *Journal of Affective Disorders, 216*, 17–29. https://doi.org/10.1016/j.jad.2016.12.046

Pelizza, L., & Ferrari, A. (2009). Anhedonia in schizophrenia and major depression; state or trait? *Annals of General Psychiatry, 8*, 22. https://doi.org/10.1186/1744-859X-8-22

Pine, D. S., Cohen, E., Cohen, P., & Brook, J. (1999). Adolescent depressive symptoms as predictors of adult depression: Moodiness or mood disorder? The. *American Journal of Psychiatry, 156*, 133–135. https://doi.org/10.1176/ajp.156.1.133

Rhodes, S., Richards, D. A., Ekers, D., McMillan, D., Byford, S., Farrand, P. A., et al. (2014). Cost and outcome of behavioural activation versus cognitive behaviour therapy for depression (COBRA): Study protocol for a randomised controlled trial. *Trials, 15*, 29. https://doi.org/10.1186/1745-6215-15-29

Richards, D. A., Ekers, D., McMillan, D., Taylor, R. S., Byford, S., Warren, F. C., et al. (2016). Cost and outcome of behavioural activation versus cognitive behaviour therapy for depression (COBRA): A randomised, controlled, non-inferiority trail. *The Lancet, 388*, 871–880. https://doi.org/10.1016/S0140-6736(16)31140-0

Rizvi, S. J., Pizzagalli, D. A., Sproule, B. A., & Kennedy, S. H. (2016). Assessing anhedonia in depression: Potentials and pitfalls. *Neuroscience & Biobehavioral Reviews, 65*, 21–35. https://doi.org/10.1016/j.neubiorev.2016.03.004

Rodrigues, N. B., McIntyre, R. S., Lipsitz, O., Cha, D. S., Lee, Y., Gill, H., et al. (2020). Changes in symptoms of anhedonia in adults with major depressive or bipolar disorder receiving IV ketamine: Results from the Candian Rapid Treatment Centre of Excellence. *Journal of Affective Disorders, 276*, 570–575. https://doi.org/10.1016/j.jad.2020.07.083

Snaith, R. P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., & Trigwell, P. (1995). A scale for the assessment of hedonic tone: The Snaith-Hamilton pleasure scale. *British Journal of Psychiatry, 167*(1), 99–103. https://doi.org/10.1192/bjp.167.1.99

Spijker, J., Bijl, R. V., de Graaf, R., & Nolen, W. A. (2001). Determinants of poor 1-year outcome of DSM-III-R major depression in the general population: Results of The Netherlands mental health survey and incidence study (NEMESIS). *Acta Psychiatrica Scandinavica, 103*, 122–130. https://doi.org/10.1034/j.1600-0447.2001.103002122.x

Spitzer, R. L., Williams, J. B., Gibbon, M., & First, M. B. (1992). The structured clinical interview for DSM-III—r (SCID): I. History, rationale, and description. *Archives of General Psychiatry, 49*(8), 624–629. https://doi.org/10.1001/archpsyc.1992.01820080032005

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

Trøstheim, M., Eikemo, M., Meir, R., Hansen, I., Paul, E., Kroll, S. L., … Leknes, S. (2020). Assessment of anhedonia in adults with and without mental illness: A systematic review and meta-analysis. *JAMA Network Open, 3*. e2013233-e2013233. https://doi.org/10.1001/jamanetworkopen.2020.13233

Vannoy, S. D., Arean, P., & Unützer, J. (2010). Advantages of using estimated depression-free days for evaluating treatment efficacy. *Psychiatric Services, 61*(2), 160–163. https://doi.org/10.1176/appi.ps.61.2.160

Vittengl, J. R., Clark, L. A., Dunn, T. W., & Jarrett, R. B. (2007). Reducing relapse and recurrence in unipolar depression: A comparative meta-analysis of cognitive-behavioral therapy's effects. *Journal of Consulting and Clinical Psychology, 75*(3), 475–488. https://doi.org/10.1037/0022-006X.75.3.475

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76*, 820–838. https://doi.org/10.1037/0022-3514.76.5.820

Widnall, E., Price, A., Trompetter, H., & Dunn, B. D. (2020). Routine cognitive behavioural therapy for anxiety and depression is more effective at repairing symptoms of psychopathology than enhancing wellbeing. *Cognitive Therapy and Research, 44*(1), 28–39. https://doi.org/10.1007/s10608-019-10041-y

Wilcox, H. C., & Anthony, J. C. (2004). Child and adolescent clinical features as forerunners of adult-onset major depressive disorder: Retrospective evidence from an epidemiological sample. *Journal of Affective Disorders, 82*, 9–20. https://doi.org/10.1016/j.jad.2003.10.007

World Health Organization. (2017). *Depression and Other Common Mental Disorders: Global Health Estimates*. Licence: Genera, World Health Organization.

Zimmerman, M., McGlinchey, J. B., Posternak, M. A., Friedman, M., Attiullah, N., & Boerescu, D. (2006). How should remission from depression be defined? The depressed patient's perspective. *American Journal of Psychiatry, 163*(1), 148–150. https://doi.org/10.1176/appi.ajp.163.1.148