ORIGINAL ARTICLE

Routine Cognitive Behavioural Therapy for Anxiety and Depression is More Effective at Repairing Symptoms of Psychopathology than Enhancing Wellbeing

Emily Widnall[1] · Alice Price[1] · Hester Trompetter[2] · Barnaby D. Dunn[1]

Abstract

The primary focus of classic cognitive behavioural therapy (CBT) for depression and anxiety is on decreasing symptoms of psychopathology. However, there is increasing recognition that it is also important to enhance wellbeing during therapy. This study investigates the extent to which classic CBT for anxiety and depression leads to symptom relief versus wellbeing enhancement, analysing routine outcomes in patients receiving CBT in high intensity Improving Access to Psychological Therapy (IAPT) Services in the UK. At intake, there were marked symptoms of anxiety and depression (a majority of participants scoring in the severe range) and deficits in wellbeing (a majority of participants classified as languishing, relative to general population normative data). CBT was more effective at reducing symptoms of anxiety and depression than repairing wellbeing. As a result, at the end of treatment, a greater proportion of participants met recovery criteria for anxiety and depression than had moved from languishing into average or flourishing levels of wellbeing. Given the importance of wellbeing to client definitions of recovery, the present results suggest a greater emphasis should be placed on enhancing wellbeing in classic CBT.

Keywords Wellbeing · Positive affect · Depression · Anxiety · Psychopathology · Cognitive behaviour therapy

Anxiety and depressive disorders are prevalent, recurrent, frequently comorbid mental health conditions that are a significant cause of worldwide disability (WHO 2017; Kessler et al. 2005, 2007). Psychological therapies for depression and anxiety predominantly focus on decreasing symptoms of psychopathology, conceptually underpinned by disease models arguing that recovery from mental illness equates to an absence of symptoms. Recovery is defined as falling beneath a cut-off on symptom measures, which moves individuals from a 'clinical' range to a 'normal' range.

✉ Barnaby D. Dunn
b.d.dunn@exeter.ac.uk

[1] Mood Disorders Centre, University of Exeter, Exeter EX4 4QG, UK

[2] Department of Medical and Clinical Psychology, Tilburg University, Tilburg, The Netherlands

When asking patients what recovery involves, a different perspective is increasingly emerging. A key component of recovery is the capacity to experience increased wellbeing, which can be defined as experiencing positive emotional states, feeling connected to and valued by others, and having a sense of meaning and purpose in life (Keyes 2002, 2005). Wellbeing enhancement is at least as important a part of recovery to patients as relief from symptoms (Zimmerman et al. 2006; Demyttenaere et al. 2015) and predicts future resilience (Garland et al. 2010; Wood and Joseph 2010). This perspective resonates with the broader recovery movement arguing that recovery means individuals living a valued and enjoyable life and minimising the extent to which symptoms impede this goal (Slade 2010).

There are different conceptual views as to how distinct wellbeing and symptoms are from one another. A single continua model sees symptoms and wellbeing as two opposite ends of a bipolar dimension. Recovery involves moving individuals from the symptomatic end of the distribution and as far as possible into the wellbeing end of the distribution (Huppert 2014). An alternative perspective is that symptoms and wellbeing represent orthogonal dimensions

(the dual-continua model: Tudor 1996; Provencher and Keyes 2011), based on findings that the two constructs are only moderately correlated in some samples and are best accounted for as two latent dimensions rather than a single latent dimension (Keyes 2005, 2006, 2007). Irrespective of which of these positions is adopted, treatment for anxiety and depression should aim to move individuals from a position of 'languishing' (low wellbeing, high mental illness) to one of 'flourishing' (high wellbeing, low mental illness) (Keyes and Lopez 2002; Coulombe et al. 2016). Moreover, regardless of which of these frameworks is correct, it is likely that different intervention strategies will be required to develop wellbeing as opposed to reduce symptoms.

One of the most robustly validated and extensively deployed treatment modalities for anxiety and depression is cognitive therapy (CT), in UK contexts more often referred to as cognitive behavioural therapy (CBT). Classic CBT approaches focus on correcting negative biases in information processing and avoidant behaviour in an effort to reduce symptoms of depression and anxiety (e.g. Beck et al. 1979; Clark and Beck 2010).[1] There is good evidence that classic CBT is effective (although not optimally so) in reducing depression symptoms during acute episodes and to some extent minimising the risk of subsequent relapse (Cuijpers et al. 2013). Similarly, there is evidence of acute and sustained benefit for CBT protocols for specific anxiety disorders, but nevertheless with a subset of clients not responding or showing a chronic, relapsing course (Hofmann and Smits 2008; Ali et al. 2017). This mirrors effective, but nevertheless sub-optimal, outcomes observed following other psychological and pharmacological treatment modalities for anxiety and depression. Moreover, recovery rates in routine practice may be substantially lower than those observed in clinical trials (Lambert 2017).

Given the disease model underpinning classic CBT, it may well be better at reducing symptoms of mental illness than building wellbeing and adaptations are likely required to existing treatments to maximise wellbeing gains. In particular, symptom relief could be a necessary but not sufficient component for enhancement of wellbeing. To optimise

wellbeing, treatment may require symptom relief *and* also systematic attention to improving day-to-day positive mood, functioning, broader quality of life, and social connection/identity. However, very few classic CBT trials have reported wellbeing outcomes, so the relative efficacy of the approach in repairing symptoms relative to wellbeing is yet to be definitively established.

We are aware of a handful of studies that indirectly examine the extent to which classic CBT repairs symptoms relative to wellbeing. These all used CBT as a control condition in small scale randomised controlled trials evaluating novel positive psychotherapies. One trial evaluated the efficacy of group CBT, relative to a group positive psychology intervention, in treating acute depression and dysthymia using a broad array of symptom and wellbeing measures (Chaves et al. 2017). Pre-post effect sizes in CBT were larger for clinical variables (including depression; Cohen's $d = 0.44$) than for positive functioning variables (including wellbeing; Cohen's $d = 0.26$). Three other studies did not report pre-post effect sizes for both wellbeing and symptoms, but we calculated these using the means and standard deviations described in the papers. We report Hedges g, as this is appropriate for smaller sample sizes.[2] Fava et al. (1998a) randomized 20 individuals with residual symptoms of affective disorder to receive 8 sessions of either group CBT or group wellbeing therapy. Individuals in the CBT arm showed large improvements on interviewer-rated depression ($g = 1.35$) and reported small improvements on self-rated wellbeing ($g = 0.36$). Fava et al. (2005) randomized 16 individuals with acute generalized anxiety disorder to receive either 8 sessions of CBT or 8 sessions of CBT combined with wellbeing therapy. In the CBT only arm, there were large improvement in interviewer rated depression ($g = 0.96$) and anxiety ($g = 1.86$) but only medium improvements in self-reported wellbeing ($g = 0.52$). Geschwind et al. (2019) allocated 40 acutely depressed participants to receive a combination of individual CBT and a novel positive form of CBT (CBT+) in different orders, using a cross-over randomized controlled trial design. It is possible to isolate the pre-post effects in the 20 depressed participants who were randomized to CBT followed by CBT+, by focusing solely on observed change in the first treatment block only. After the first 8 CBT sessions, these participants showed a large effect size improvement in depression symptoms ($g = 1.01$) and a medium effect size improvement in wellbeing ($g = 0.67$).

This pattern of findings suggests that CBT is more effective at repairing symptoms than wellbeing. However, this conclusion is undermined by a number of methodological issues. None of these studies directly compared the

---

[1] Over the past forty years cognitive protocols have evolved and proliferated. CBT can therefore be broadly defined as a family of interventions that aim to change how an individual thinks and behaves in an effort to repair mood. We are restricting our analysis to 'classic' Beckian protocols. By this we mean 'second wave' treatments that combine behavioural and cognitive approaches to modify negatively biased cognitive content in anxiety and depression. We distinguish these from more recent 'third wave' cognitive protocols that change the relationship to as well as the content of cognition. These 'third wave approaches include Acceptance and Commitment Therapy, Dialectical Behaviour Therapy, and Mindfulness Based Cognitive Therapy. CBT protocols have also recently been adapted to have a more explicit recovery focus.

[2] In these studies, the pre-post correlation for each measure was never described, so we assumed in all cases it was 0.5.

magnitude of symptom versus wellbeing repair. As all of the studies had relatively small sample sizes, there are wide confidence intervals around the effect sizes reported above. Some of the studies used interviewer scales to assess symptoms and self-report scales to assess wellbeing, which may influence the size of the effects observed.

There is also a parallel body of work looking at how well CBT enhances quality of life (QoL), which has significant overlap with wellbeing as a concept. A recent meta-analysis examined the impact of classic CBT and drug treatment for depression on symptom and QoL measures in 37 randomised controlled trials (Hofmann et al. 2017). This found a large pre-post effect on depression severity ($g = 1.30$; 95% CI 1.16–1.45). In contrast, there were only medium effects on QoL ($g = 0.69$; 95% CI 0.61–0.78). However, while QoL has clear overlap with wellbeing as a construct, it is not directly analogous (Salvador-Carulla et al. 2014). In particular, many QoL measures have a relatively narrow health/disease focus rather than a more holistic wellbeing focus, and so may miss broader benefits of treatment (e.g. see Al-Janabi et al. 2012).

Therefore, it is not possible to conclude to what extent classic CBT repairs wellbeing relative to symptoms on the basis of the extant literature and further research is needed. One efficient way to achieve this goal is to look at outcomes in routine clinical practice. The Improving Access to Psychological Therapy (IAPT) initiative in the UK presents a good way to achieve this, as high volumes of patients are given protocol adherent classic CBT for depression and anxiety and routine outcome data are collected. IAPT services run a stepped care model, where mild to moderate presentations are first treated with low intensity evidence-based approaches (for example, brief guided self-help) and are only 'stepped up' to higher intensity interventions like individual CBT if they fail to respond. Individuals with a more complex, recurrent or severe presentation can be offered high intensity as a first line treatment. Therefore, high intensity waiting lists are characterised by more severe, complex, and often treatment resistant presentations of depression and anxiety.

While wellbeing measures are not a standard part of the national IAPT outcome data set, they have been included as an additional outcome measure in some services. In particular, Somerset Partnership Foundation Trust (SPFT) supplemented the minimum IAPT data set with a wellbeing measure (the Warwick Edinburgh Mental Wellbeing Scale; WEMWBS, Tennant et al. 2007) between 2012 and 2017. The present study analyses wellbeing relative to symptom outcomes in this service, focusing on those individuals allocated to high intensity CBT treatment for either depression or anxiety (typically between 8 and 20 sessions of individual therapy). The primary aims are: (i) to evaluate to what extent 'high intensity' CBT repairs symptoms versus wellbeing;

and (ii) to assess the extent to which wellbeing and symptom deficits are 'normalised' by the end of treatment.

Some thought is required about how to operationalise change and recovery on wellbeing measures. When using symptom-focused measures, the objective is to eliminate symptoms and to ensure individuals fall under some cut-off that indicates recovery. For example, the depression outcome measure used in IAPT is the Patient Health Questionnaire (PHQ-9; Kroenke et al. 2001) and scoring nine or less is used to indicate remission. It is less clear cut what indicates sufficient repair of wellbeing. One approach is to examine where a patient falls in the general population distribution before and after treatment using large scale normative data. The WEMWBS has been well validated on the UK general population, with data collected on over 7000 individuals as part of the UK Health of the Nation Survey. Therefore, it is possible to express WEMWBS scores for individual IAPT clients in terms of where they sit in this general population distribution. Individuals scoring in the bottom third of the distribution can be viewed as 'languishing' and those scoring in the top third of the distribution can be viewed as 'flourishing'. Recovery can be defined as scoring in the average or flourishing parts of the general population distribution.

We predict that CBT will lead to a greater magnitude repair of symptoms than wellbeing (Hypothesis One) and that at the end of treatment more individuals will meet recovery criteria for symptoms than wellbeing (Hypothesis Two). We will additionally explore to what extent wellbeing and symptom measures are associated with one another and whether the extent of wellbeing and symptom repair is related to number of sessions attended. We had no a priori hypotheses for these exploratory analyses.

Method

Participants

Routine outcome data were collected on individuals being treated with high intensity CBT for depression and/or anxiety in the Somerset IAPT service between 2012 and 2017. Inclusion criteria for the IAPT service were being over 18 years of age and presenting with a primary problem of depression or anxiety that met IAPT clinical criteria at the point they were put on the waiting list (PHQ-9 $\geq$ 10 or GAD-7 $\geq$ 8). A subset of individuals no longer met these caseness criteria at the point they started treatment. Following national IAPT guidelines, participants were not offered treatment if they presented with comorbid psychosis, bipolar disorder; if drug and alcohol misuse was the primary presenting problem; if there was a moderate to severe impairment of cognitive function; and if they presented with a high level of risk to

self or others that could not be safely managed in the service context. We extracted from the database the subset of data for clients who were allocated to high intensity CBT. South West Frenchay Health Research Authority granted ethical approval for the study (15/SW/0352, IRAS ID 163179). As patient data were anonymised and could not be linked back to individuals, patient consent was not required to access the data.

Measures and Procedure

The Warwick-Edinburgh Mental Wellbeing Scale (WEM-WBS, Tennant et al. 2007) was used to measure wellbeing. Participants rate to what extent they have felt the way described in 14 wellbeing statements (e.g. "I've been feeling optimistic about the future") over the last 2 weeks, on a scale ranging from 1 (none of the time) to 5 (all of the time). Scores range from 14 (low wellbeing) to 70 (high wellbeing). Normative WEMWBS data are available for 7020 individuals in the UK general population as part of the Health Survey for England 2011 (Mean = 51.61, SD = 8.71). As far as we are aware, there are no normative data for a depressed population. The WEMWBS has been found in previous studies to have good internal reliability (Cronbach's $\alpha = 0.91$) and acceptable one-week test–retest reliability (intraclass correlation = 0.83) (Tennant et al. 2007; Stewart-Brown et al. 2009, 2011). In the present sample, internal reliability was also acceptable (intake $\alpha = .91$). There are no agreed reliable or clinically significant change criteria for the WEMWBS.

The Patient Health Questionnaire (PHQ-9; Kroenke et al. 2001) was used to measure depression symptom severity. Participants rate how many days over the past 2 weeks they have experienced the nine DSM-V symptoms of depression (e.g. "little interest or pleasure in doing things"), on a scale ranging from 0 (not at all) to 3 (nearly every day). Scores range from 0 (asymptomatic) to 27 (severely depressed), with scores of 5, 10, 15 and 20 representing mild, moderate, moderately severe and severe depression respectively. A cut off score of 10 has been found to be a good proxy for meeting diagnostic criteria for a major depressive episode as measured by structured clinical interview, with 88% specificity and 88% sensitivity (Kroenke et al. 2010). This is also the cut-off used in UK IAPT services to indicate a clinical presentation. Normative data are available on 5018 individuals in the general population from face-to-face household surveys conducted in Germany between 2003 and 2008 (Mean = 2.91, SD = 3.52) (Kocalevent et al. 2013). Studies find the PHQ-9 has good internal reliability ($\alpha = .89$) and test–retest reliability (intraclass correlation = .84) (Kroenke et al. 2001). In the present sample, internal reliability was also acceptable (intake $\alpha = .85$).

The Generalized Anxiety Disorder scale (GAD-7; Spitzer et al. 2006) was used to measure anxiety symptom severity. Participants rate how many days over the past 2 weeks they have experienced seven symptoms of anxiety (e.g. "feeling nervous, anxious, or on edge"), on a scale ranging from 0 (not at all) to 3 (nearly every day). Scores range from zero (asymptomatic) to 21 (severely anxious), with scores of 5, 10 and 15 representing mild, moderate and severe symptoms respectively. A cut off score of 10 was reported to have optimal sensitivity (89%) and specificity (82%) to confirm a diagnosis of generalized anxiety disorder based on structured clinical interview (although in UK IAPT services, a cut-off of ≥ 8 is used to indicate a clinically significant presentation). Normative data are available on 5030 individuals in the general population from a nationally representative face-to-face household survey conducted in Germany (Löwe et al. 2008) (Mean = 2.95, SD = 3.41). The GAD-7 has been found to have good internal reliability ($\alpha = .92$) and test–retest reliability (intraclass correlation = .83) (Spitzer et al. 2006). In the present sample, internal reliability was also acceptable (intake $\alpha = .85$).

The PHQ-9 and GAD-7 form part of the routine national data set administered prior to each session in IAPT, while the WEMWBS was only administered at first and last treatment session.

Analysis Plan

All analyses used two-tailed tests with an alpha of .05. The proportion of individuals scoring above clinical cut-offs for anxiety and depression at each time point was reported (using standard IAPT criteria of PHQ-9 scores ≥ 10 and GAD-7 scores ≥ 8). We also describe the proportion of individuals who were languishing, showing average wellbeing, and flourishing (scoring in the bottom third [< 47], middle third [47–57], and top third [> 57] of the general population distribution respectively). We used these cut-offs to determine the proportion of individuals at each time point who fell into each category in the Provencher and Keyes (2011) model of complete mental health. This is a two (symptomatic, asymptomatic in terms of anxiety and depression symptoms) by three (languishing, averaging, flourishing in terms of wellbeing) affective space. To allow a direct comparison of recovery rates for symptoms versus wellbeing, we compared the proportion of participants at each assessment point who met recovery criteria for symptoms (defined as scoring < 10 on the PHQ-9 *and* < 8 on the GAD-7) and wellbeing (defined as scoring in the average or flourishing part of general distribution; > 46) using McNemar tests for paired samples.

To analyse the extent of repair in each measure, we used two different analytic methods. If the same conclusions emerge across these different strategies, this

suggests they are likely to be robust. First, paired sample t-tests were run on pre- and post- scores. We reported Hedges g (and its 95% confidence interval) as a measure of effect size for each analysis. Second, the importance of tracking individual level as well as group level outcomes is increasingly realised (Guidi et al. 2018). Therefore, the proportion of individuals showing reliable improvement and reliable deterioration on each measure from first to last session was calculated (cf. Jacobson and Truax 1991). Reliable improvement/deterioration was defined as an improvement/deterioration of more than 1.96 times the standard error of difference for the scale. We used the standard deviation estimates from the present sample and estimates of test–retest reliability from scale validation studies in these analyses. The proportion of reliable improvement/deterioration on each measure was compared using a series of pairwise McNemar tests.

Association between intake scores, change scores, and number of sessions were analysed using Pearson's correlation coefficients (reporting simple r and attenuation corrected r).

There was a relatively high degree of missing WEMWBS data at the final session assessment, so analyses were run on both a complete case basis and a multiple imputation basis (to simulate missing values). This is because there is ongoing debate in the literature about how best to analyse data were there is a relatively high proportion of missing data, where there is reason to think data may be 'missing not at random' (MNAR), and where there is a limited pool of auxiliary variables to use to predict missing values (Jakobsen et al. 2017; Madley-Dowd et al. 2019; van Ginkel et al. 2019). If an identical pattern of findings emerges with both analytic approaches, this suggests that the bias inherent in either method is unlikely to have substantially contaminated the results. Imputation was conducted using a Markov Chain Monte Carlo (MCMC) algorithm, with 70 imputation runs (based on guidance that the number of imputations should exceed the percentage of missing data; White et al. 2010). All variables used in subsequent analysis models (intake and final-session WEMWBS, PHQ-9, and GAD-7) as well as auxiliary variables that might predict variables with missing data were included in the imputation model. Auxiliary variables were age, gender, ethnicity, sessions attended, intake score on the IAPT phobia scale, intake score on the Work and Social Adjustment Scale (WSAS; Mundt et al. 2002), and intake score on the Standard Assessment of Personality abbreviated scale (SAPAS; Moran et al. 2003). Multiple imputation analyses used pooled data across these 70 imputations. The imputed data set can be viewed as an intention-to-treat sample.

Table 1 Demographic and clinical characteristics of individuals with and without post WEMWBS

	With post treatment WEMWBS (n=618)	Without post treatment WEMWBS (n=980)
Age	39.38 (13.8)	37.64 (13.7)
Gender	67% female	70% female
Session 1 PHQ-9	15.54 (6.07)	16.47 (6.13)
Session 1 GAD-7	14.03 (4.90)	14.51 (4.98)
Session 1 WEMWBS	33.80 (9.35)	32.23 (9.19)
Sessions completed	8.98 (3.47)	6.38 (3.87)
Δ PHQ-9	−7.78 (6.20)	−4.57 (6.08)
Δ GAD-7	−7.02 (5.73)	−4.19 (5.63)

Data are mean (SD) values

Results

Data Completeness

Intake WEMWBS, PHQ-9 and GAD-7 data were available for 1854 participants. Of these individuals, only 618 (39%) had final session data on all three measures. Complete case change and final session analyses were conducted on the subset of 618 individuals with complete data at final session; all other analyses were implemented on the entire sample.

Independent sample t-tests were run comparing intake demographic and clinical characteristics and symptom change score in individuals with and without complete data (see Table 1). At intake, there were small but nevertheless statistically significant differences between these groups. Those with complete WEMWBS data were significantly older, t(1596)=2.45, p=.01; had significantly lower PHQ-9 scores, t(1596)=2,95, p<.001, had significantly higher WEMWBS scores, t(1596)=3.29, p<.001, and tended to have lower GAD-7 scores at the level of a non-significant trend, t(1596)=1.87, p=.06, compared to those without complete WEMWBS data. There were no significant gender differences, $\chi^2<1$. There were significant and slightly greater magnitude differences in treatment responsiveness between the groups also. Those with complete WEMWBS data showed a greater improvement during treatment in PHQ-9 depression, t(1579)=10.16, p<.001, and GAD-7 anxiety, t(1579)=9.70, p<.001, than those without complete WEMWBS data. Those with complete WEMWBS data also attended a greater number of treatment sessions, t(1596)=13.59, p<.001, than those without complete WEMWBS data. In summary, participants with WEMWBS data are a subgroup of individuals who are slightly older, less depressed and anxious at intake; attend more sessions; and are more treatment responsive.

Table 2 Depression, anxiety and wellbeing at intake and final treatment session (n = 618 with complete data, 1854 with imputed data)

	Intake session	Final session (complete case)	Final session (imputed)
PHQ-9	15.54 (6.07)	7.76 (6.12)	10.96 (7.22)
GAD-7	14.03 (4.90)	7.00 (5.52)	9.57 (6.23)
WEMWBS	33.80 (9.35)	44.75 (10.91)	40.04 (11.93)

Data are mean (SD) values

Wellbeing and Symptom Levels at Assessment

Table 2 reports intake and final session wellbeing, depression, and anxiety scores for participants (including both complete case and imputed estimates for final session data). The sample scored below the wellbeing general population levels (mean score of 33 compared to general population average of 51). 1707/1854 (92%) were classified as languishing; 125/1854 (7%) were classified as having adequate mental health; and 22/1854 (1%) were classified as flourishing (8% in total meeting recovery criteria for WEMWBS). Similarly, the sample had elevated levels of depression and anxiety at intake. 1724 participants scored above IAPT cut-offs for either GAD-7 ($\geq$ 8) or PHQ-9 ($\geq$ 10), while 130 met symptom recovery criteria. 622 participants (33%) were in the severe range for depression (PHQ-9 > 19) and 1020 participants (55%) were in the severe range for anxiety (GAD-7 > 14). The proportion of individuals meeting recovery criteria at intake did not significantly differ for symptoms versus wellbeing, McNemar test p = .178.

Wellbeing and Symptom Change During High Intensity CBT

Paired sample t-tests showed a significant WEMWBS improvement (average increase of 10.94 points [SD = 10.99] for complete case data; average increase of 7.12 points [SD = 10.46] for imputed data), complete case paired-sample t(617) = 24.75, p < .001, imputed paired-sample t(1853) = 23.926, p < .001. There was also a significant reduction in depression symptoms (mean drop of 7.78 points [SD = 6.20] for complete case data; mean drop of 5.26 points [SD = 6.25] for imputed data), complete case t(617) = 31.21, p < .001, imputed t(1853) = 36.07, p < .001. Similarly, there was a significant reduction in anxiety symptoms (mean drop of 7.02 points [SD = 5.73] for complete case data; mean drop of 4.77 points [SD = 5.79] for imputed data), complete case t(617) = 30.48, p < .001, imputed t(1853) = 35.38, p < .001. Figure 1 plots the effect size (and 95% confidence interval) for change in each outcome measure for complete case data and imputed data. All effects were of a large magnitude (g > 0.8) for complete case data and of a large or medium effect size (g > 0.5 < 0.8) for imputed data according to rules of thumb (Cohen 1988). The anxiety and depression effects
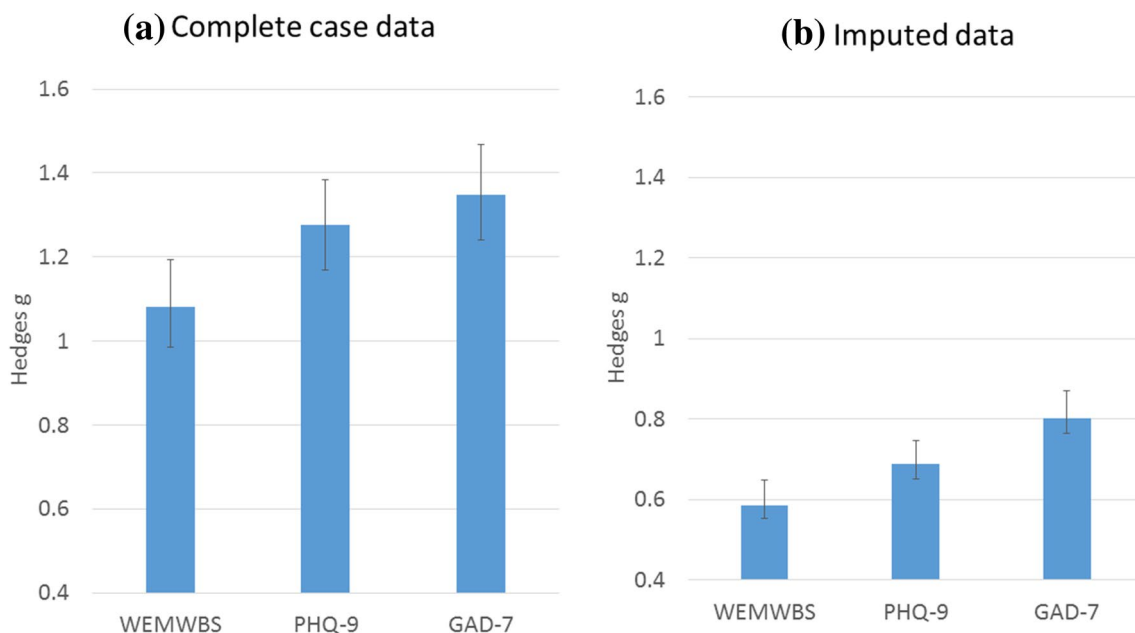

Fig. 1 Pre-post effect sizes for each measure on complete data (n = 618; **a**) and imputed data (n = 1854; **b**). *Note*: Data are mean (95% confidence interval) values

sizes were numerically greater than the wellbeing effect size in both the complete case and the imputed analyses.

Reliable improvement was observed for 291 individuals (48%) for wellbeing, 339 individuals (55%) for depression and 364 individuals (59%) for anxiety in complete case analyses. A greater proportion of individuals showed reliable improvement in depression relative to wellbeing, McNemar, p < .001, anxiety relative to wellbeing, McNemar p < .001, and anxiety relative to depression, McNemar p = .046. Reliable deterioration was observed for 9 individuals (1%) for the WEMWBS, 8 individuals (1%) for depression, and 4 individuals (< 1%) for anxiety. There were no significant differences in the rates of deterioration for each outcome measure, McNemar ps > .266.

When using imputed data, reliable improvement was seen for 622.6 individuals (34%) for wellbeing, 727.1 individuals (39%) for depression, and 777.1 individuals (42%) for anxiety. More individuals improved for depression than wellbeing, McNemar p < .001, for anxiety than wellbeing, McNemar p < .001, and for anxiety than depression, McNemar p = .013. Reliable deterioration was seen for 52.4 individuals (3%) for wellbeing, for 44.3 individuals (2%) for depression, and for 33.7 individuals (2%) for anxiety. There were no significant differences in the rates of deterioration for wellbeing and depression, McNemar p = .500, anxiety and depression, McNemar p = .206, or wellbeing and anxiety, McNemar p = .066.

Wellbeing and Symptom Levels at Final Treatment Session

At the final treatment session, on average the sample continued to score below the wellbeing general population average (complete case mean score = 45; imputed mean score = 40). Using complete case data, 339 individuals (55%) were languishing, 191 individuals (31%) were in the average range, and only 88 individuals (14%) were flourishing. Using scoring in the average or flourishing range as a proxy for recovery, in total 44% had recovered on the WEMWBS. Using multiple imputation data, 1263.1 individuals (68%) were languishing, 429.6 individuals (23%) were in the average range, and 161.3 individuals (9%) were flourishing, with 32% in total in recovery.

In terms of symptom outcomes, using complete case data 242 individuals (40%) fell in the clinical range on one or both of the PHQ-9 and GAD-7, while the remaining 376 individuals (60%) scored beneath caseness on both measures (the IAPT definition of recovery, Clark and Oates 2014) at the final treatment session. In total, 206 individuals failed to recover on either symptoms or wellbeing, 243 individuals recovered on both symptoms and wellbeing, 36 individuals recovered just on wellbeing, and 133 individuals recovered just on symptoms. A significantly greater proportion of

individuals met recovery criteria for symptoms relative to wellbeing, McNemar p < .001.

Using imputed data, 1104.6 individuals (60%) fell in the clinical range on one or both symptom measures, while 749.4 individuals (40%) scored beneath caseness (i.e. had recovered in IAPT terms) at the final treatment session. 975.6 individuals failed to recover on either symptoms or wellbeing, 287.5 individuals recovered on symptoms but not wellbeing, 129.0 individuals recovered on wellbeing but not symptoms, and 461.9 individuals recovered on both symptoms and wellbeing. A greater proportion of individuals recovered for symptoms than wellbeing, McNemar test, p < .001.

Finally, we considered the percentage of participants falling in each space of the Provencher and Keyes model at final treatment session. Only 13% were in the optimal space of asymptomatic and flourishing (8% using imputed data). 26% were asymptomatic and had average wellbeing (17% using imputed data); 22% were asymptomatic and were languishing (15% using imputed data); < 1% were flourishing but symptomatic (< 1% using imputed data); 5% were symptomatic and had average wellbeing (6% using imputed data); and 33% were symptomatic and languishing (53% using imputed data).[3]

Exploratory Association Analyses

Greater intake WEMWBS was significantly associated with lower depression (Pearsons's r = − .645, p < .001; correcting for attenuation r = − .739) and anxiety (r = − .515, p < .001; correcting for attenuation r = − .602). Greater anxiety was also significantly associated with greater depression (r = .661, p < .001; correcting for attenuation r = .752). Greater increase in wellbeing over the course of treatment was significantly associated with a greater drop in depression (Pearson's r on complete case data = − .604, p < .001, correcting for attenuation r = − .682, using imputed data

---

[3] The analyses comparing the magnitude of recovery for symptoms versus wellbeing are relying on essentially arbitrary definitions of recovery, particularly for the WEMWBS where no cut-off has yet been established. Another criteria that has been used in the literature is to define anyone who scores no more than one SD below the general population average as being in recovery (which results in a lower cut-off score of 44 for the WEMWBS). Using this approach, at intake a greater proportion of individuals were in recovery for WEMWBS relative to symptoms, McNemar p<.001. At post-treatment, a greater proportion of individuals met recovery for symptoms relative to wellbeing, McNemar, p=.002. We also conducted additional continuous analyses on change on each measure in Z-score units (relative to general population averages) and proportion change units (relative to scale range). These continuous analyses do not rely on arbitrary cut-offs to define recovery. These reached an identical conclusion to the categorical analyses, with greater repair of symptoms relative to wellbeing (see supplementary materials for details).

r = − .631, p < .001) and anxiety (complete case r = − .599, p < .001; correcting for attenuation r = − .671; imputed data r = − .613, p < .001). Greater depression repair was significantly associated with greater anxiety repair (complete case r = .713, p < .001; correcting for attenuation r = .807; imputed data r = .749, p < .001). Attending more sessions was significantly linked to greater PHQ-9 repair (complete case r = − .130, p = .001; imputed data r = − .095, p < .001) and GAD-7 repair (complete case r = − .090, p = .025; imputed data r = − .095, p < .001) but was not significantly linked to wellbeing improvements (complete case r = .038, p = .344; imputed data r = .024, p = .440).

Discussion

This study investigated the extent to which routinely delivered classic CBT repairs symptoms versus enhances wellbeing in a sample of individuals with anxiety and depression (using registry data from a high intensity UK IAPT service). At intake, there were marked deficits for both wellbeing and symptoms, with a vast majority of individuals showing clinically significant levels of anxiety (GAD-7 > 7) or depression (PHQ-9 > 9) and languishing in wellbeing terms (falling in the bottom third of the general population distribution). Participants on average attended around 9 treatment sessions and there was a significant increase in wellbeing and decrease in depression and anxiety symptoms between first and last treatment session.

Two different analytic techniques showed CBT had a larger effect on symptoms than wellbeing (supporting Hypothesis One). Pre-post effect sizes were numerically larger and a greater proportion of individuals showed reliable improvement of symptoms relative to wellbeing. While these two analytic methods both have their strengths and limitations, all reach the same conclusion. Unsurprisingly given the smaller magnitude of wellbeing relative to symptom repair, at the final treatment session a smaller proportion of the sample met recovery criteria on the wellbeing (no longer falling in the languishing third of the general population distribution) than met recovery criteria for symptoms (no longer showing clinically significant levels of depression and anxiety). Only a very small number of individuals were flourishing at the end of treatment. Therefore, Hypothesis Two was also supported. The same pattern of results emerged in both complete case analyses and when using multiple imputation to simulate missing data. Overall outcomes were superior in the complete-case analyses relative to the imputed data set on all variables (large versus medium pre-post effect sizes), but critically the *relative* magnitude of symptom versus wellbeing repair was the same in both sets of analyses.

This result broadly mirrors findings in previous trials that have included classic CBT as a comparator for novel positive psychology or wellbeing interventions (Chaves et al. 2017; Geschwind et al. 2019; Fava et al. 1998a; Fava et al. 2005) and also is consistent with meta-analytic findings that CBT and drug treatments for depression have a bigger effect on symptoms than quality of life (Hofmann et al. 2017). It extends them by directly comparing wellbeing and symptom deficits in a large clinical sample, giving greater precision to the estimates observed.

We also explored whether number of sessions predicted wellbeing and symptom repair. Attending a greater number of sessions attended was not significantly associated with wellbeing repair, despite being robustly associated with greater anxiety and depression repair. Therefore, it is unlikely to be sufficient to enhance wellbeing outcomes simply by offering a longer treatment dose of existing CBT protocols (although of course correlational data of this kind cannot be used to test causal claims).

Given the present results show CBT is less effective at building wellbeing than reducing depression (and that increasing number of sessions alone will not resolve this issue), it is important to consider alternative treatment approaches. One way forward is to adapt the delivery of classic CT so that it more explicitly addresses positive affect and wellbeing deficits (see Dunn, in press). This could include a greater focus on identifying values to help clients build meaning, more systematic targeting of underlying mechanisms that block positive affect (e.g. dampening: Burr et al. 2017; experiential processing; Gadeikis et al. 2017), and incorporation of a more explicit recovery focus (see Medalia et al. 2019).

A second way forward way could be to evaluate the potency of other established acute mood disorder treatments in repairing wellbeing, for example behavioural activation (BA; see Mazzucchelli et al. 2010). However, given there is significant content overlap between the activity scheduling aspects of CBT and BA protocols, it seems unlikely that BA will lead to significantly enhanced wellbeing outcomes relative to CBT. There may be greater potential in 'third wave' cognitive treatments like Acceptance and Commitment Therapy (ACT; Hayes et al. 2012). However, what little evidence exists suggests that ACT may not optimally repair wellbeing. For example, a recent secondary analysis of an RCT found that a brief online version of ACT is also less effective at repairing wellbeing relative to symptoms (Trompetter et al. 2017).

A third approach could be to offer staged treatments, where classic CBT is used to treat symptoms and then a bespoke wellbeing therapy is offered afterwards as a second step treatment (for example, Wellbeing Therapy; Fava 2016). There is good clinical trial evidence that Wellbeing Therapy helps prevent relapse in depression (Fava et al.

1998b; Fava et al. 2004; Stangier et al. 2013), but wellbeing outcomes were not reported in these trials unfortunately. In two pilot trials that did include wellbeing measures, mixed results have emerged. Fava et al. (2005) showed superior wellbeing outcomes for Wellbeing Therapy relative to CBT in treating acute generalized anxiety disorder. The Wellbeing Therapy arm generated large effect size pre-post wellbeing improvement, whereas the CBT arm only led to small effect size improvement. However, Fava et al. (1998a) found no superiority of Wellbeing Therapy relative to CBT in treating residual affective disorders (small wellbeing pre-post effect sizes observed in each arm). Alternatives to Wellbeing Therapy as staging treatments are also starting to emerge. For example, Geschwind et al. (2019) demonstrate a large pre-post wellbeing effect size for CBT followed by a positive CBT protocol (albeit this wellbeing improvement was smaller than that observed for symptom relief).

Finally, novel acute therapies could be delivered that explicitly target wellbeing from the outset. Application of positive psychology approaches in clinical populations have had limited success in repairing wellbeing. For example, Chaves et al. (2017) found no significant difference between group CBT and a group positive psychology intervention for acute depression, with both producing small to medium pre-post effect sizes. However, other treatments are being developed to target positive valence system deficits that have encouraging preliminary results on positive affect and wellbeing (e.g. Taylor et al. 2017; Positive Affect Treatment, Craske et al. 2019; Augmented Depression Therapy, Dunn et al. 2019). These novel therapies now require more robust evaluation in definitive trials in diagnosed depressed and anxious populations.

The findings also open opportunities for personalised approaches to care, whereby treatment selection could be tailored depending on the type of deficits the client presents with. For example, clients with particularly marked wellbeing deficits may benefit from approaches explicitly targeting positive outcomes, whereas clients presenting with particularly marked deficits in symptoms may benefit from existing approaches such as CBT. Emerging analytical techniques have the potential to match patients to the most effective treatments in this way (e.g. The Personalised Advantage Index; DeRubeis et al. 2014).

We provide an estimate of wellbeing deficits at intake as measured by WEMWBS in a real world clinical sample (a high intensity IAPT population). Our results suggest wellbeing levels in this high intensity IAPT sample are over two standard deviations on average below general population averages. This may be useful for benchmarking purposes for future work (although see the caveats raised below about generalisability of the present sample).

The present data speak to the debate about whether wellbeing and symptoms should be viewed as at two ends of a single dimension or as orthogonal constructs. Wellbeing showed a strong (but not perfect) negative association with anxiety and depression symptoms at intake, which became even stronger when correcting for attenuation. Similarly, there was a strong negative association between wellbeing change and symptom change during treatment, which became more marked when attenuation was corrected for. This is more consistent with the single dimension account (Huppert 2014). This deviates from other studies showing that wellbeing and symptom repair are only weakly correlated (for example, Trompetter et al. 2017). It may be that this discrepancy in part reflects the choice of wellbeing measure used. We deployed the WEMWBS to measure wellbeing, which has a high degree of content overlap with symptom measures (particularly with the PHQ-9). The WEMWBS can be critiqued as simply a positively rather than negatively framed measure of the same underlying symptom features tapped by the PHQ-9. Bifactor analyses have not shown the WEMWBS to be distinct from symptom measures in a large community sample (Böhnke and Croudace 2016). In contrast, Trompetter et al. (2017) used the Mental Health Continuum short-form, which may be a purer measure of wellbeing that is more distinct from psychopathology.

Even if wellbeing and symptoms are to some extent measuring the same underlying construct, it is still beneficial to measure them separately. Goal setting theory argues objectives are more likely to be achieved if they are couched in approach rather than avoidance terms (Elliot et al. 1997; Roskes et al. 2014). Reducing levels of depression and anxiety is an avoidance goal, whereas enhancing levels of wellbeing is an approach goal. Moreover, symptom measures may be relatively insensitive to the upper half of a single mental health continuum (measuring movement from negative to neutral mental health but not from neutral to positive mental health). It is noteworthy that of the individuals who met recovery criteria for symptoms, many were still languishing and very few were flourishing in wellbeing terms.

Limitations of the present study need to be held in mind. First, and most critically, there was a substantial degree of missing WEMWBS data at follow-up and these data were potentially 'missing not at random' (MNAR). Those with complete data were older, had less marked symptoms and higher wellbeing, were more treatment responsive, and attended more sessions. These differences may reflect client willingness to fill in extra measures as a function of whether they felt they benefitted, therapist implicit biases in who they gave post-treatment WEMWBS to, or the fact that WEMWBS was only given at planned discharge (where response rates are likely to be superior). These characteristics of the completer sample may be biasing results, in particular artificially inflating effect sizes on all outcome variables. Partially offsetting this concern, it is reassuring

that an identical pattern of findings emerged when analysing the imputed dataset (effectively an intention-to-treat sample). There is ongoing debate about the reliability of multiple imputation with a high degree of missing data, when the data may be MNAR and when only a limited set of auxiliary variables are available to predict missing values (van Ginkel et al. 2019; Jakobsen et al. 2017; Madley-Dowd et al. 2019). Nevertheless, multiple imputation may provide less biased results than list wise deletion when data are MNAR. Overall, while it is encouraging that the same pattern of findings (greater repair of symptoms relative to wellbeing) emerged in both complete-case and multiple imputation-based analyses, neither method are free from bias when data are MNAR. Therefore, caution needs to be taken when generalising these findings to other samples.

Second, this is an uncontrolled study using routine registry data, meaning that factors such as spontaneous recovery over time and treatment expectancy are not accounted for. Moreover, we only had the demographic and clinical information that is routinely collected in IAPT services (which is less extensive than that which is typically captured in research studies). This limits the extent to which detailed comparisons can be made to other research samples and also precluded moderation analyses being conducted. Future research should examine which baseline variables predict wellbeing repair during CBT. Third, the sample is also not entirely representative of IAPT services, as it was part of a personality disorder demonstration site rather than routine IAPT care. This meant that the service had latitude to take on clients with a greater degree of comorbid personality pathology than is typically deemed eligible for IAPT and also had more flexibility about the number of sessions they could offer to clients. This should be taken into account when using this data to estimate how well high intensity CBT in more routine IAPT settings will perform in repairing wellbeing (given that previous work suggests those with more marked personality disorder features do less well in IAPT generally; Goddard et al. 2015).

Fourth, it is important to recognise that the pattern of results regarding magnitude of change of wellbeing versus symptoms do not necessarily generalise beyond the specific measures used (the PHQ-9, GAD-7, and WEMWBS). It may simply be the case that the measures are differentially sensitive to change, with the PHQ-9 and GAD-7 being more sensitive than the WEMWBS. The findings require replication using a broader array of outcome measures. Fifth, the sample was selected on the basis of scoring above cut-offs on the PHQ-9 and GAD-7, reflecting standard practice in IAPT settings. While these scores have been shown to predict diagnostic status with a reasonable degree of accuracy, it would have been methodologically stronger to have diagnostically interviewed participants. Sixth, greater repair of symptoms relative to wellbeing could simply reflect regression to the mean, in that the sample was further away from the general population average scores for symptoms than wellbeing at intake. Seventh, definitions of positive recovery focus on optimising functioning (i.e. being able to perform in key life areas like vocation, family, relationships, and hobbies) as well as wellbeing. We did not consider functioning outcomes here. Eighth, reflecting the way in which the GAD-7 is used in IAPT services, we are equating changes in this scale with changes in overall anxiety. However, the GAD-7 was originally designed as a screening tool and severity indicator for generalized anxiety disorder specifically rather than anxiety more broadly. Finally, we did not have accurate data on the use of psychotropic medications, so possible interactions between drug and psychological therapy could not be examined.

In conclusion, notwithstanding the above caveats, our findings suggest that classic CBT does a better job of repairing symptoms than building wellbeing. There is a need to enhance existing treatments, and potentially to develop novel treatments, that better target wellbeing enhancement in addition to symptom relief.

Compliance with Ethical Standards

References

Ali, S., Rhodes, L., Moreea, O., McMillan, D., Gilbody, S., Leach, C., et al. (2017). How durable is the effect of low intensity CBT for depression and anxiety? Remission and relapse in a longitudinal cohort study. *Behaviour Research and Therapy, 94,* 1–8.

Al-Janabi, H., Flynn, T. N., & Coast, J. (2012). Development of a self-report measure of capability wellbeing for adults: The ICECAP-A. *Quality of Life Research, 21,* 167–176.

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

Böhnke, J. R., & Croudace, T. J. (2016). Calibrating well-being, quality of life and common mental disorder items: Psychometric epidemiology in public mental health research. *The British Journal of Psychiatry, 209,* 162–168.

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.

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.

Clark, D. A., & Beck, A. T. (2010). *Cognitive therapy of anxiety disorders: Science and practice*. New York: Guilford Press.

Clark, D., & Oates, M. (2014). Improving access to psychological therapies: Measuring improvement and recovery in adult services, version 2. Accessed, 21 June 16

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Coulombe, S., Radziszewski, S., Meunier, S., Provencher, H., Hudon, C., Roberge, P., et al. (2016). Profiles of recovery from mood and anxiety disorders: A person-centered exploration of people's engagement in self-management. *Frontiers in Psychology, 7,* 584.

Craske, M., Meuret, A. E., Ritz, T., Treanor, M., Dour, H., & Rosenfield, D. (2019). Positive affect treatment for depression and anxiety. A randomized controlled trial for a core feature of anhedonia. *Journal of Consulting and Clinical Psychology, 87,* 457–541.

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.

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.

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: Translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS ONE, 9,* e83875.

Dunn, B. D. (in press). Augmenting cognitive behavioural therapy to build positive mood in depression. In J. Gruber (Ed.), *Oxford handbook of positive emotion and psychopathology*. New York: Oxford University Press.

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.* https://doi.org/10.1016/j.brat.2019.103418.

Elliot, A. J., Sheldon, K. M., & Church, M. A. (1997). Avoidance personal goals and subjective well-being. *Personality and Social Psychology Bulletin, 23,* 915–927.

Fava, G. A. (2016). Well-being therapy: Current indications and emerging perspectives. *Psychotherapy and Psychosomatics, 85,* 136–145.

Fava, G. A., Rafanelli, C., Cazzaro, M., Conti, S., & Grandi, S. (1998a). Well-being therapy. A novel psychotherapeutic approach for residual symptoms of affective disorders. *Psychological Medicine, 28,* 475–480.

Fava, G. A., Rafanelli, C., Grandi, S., et al. (1998b). Prevention of recurrent depression with cognitive behavioural therapy: Preliminary findings. *Archives of General Psychiatry, 55,* 816–820.

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,* 1872–1876.

Fava, G. A., Ruini, C., Rafanelli, C., Finos, L., Salmaso, L., Mangelli, L., et al. (2005). Well-being therapy of generalized anxiety disorder. *Psychotherapy and Psychosomatics, 74,* 26–30.

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

Garland, E. L., Fredrickson, B., Kring, A. M., Johnson, D. P., Meyer, P. S., & Penn, D. L. (2010). Upward spirals of positive emotions counter downward spirals of negativity: Insights from the broaden-and-build theory and affective neuroscience on the treatment of emotion dysfunctions and deficits in psychopathology. *Clinical Psychology Review, 30,* 849–864.

Geschwind, N., Arntz, A., Bannink, F., & Peeters, F. (2019). Positive cognitive behaviour therapy in the treatment of depression: A randomized order within-subject comparison with traditional cognitive behaviour therapy. *Behaviour Research and Therapy, 116,* 119–130.

Goddard, E., Wingrove, J., & Moran, P. (2015). The impact of comorbid personality difficulties in response to IAPT treatment for depression and anxiety. *Behaviour Research and Therapy, 73,* 1–7.

Guidi, J., Brakemeier, B., Bockting, C. L. H., Fava, G. A., et al. (2018). Methodological recommendations for trials of psychological interventions. *Psychotherapy and Psychosomatics, 87,* 276–284.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2012). *Acceptance and commitment therapy: The process and practice of mindful change*. New York: Guilford Press.

Hofmann, S. G., Curtiss, J., Carpenter, J. K., & Kind, S. (2017). Effect of treatments for depression on quality of life: A meta-analysis. *Cognitive and Behaviour Therapy, 46,* 265–286.

Hofmann, S. G., & Smits, J. A. (2008). Cognitive-behavioral therapy for adult anxiety disorders: a meta-analysis of randomized placebo-controlled trials. *The Journal of Clinical Psychiatry, 69,* 621–632.

Huppert, F. A. (2014). The state of wellbeing science: Concepts, measures, interventions and policies. In G. Cooper (Ed.), *Introduction to wellbeing: A complete reference guide* (Vol. 6). New Jersey: Wiley.

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.

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials—A practical guide with flowcharts. *BMC Medical Research Methodology, 17,* 162.

Kessler, R. C., Angermeyer, M., Anthony, J. C., De Graaf, R. O. N., Demyttenaere, K., Gasquet, I., et al. (2007). Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry, 6,* 168–176.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62,* 593–602.

Keyes, C. L. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Social Behavior, 43,* 207–222.

Keyes, C. L. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology, 73,* 539–548.

Keyes, C. L., & Lopez, S. J. (2002). *Toward a science of mental health: Positive directions in diagnosis and intervention*. New York: Oxford University Press.

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.

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.

Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2010). The patient health questionnaire somatic, anxiety, and depressive symptom scales: A systematic review. *General Hospital Psychiatry, 32,* 345–359.

Lambert, M. J. (2017). Maximizing psychotherapy outcomes beyond evidence-based medicine. *Psychotherapy and Psychosomatics, 86,* 80–89.

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., et al. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care, 46,* 266–274.

Madley-Dowd, P., Hughes, R., Tilling, K., & Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology, 110,* 63–73.

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.

Medalia, A., Beck, A. T., & Grant, P. M. (2019). Cognitive therapies for psychosis: Advances and challenges. *Schizophrenia Research, 203,* 1–2.

Moran, P., Leese, M., Lee, T., Walters, P., Thornicroft, G., & Mann, A. (2003). Standardised assessment of personality—Abbreviated scale (SAPAS): Preliminary validation of a brief screen for personality disorder. *British Journal of Psychiatry, 183,* 228–232.

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

Provencher, H. L., & Keyes, C. L. (2011). Complete mental health recovery: Bridging mental illness with positive mental health. *Journal of Public Mental Health, 10,* 57–69.

Roskes, M., Elliot, A. J., & De Dreu, C. K. W. (2014). Why is avoidance motivation problematic, and what can be done about it? *Current Directions in Psychological Science, 23,* 133–138.

Salvador-Carulla, L., Lucas, R., Ayuso-Mateos, J. L., & Miret, M. (2014). Use of the terms "wellbeing" and "quality of life" in health sciences: A conceptual framework. *European Journal of Psychiatry, 28,* 50–65.

Slade, M. (2010). Mental illness and well-being: The central importance of positive psychology and recovery approaches. *BMC Health Services Research, 10,* 26.

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

Stangier, U., Hilling, C., Heidenreich, T., Risch, A. K., Barocka, A., Schlösser, R., et al. (2013). Maintenance cognitive-behavioural therapy and manualized psychoeducation in the treatment of recurrent depression: A multicenter prospective randomized controlled trial. *American Journal of Psychiatry, 170,* 624–632.

Stewart-Brown, S. L., Platt, S., Tennant, A., Maheswaran, H., Parkinson, J., Weich, S., et al. (2011). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): A valid and reliable tool for measuring mental well-being in diverse populations and projects. *Journal of Epidemiology and Community Health, 65*(Supplement 2), A38–A39.

Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009). Internal construct validity of the Warwick-Edinburgh mental well-being scale (WEMWBS): A Rasch analysis using data from the Scottish health education population survey. *Health and Quality of Life Outcomes, 7,* 15.

Taylor, C. T., Lyubomirsky, S., & Stein, M. B. (2017). Upregulating the positive affect system in anxiety and depression: Outcomes of a positive activity intervention. *Depression and Anxiety, 34,* 267–280.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., et al. (2007). The Warwick-Edinburgh mental well-being scale (WEMWBS): Development and UK validation. *Health and Quality of life Outcomes, 5,* 63.

Trompetter, H. R., Lamers, S. M. A., Westerhof, G. J., Fledderus, M., & Bohlmeijer, E. T. (2017). Both positive mental health and psychopathology should be monitored in psychotherapy: Confirmation for the dual-factor model in acceptance and commitment therapy. *Behaviour Research and Therapy, 91,* 58–63.

Tudor, K. (1996). *Mental health promotion: Paradigms and practice*. East Sussex: Routledge.

Van Ginkel, J. R., Linting, M., Rippe, R. C. A., & van der Voort, A. (2019). Rebutting existing misconceptions about multiple imputation as a method for handling missing data. *Journal of Personality Assessment*. https://doi.org/10.1080/00223891.2018.1530680.

White, I. R., Royston, P., & Wood, A. M. (2010). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine, 30,* 377–399.

Wood, A. M., & Joseph, S. (2010). The absence of positive psychological (eudemonic) well-being as a risk factor for depression: A ten year cohort study. *Journal of Affective Disorders, 122,* 213–217.

World Health Organization. (2017). *Depression and other common mental disorders: Global health estimates*. Geneva: 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,* 148–150.