

Original Article

Cite this article: Bauer-Staeb C, Griffith E, Faraway JJ, Button KS (2023). Trajectories of depression and generalised anxiety symptoms over the course of cognitive behaviour therapy in primary care: an observational, retrospective cohort. *Psychological Medicine* 53, 4648–4656. <https://doi.org/10.1017/S0033291722001556>

Received: 3 June 2021
Revised: 18 March 2022
Accepted: 10 May 2022
First published online: 16 June 2022

Key words:

CBT; depression; generalised anxiety; primary care; psychotherapy; symptom trajectories

Author for correspondence:

Katherine S. Button,
E-mail: k.s.button@bath.ac.uk

Trajectories of depression and generalised anxiety symptoms over the course of cognitive behaviour therapy in primary care: an observational, retrospective cohort

Clarissa Bauer-Staeb¹, Emma Griffith^{1,2}, Julian J. Faraway³
and Katherine S. Button¹ 

¹Department of Psychology, University of Bath, Bath, UK; ²Avon and Wiltshire NHS Mental Health Partnership Trust, Bath, UK and ³Department of Mathematical Sciences, University of Bath, Bath, UK

Abstract

Background. Cognitive-behavioural therapy (CBT) has been shown to be an effective treatment for depression and anxiety. However, most research has focused on the sum scores of symptoms. Relatively little is known about how individual symptoms respond.

Methods. Longitudinal models were used to explore how depression and generalised anxiety symptoms behave over the course of CBT in a retrospective, observational cohort of patients from primary care settings ($n = 5306$). Logistic mixed models were used to examine the probability of being symptom-free across CBT appointments, using the 9-item Patient Health Questionnaire and the 7-item Generalised Anxiety Disorder scale as measures.

Results. All symptoms improve across CBT treatment. The results suggest that *low mood/hopelessness and guilt/worthlessness* improved quickest relative to other depressive symptoms, *with sleeping problems, appetite changes, and psychomotor retardation/agitation* improving relatively slower. *Uncontrollable worry* and *too much worry* were the anxiety symptoms that improved fastest; *irritability* and *restlessness* improved the slowest.

Conclusions. This research suggests there is a benefit to examining symptoms rather than sum scores alone. Investigations of symptoms provide the potential for precision psychiatry and may explain some of the heterogeneity observed in clinical outcomes when only sum scores are considered.

Introduction

Depression and anxiety are leading causes of disability worldwide (GBD 2019 Mental Disorders Collaborators, 2022). Despite effective interventions existing, the current treatment success leaves room for improvement (Cuijpers, van Straten, Bohlmeijer, Hollon, & Andersson, 2010). Much of the research efforts evaluating treatments to date have focused on sum scores. When sum scores are used, individual symptoms are assumed to be equivalent in value as common indicators of a disorder. Subsequently, individual symptoms have received comparably little attention (Fried & Nesse, 2015a). However, there may be important insights to be gained from better understanding individual symptoms and how they respond over the course of treatment (Fried & Nesse, 2015a).

Depression and anxiety are heterogeneous disorders. For example, the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) lists nine depressive symptoms (APA, 2013). In order to meet the criteria for a diagnosis of depression, patients have to meet at least five of the nine symptom criteria, including at least one of the core symptoms – depressed mood or diminished interest/pleasure (APA, 2013). These criteria include compounded symptoms that are grouped together such as worthlessness or inappropriate guilt as well as symptoms that lie on the opposite spectrum such as psychomotor agitation or retardation. Given the multitude of possible symptom combinations for depression alone, it is perhaps unsurprising that considerable symptom profile variability has been reported with few exact symptom combinations being shared between patients (Fried & Nesse, 2015b). When considering individual symptoms, they do not appear to be equal, with evidence suggesting that individual symptoms may vary in the degree to which they contribute to functional impairment (Fried & Nesse, 2014), their risk factors (Fried, Nesse, Zivin, Guille, & Sen, 2014), and their heritability (Jang, Livesley, Taylor, Stein, & Moon, 2004). Furthermore, they may differ in their response to different treatment (Bekhuis et al., 2018; Boschloo et al., 2019a) and in terms of their association with clinical outcomes (O'Driscoll et al., 2021). High variability in the presence of symptoms, combined with evidence suggesting that symptoms are not interchangeable implies that rather than being clearly bound disorders, depression and anxiety have

a complex and heterogenous presentation. Sum scores, which assume equivalence of symptoms, may therefore hide important differences.

Outcome research for depression and anxiety is further complicated by the fact that psychotherapies such as cognitive-behavioural therapy (CBT) are often considered to be a 'black box' (Huibers, Lorenzo-Luaces, Cuijpers, & Kazantzis, 2021). They commonly target a broad group of symptoms (Eronen, 2020) and identifying how treatments works, and for whom, is arguably the main scientific challenge facing depression and anxiety outcome researchers (Carey, Griffiths, Dixon, & Hines, 2020; Huibers et al., 2021; Paul, 1967). Despite multiple putative mechanisms being proposed, that reflect both 'active ingredients' of specific psychotherapies and non-specific effects shared across psychotherapies, the evidence remains ambiguous with no universal consensus concerning specific mechanisms of change (Carey et al., 2020; Huibers et al., 2021).

Exploring how symptoms respond to treatment, rather than sum scores alone, may lead to important insights and provide first steps toward precision psychiatry. If specific therapeutic interventions could be mapped onto change in specific symptoms it may begin to explain how psychotherapies work and for whom. More pragmatically perhaps, understanding possible symptom-specific effects of treatments could also potentially lead to a more personalised matching of symptoms profiles to treatments as well as having the potential to explain heterogeneity in clinical outcomes. Some treatments may cause side effects that increase some of the very symptoms used to measure depression, such as antidepressant medication and the possible side effect of sleep disturbances (Wichniak, Wierzbicka, Wałęcka, & Jernajczyk, 2017). These could potentially mask or dilute specific benefits on other symptoms when sum scores are used (Fried & Nesse, 2015a).

Overall, there is a strong argument to be made around exploring symptom-specific effects. The evidence suggests that CBT is an effective treatment for depression and anxiety (Cuijpers, Cristea, Karyotaki, Reijnders, & Huibers, 2016). However, most evaluations of CBT focus on sum scores. As such, less is known about the symptom-specificity of CBT. In the present study, we aim to explore how individual symptoms of depression and generalised anxiety behave across a course of CBT in an observational, retrospective cohort by examining symptom trajectories and how they compare to one another.

Methods

Settings

Improving Access to Psychological Therapies (IAPT) is a national programme that delivers psychological therapy for depression and anxiety across England in primary care settings (Clark, 2011; Clark et al., 2018). IAPT offers a variety of psychological therapies, including both low-intensity therapy (LIT) and high-intensity therapy (HIT) (Clark, 2011; Clark et al., 2018). IAPT has implemented routine outcome monitoring, where detailed information is gathered about patients, their treatment, and their clinical outcomes (Clark, 2011). These data are collected on a session-by-session basis to increase complete-case recording.

Sample

The clinical records for the present study were obtained from ten IAPT practices across the southwest of England and London, who

consented to share their data via Mayden, the providers of a patient management software used in IAPT, for the purposes of this research. The data from each participating services were extracted and fully anonymised by Mayden before sharing with us for processing and analysis. The present analysis contained data for referrals being referred from the year 2014 to mid-2019 – earlier years were excluded due to operational issue early in IAPT such as missing or inappropriate diagnostic labels which lead to poorer outcomes as a result of inappropriate matching of treatment protocols to clinical needs of the patient (Saunders et al., 2020b).

Referrals were included if they had received high-intensity CBT for depression or generalised anxiety. Patients receiving treatment for depression and generalised anxiety were identified by their diagnostic labels of having a depressive episode, recurrent depressive disorder, or generalised anxiety disorder as well as having their primary outcomes measures as the PHQ-9 (Kroenke, Spitzer, & Williams, 2001) and GAD-7 (Spitzer, Kroenke, & Williams, 2006), which are used to monitor treatment process and evaluate treatment response – other anxiety measures indicate treatment for different anxiety disorders. Treatment was defined as CBT if all the recorded treatment labels were labelled as CBT. Referrals receiving other therapies in addition to CBT were excluded due to difficulties isolating the effects of CBT. In order to examine the effect of CBT on specific questionnaire items, only patients who had at least one full set of item-level questionnaire scores recorded were included in the analyses. Due to the observational nature of the data no fixed amount of treatment is provided. To counteract the issue of unequal treatment doses, a minimum treatment dose was defined as eight appointments – the overall average number of delivered appointments in IAPT (NHS Digital, 2021). Eight appointments may also be more likely to at least partially capture some of the later responses which are overserved for some subgroups of patients in therapy which can occur approximately after six appointments (Saunders et al., 2019). As such, all subsequent analyses examine the first eight appointments from each referral that had at least eight appointments. To minimise the impact of previous treatment effects, each patient's first referral, that met the criteria above, was used.

Intervention

CBT was delivered by mental health professionals trained in accordance with the national curriculum (Department of Health, 2019). CBT is a time-limited, structured, and problem-focused psychological therapy (Fenn & Byrne, 2013). CBT explores the links between cognitions, emotions, and behaviours and supports patients in identifying and modifying less helpful cognitions and behaviours and develop alternative, more adaptive ones (Fenn & Byrne, 2013).

Measures

Patient Health Questionnaire 9 item (PHQ-9)

The PHQ-9 is a 9-item self-report questionnaire that assesses the severity of depressive symptoms over the past 2 weeks (Kroenke et al., 2001). Each item is rated on a 4-point Likert scale ranging from 0 ('not at all') to 3 ('nearly every day'). Patients are asked to rate the severity of symptoms relating to: (1) *anhedonia*, (2) *low mood/hopelessness*, (3) *sleeping problems* (4) *tired/low energy* (5) *appetite*, (6) *guilt/worthlessness*, (7) *concentration*, (8) *psychomotor retardation/agitation*, and (9) *suicidal thoughts*.

Generalised Anxiety Disorder 7 item scale (GAD-7)

The GAD-7 is a 7-item self-report questionnaire assesses the severity of generalised anxiety over the past 2 weeks (Spitzer et al., 2006). Each item is rated on a 4-point Likert scale ranging from 0 (not at all) to 3 (nearly every day). Patients are asked to rate the severity of symptoms relating to: (1) *nervous/anxious*, (2) *uncontrollable worry*, (3) *too much worry*, (4) *trouble relaxing*, (5) *restlessness*, (6) *irritability*, and (7) *fear*.

Statistical analysis

All analyses were performed in the R statistical programming language (R Core Team., 2013). We used a generalised logistic mixed model to examine how items of the PHQ-9 and GAD-7 responded over the course of CBT treatment appointments using the `glmer` function from the 'lme4' package (Bates, Maechler, Bolker, & Walker, 2014). To simplify modelling, the item scores were dichotomised to model the probability of being symptom free, where a score of 0 denotes being symptom free and scores of 1–3 denotes having symptoms. The default call for `glmer()` uses Gaussian-Hermite quadrature which we used to estimate the model parameters but is computationally expensive. To obtain the confidence intervals we used bootstrap computation with the approximation $nAGQ = 0$ to make the calculation feasible.

Two separate models were built for patients being treated for depression and generalised anxiety, using the PHQ-9 and GAD-7 respectively as outcome measures. All dichotomised item scores for each questionnaire were specified as the outcome, with the appointments and question item being specified as predictors with an interaction term. As such, all PHQ-9 items were analysed in one model and all GAD-7 items being analysed in a separate model. The models were adjusted for baseline characteristics: gender, age, ethnicity, employment status, Index of Multiple Deprivation (McLennan et al., 2019), disability status, long-term health condition status, diagnosis, sum score baseline PHQ-9, sum score baseline GAD-7, sum score baseline Work and Social Adjustment Scale (Mundt, Marks, Shear, & Greist, 2002) medication status, referral number, referral source, service and year. IAPT services collect various information about patients at baseline in a core dataset, with some services collecting additional information. Baseline characteristics were chosen on the data availability such as variables that were most consistently recorded across services, data quality of the recorded variables, and informed by considered variables in previous literature using IAPT records (see for example Delgadoillo & Gonzalez Salas Duhne, 2020; Delgadoillo, Moreea, & Lutz, 2016; Green et al., 2015; Saunders, Buckman, & Pilling, 2020a). Missing baseline covariates were singly imputed using random forest with 'missForest' (Stekhoven & Bühlmann, 2012). The cost of computation and the complexity of the model made the preferred option of multiple imputations impractical. Single imputation is adequate here because (a) the small amount of missing baseline data (maximum 7.4%), (b) evidence that missForest is effective in similar scenarios (Stekhoven & Bühlmann, 2012; Waljee et al., 2013) and (c) satisfactory out-of-bag error suggests the imputation was successful with a normalised root mean square error of 0.39 for continuous variables and proportion of falsely classified of 0.15 for categorical variables. Random intercepts were specified in all models for referral id and appointment id to account for repeated observations. As there is no natural reference group on questionnaires which would provide an appropriate comparator

to measure all other items against, sum coding was used. Sum coding allows the trajectory of each individual question to be compared to the mean trajectory of all other questions. Thus, proving a comparison of how specific trajectories compare relative to all others. Variability estimates for all questions relative to other questions are provided, except the last question of each questionnaire as these are estimated by the inverse of the sum of the log odds of all other questions. Longitudinal models were used to predict the probability of being symptom free on the questionnaire items across CBT treatment appointments. Predictions were made for a hypothetical 'average' patient with continuous variables fixed at the mean and categorical covariate set to the most frequently occurring group. The first eight appointments were used as a basis for predictions of up to 20 and 15 appointments to depict how item trajectories might respond for the NICE-recommend durations of treatment for depression and generalised anxiety respectively (NICE, 2009, 2011). Bootstraps with 5000 simulations were used to approximate 95% confidence intervals for predictions. The 'ggplot2' package was used to visualise questionnaire item trajectories (Wickham, 2011).

Results

Baseline characteristics

Of the 5306 referrals included in the main analysis, the majority of patients were female, White, with an average age of 38 years, experiencing severe depression and moderate/severe anxiety (Table 1).

All PHQ-9 and GAD-7 symptoms appeared to improve across CBT appointments. However, the rate at which they improved relative to the average rate of improvement across all questionnaire items varied.

Patient Health Questionnaire-9

We found the strongest evidence that PHQ-9 items 2 (*low mood/hopelessness*) and 6 (*guilt/worthlessness*) improved fastest across CBT appointments in comparison to the average rate of all questionnaire items (Table 2). Patients had 7% higher odds of having no *low mood/hopelessness* with every appointment compared to all other depression symptoms, or 56% across 8 appointments. Patients had 6% higher odds of having no *guilt/worthlessness* with every appointment compared to all other depression symptoms, or 48% across 8 appointments. We also found the strongest evidence that PHQ-9 items 3 (*sleeping problems*), 5 (*appetite*) and 8 (*psychomotor retardation/agitation*) improved at a slower rate across CBT appointments in comparison to the average response of all other questionnaire items. Patients had 6, 7, and 6% lower odds of having no symptoms with every additional CBT appointment on *sleeping problems*, *appetite* and *psychomotor retardation/agitation* respectively, compared to all other depression symptoms; or 48, 56 and 48% across 8 appointments. Some evidence suggested that the odds of being symptom free were higher for PHQ-9 item 1 (*anhedonia*) and PHQ-9 item 4 (*tired/low energy*); however, these associations were weaker. PHQ-9 item 7 (*concentration*) and PHQ-9 item 9 (*suicidal thoughts*) did not appear to differ in their rate of improvement compared to the average rate.

Figure 1 shows that the probability of being symptom free has different starting points for the PHQ-9 items. The predicted probability of being symptom free is much higher at baseline for PHQ-9 items 8 (*psychomotor retardation/agitation*) and 9

Table 1. Baseline clinical and sociodemographic patient characteristics

	Mean (s.d.)/Count (%)
<i>N</i>	5306
Gender	
Female	3675 (69.3)
Male	1631 (30.7)
Age	37.56 (13.41)
Ethnicity	
White	4648 (87.6)
Black, Asian, and ethnic minorities	658 (12.4)
Employment status	
Employed	3256 (61.4)
Not working	2050 (38.6)
Index of Multiple Deprivation	20.46 (12.08)
Disability status	
None	4664 (87.9)
Yes	642 (12.1)
Long-Term health condition status	
None	3651 (68.8)
Yes	1655 (31.2)
Diagnosis	
Depressive episode	2437 (45.9)
Recurrent depressive disorder	977 (18.4)
Generalised anxiety disorder	1892 (35.7)
Medication status	
Yes	2909 (54.8)
No	2397 (45.2)
Referral number	1.97 (1.38)
Referral source	
Self	3437 (64.8)
Primary care	1572 (29.6)
Other	297 (5.6)
Patient Health Questionnaire-9	16.61 (5.61)
Generalised Anxiety Disorder Scale-7	14.70 (4.47)
Work and Social Adjustment Scale	22.19 (8.82)

(suicidal thoughts). This is a result of the distribution of the response scores on each item at baseline, with items 8 (*psychomotor retardation/agitation*) and 9 (*suicidal thoughts*) having a higher frequency of 0 scores at baseline and through treatment.

Generalised Anxiety Disorder scale-7

We found the strongest evidence that GAD-7 items 2 (*uncontrollable worry*) and 3 (*too much worry*) improved fastest across CBT in comparison to the mean response of all questionnaire items (Table 2). For both items, patients had 9% higher odds of having no symptoms compared to the average of all other anxiety

Table 2. Effects of appointment on the odds of being symptom free for item-level scores stratified by questionnaire

	Odds ratio	95% Confidence Intervals		<i>p</i> value
Patient Health Questionnaire-9				
Question 1 × Appointment	1.04	1.01	1.07	0.014
Question 2 × Appointment	1.07	1.04	1.11	<0.001
Question 3 × Appointment	0.94	0.91	0.96	<0.001
Question 4 × Appointment	1.04	1.00	1.08	0.027
Question 5 × Appointment	0.93	0.91	0.95	<0.001
Question 6 × Appointment	1.06	1.03	1.09	<0.001
Question 7 × Appointment	0.99	0.97	1.01	0.335
Question 8 × Appointment	0.94	0.92	0.96	<0.001
Question 9 × Appointment	1.00	0.98	1.03	–
Generalised Anxiety Disorder Scale-7				
Question 1 × Appointment	1.07	1.00	1.13	0.044
Question 2 × Appointment	1.09	1.04	1.14	<0.001
Question 3 × Appointment	1.09	1.04	1.15	<0.001
Question 4 × Appointment	1.03	0.99	1.07	0.169
Question 5 × Appointment	0.89	0.87	0.92	<0.001
Question 6 × Appointment	0.93	0.90	0.96	<0.001
Question 7 × Appointment	0.93	0.89	0.97	–

Adjusted for main effects: age, gender, ethnicity, employment status, Index of Multiple Deprivation, disability status, long-term health conditions, diagnosis (for depression only), baseline Patient Health Questionnaire-9, baseline Generalised Anxiety Disorder Scale-7, baseline Work and Social Adjustment Scale, medication status, referral number, referral source, service, year, question, appointment as well as random effects of referral id and appointment id. *p* values for the reference group of the sum coding can not be estimated.

symptoms, or 72% across 8 appointments. We also found the strongest evidence that GAD-7 items 5 (*restlessness*) and 6 (*irritability*) improved slower across CBT appointments in comparison to all other questionnaire items. Patients were 11% and 7% less likely to have no symptoms of *restlessness* and *irritability* respectively, compared to all other anxiety symptoms; or 88% and 56% across 8 appointments. There was some evidence to suggest that the odds of being symptom free were higher for GAD-7 item 1 (*nervous/anxious*); however, this association was weak. While no variability estimates are computed for the last question, it appears that the odds of being symptom free on the GAD-7 item 7 (*fear*) improved at a slower pace compared to all other questions. GAD-7 item 4 (*trouble relaxing*) did not appear to differ in the rate of improvement compared to the average rate.

Figure 2 shows that the probability of being symptom free has different starting points for the GAD-7 items. The predicted probability of being symptom free is somewhat higher at baseline for GAD-7 items 5 (*restlessness*), 6 (*irritability*), and 7 (*fear*). This is a result of the distribution of the response scores on each item at baseline, these items having a somewhat higher frequency of 0 scores at baseline and throughout treatment.

We performed a sensitivity analysis in a subset of patients who had all their item-level scores recorded and found effects of a similar magnitude and direction (online Supplementary Tables S5 and S6). We further examined whether the slope of the

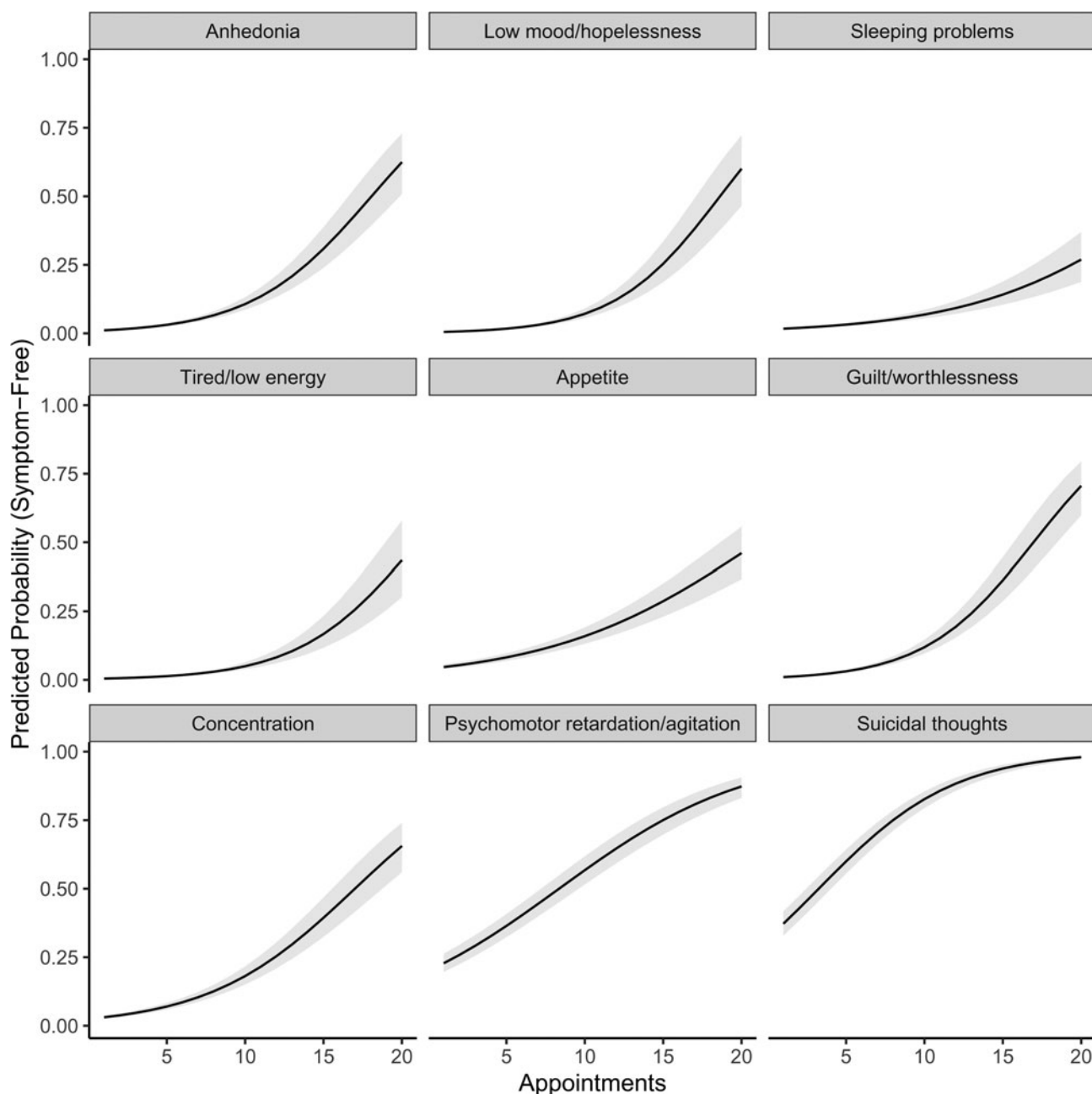


Fig. 1. Predicted probability of being symptom free on individual Patient Health Questionnaire-9 items across cognitive-behavioural therapy appointments.

trajectories for each item varied by baseline medication status and found no evidence of a differential improvement across items by medication status (PHQ-9: $p = 0.769$ and GAD-7: $p = 0.561$). Furthermore, we assessed whether there were inherent differences in baseline patient characteristics amongst those who had item-level data recorded and those who did not but found no evidence to suggest that this was the case, with all standardised mean differences between baseline characteristics falling < 0.25 (Panos & Mavridis, 2020; Rubin, 2001).

Discussion

We examined the trajectories of individual depression and anxiety symptoms as measured by two widely used measures, the PHQ-9

and the GAD-7, in a large, retrospective, observational cohort of patients receiving CBT. We found evidence to suggest *low mood/hopelessness* and *guilt/worthlessness* on the PHQ-9, as well as *uncontrollable worry* and *too much worry* on the GAD-7 improved at a faster rate relative to other symptoms. We found that *sleeping problems*, *appetite*, and *psychomotor agitation/retardation* on the PHQ-9 and *restlessness* and *irritability* on the GAD-7 improved at a slower rate compared to the average response of all other symptoms.

Worry is clinically central to generalised anxiety (APA, 2013; NICE, 2011) and working with worry and tolerance of uncertainty forms a critical part of treatment protocols for generalised anxiety disorder (Department of Health, 2019; University College London, n.d.). Low mood is considered one of the core clinical

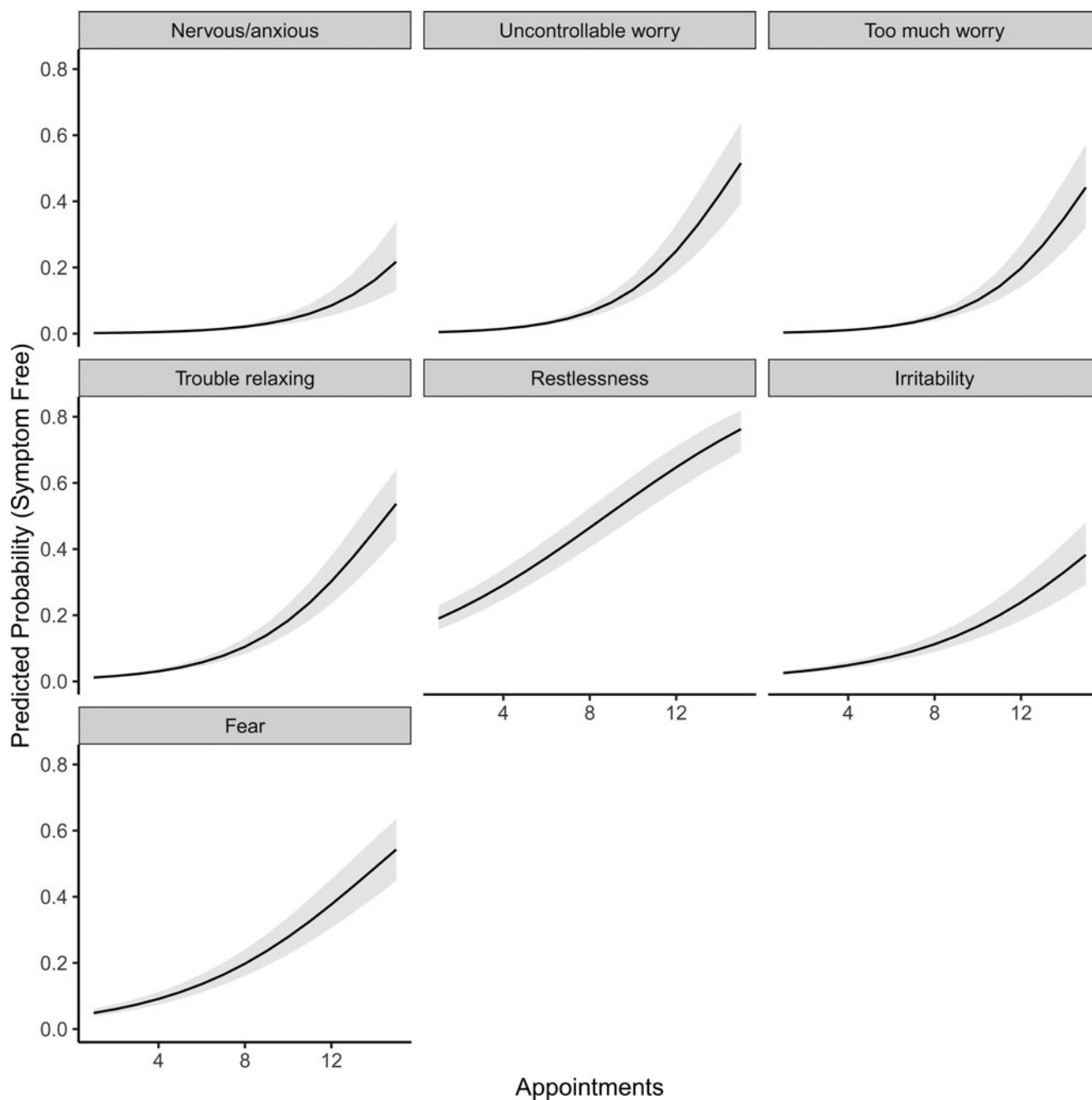


Fig. 2. Predicted probability of being symptom free on individual Generalised Anxiety Disorder-7 items across cognitive-behavioural therapy appointments.

features of depression (APA, 2013; NICE, 2009) and has been shown to be the biggest contributor of depressive symptoms to functional impairment explaining ~20% of the variance, with self-blame accounting for ~6% (Fried & Nesse, 2014). Various cognitive and behavioural components are delivered during CBT for depression, commonly starting with behavioural work in severe depression (Department of Health, 2019). However, working with negative automatic thoughts and themes of guilt or self-blame are also emphasised (Department of Health, 2019). Working with hopelessness may also be relevant to clinicians due to its association with clinical risk (McMillan, Gilbody, Beresford, & Neilly, 2007). Our research suggests that some of the symptoms that traditionally feature strongly in the

clinical conceptualisation and treatment protocols improved relatively faster than other symptoms. However, *sleeping problems, poor appetite or overeating, and psychomotor agitation/retardation* appeared to improve to a lesser degree in the present research yet account for ~4, 11 and 8% of impairment respectively and may still be relevant symptoms to patients that may require additional attention (Fried & Nesse, 2014).

Our findings that depressive symptoms like *low mood/hopelessness* and *guilt/worthlessness* as well as generalised anxiety symptoms such as *uncontrollable worry* and *too much worry* change comparatively the most during early treatment resonate with findings from network analyses. These have previously suggested that symptoms relating to low mood or hopelessness and failure, guilt

or worthlessness are amongst the most central symptoms (Beard et al., 2016; O'Driscoll et al., 2021). Symptoms relating to worry have been reported to be the most central anxiety symptoms (Beard et al., 2016). Some arguments suggest that central symptoms are critical elements within networks as they interlink closely with other symptoms (Beard et al., 2016). They are often considered to be highly relevant in the maintenance of a network and are potentially important treatment targets (Beard et al., 2016). However, caution should be taken to avoid overinterpretation given that centrality does not necessarily equate to clinical importance (Bringmann et al., 2019; O'Driscoll et al., 2021) or clinical utility, such as prognostic capacity (Buckman et al., 2021; O'Driscoll et al., 2021). As such, the literature appears ambiguous regarding the practical meaning and application of centrality metrics. It is nonetheless encouraging that our findings suggest several of the symptoms thought to be central in networks, as well as clinical conceptualisations, change comparatively more during treatment. However, given these noted limitations more research is needed.

The present research has primarily focused on how symptoms change across CBT broadly. Ideally, these changes could be mapped onto specific therapeutic techniques. This may begin to explain which elements of CBT are acting on specific symptoms and why some symptoms may increase at a faster pace. Disentangling the complex nature of psychotherapies and how these relate to specific symptoms would have important implications for precision psychiatry – a more detailed understanding of which therapeutic techniques have an effect on specific symptoms could potentially lead to better treatment matching on both empirical and theoretical grounds. However, having no detailed information regarding the order/nature of specific techniques and the lack of fidelity measures in IAPT make this difficult to assess (Martin, Iqbal, Airey, & Marks, 2022). Component studies of CBT, which also look at symptoms rather than sum scores alone, may be beneficial in gaining more granular insights into CBT; however, to date, efforts to understand specific therapeutic techniques appear underpowered (Cuijpers, Cristea, Karyotaki, Reijnders, & Hollon, 2019). Including more detailed recording of therapeutic interventions as part of routine care may be a feasible way to generate large and rich datasets to begin to address some of these questions with sufficient statistical power.

Strengths & limitations

The present study was based on data from a large retrospective cohort of patients receiving CBT that was obtained from clinical practice to examine individual symptoms. The naturalistic setting adds to the ecological validity of the findings. We also used an interpretable model to examine how individual symptoms change relative to one another and used a treatment duration that is reasonably representative of the average adult being treated for depression and anxiety in primary care in England.

Despite these strengths, several limitations should be considered. The reliability and validity of using item-level questionnaire responses, rather than sum scores, is still under debate (Boschloo et al., 2019b). While examining how symptoms improve relative to one another is strength, the lack of a control group limits causal conclusions regarding the symptom-specific effects of CBT. For example, the differential trajectories of specific symptoms might be explained by items with higher baseline scores, such as *guilt/worthless* and *uncontrollable worry*, having more scope to regress to the mean. However, not all symptoms that improved

comparatively less had lower baseline scores, i.e. *sleeping problems* improved comparably less than other symptoms but had a similar baseline severity to symptoms that improved most. A further explanation may be that symptoms that improved the most may be more amendable to natural change/recovery. However, this seems unlikely given that symptoms that improved the most included low mood and worry, which are considered core symptoms of depression and anxiety (APA, 2013). We also model the probability of being symptom free. It may nonetheless be clinically meaningful if patients move from a score of 3 (experiencing symptoms 'nearly every day') to a score of 1 (experiencing symptoms 'several days'). Due to the modelling choice, we do not capture this nuance.

While efforts were made to capture a sample of patients who received high-intensity CBT, there are limitations of examining treatment in observational data. While clinicians are trained in accordance with clinical protocols, the lack of treatment fidelity measures in IAPT (Martin et al., 2022) raises possible concerns about the consistency and fidelity to protocol of treatment that are delivered in routine care. A lack fidelity is likely to dilute any observed effects. Relatedly, clinicians may already be tailoring interventions to individual clients with clinical work placing importance on the prioritisation of problem areas (Department of Health, 2019). We also cannot exclude the possibility of any data entry errors or variations between clinicians and services on how data is recorded, related to treatment but also for other recorded characteristics.

The main analysis was based on eight appointments; this has multiple implications. A proportion of patients will require less than eight appointments to improve whilst others drop out of treatment. This likely affects the probability of overall response but is less likely to influence relative symptom trajectories, unless specific symptoms are predictors of faster response or drop out. Secondly, other symptoms might improve at a later stage of treatment either by being targeted by specific CBT techniques which may be scheduled later in treatment or as a consequence of cascading effects of improvements in other symptoms. Given our choice of eight appointments, which is the average number of appointments delivered in IAPT, and potentially covers some of the later responses observed for some patients after approximately six appointments (Saunders et al., 2019) we can be somewhat confident that these results may generalise to the average patient. However, we cannot exclude the possibility that the rate of change for symptoms may differ over the course of longer treatment. As predictions for appointments beyond eight are extrapolations based on the first eight appointments, they remain speculative and are primarily for illustrative purposes.

Implications

This study suggests that sum scores may obscure important differences of the effect of interventions on specific symptoms. However, we only captured symptoms recorded on the PHQ-9 and GAD-7, which are less extensive than other questionnaires. Given their suggested relevance, questionnaires that include non-DSM symptoms may be insightful (Fried, Epskamp, Nesse, Tuerlinckx, & Borsboom, 2016). Similarly, questionnaires that capture symptoms more granularly, such as disaggregating symptoms like psychomotor agitation and retardation, may be beneficial given their differential contribution to impairment (Fried & Nesse, 2014). A particular focus on anxiety may also be beneficial given that there is comparably less focus on individual symptoms.

While the findings of the present research require replication and more rigorous evaluation, they do provide a promising avenue for understanding the heterogeneity in clinical outcomes as well precision psychiatry/psychotherapy. It could potentially be possible to match patients to treatment based on the baseline presentation of symptoms; particularly if the symptom-specific effects of CBT and other interventions are better understood and evaluated in the future. Furthermore, it provides the opportunity of more targeted combinations of interventions that may work on symptoms which respond comparatively less. That is, in the present research, sleeping problems appeared to improve comparatively less. As such, augmenting CBT with an intervention that targets sleep problems, such as an internet-delivered insomnia programme, amongst patients with higher levels of insomnia may provide a relatively cost-effective way to improve treatment outcomes (Darden et al., 2021; Ye et al., 2016).

However, much of the evidence that will likely lead to making such improvements in personalised psychotherapy more tangible is yet to be robustly established and replicated. In addition to a lack of consensus on how psychological therapy works, the pathophysiology and aetiological mechanisms of depression and anxiety is also poorly understood (Nemeroff, 2020). Further research that addresses both areas will undoubtedly contribute to improving clinical outcomes if therapeutic interventions can then be matched to clinical presentations based on their mechanisms.

Conclusion

Trajectories of symptoms appear to differ during treatment with CBT, highlighting the importance of examining individual symptoms rather than sum scores alone. It appears that CBT improves at least some of the core symptoms of depression and anxiety, or symptoms that could be considered very clinically/theoretically important, fastest. While the research is observational and warrants further evaluation, there are possible implications for clinical practice and provide potential avenues for future research. Examining symptoms may provide an avenue for precision psychiatry/psychotherapy, allowing for more targeted prescription of treatments and/or more evidence-based augmentation of treatments.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291722001556>.

Financial support. This work was conducted as part of a PhD studentship from the University of Bath awarded to Clarissa Bauer-Staeb.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. Retrieved from <http://CRAN.R-project.org/package=lme4>.
- Beard, C., Millner, A. J., Forgeard, M. J., Fried, E. I., Hsu, K. J., Treadway, M. T., ... Björngvinsson, T. (2016). Network analysis of depression and anxiety symptom relationships in a psychiatric sample. *Psychological Medicine*, 46(16), 3359–3369. doi:10.1017/s0033291716002300.
- Bekhuis, E., Schoevers, R., de Boer, M., Peen, J., Dekker, J., Van, H., & Boschloo, L. (2018). Symptom-specific effects of psychotherapy versus combined therapy in the treatment of mild to moderate depression: A network approach. *Psychotherapy and Psychosomatics*, 87(2), 121–123. <https://doi.org/10.1159/000486793>.
- Boschloo, L., Bekhuis, E., Weitz, E. S., Reijnders, M., DeRubeis, R. J., Dimidjian, S., ... Cuijpers, P. (2019a). The symptom-specific efficacy of antidepressant medication vs. cognitive behavioral therapy in the treatment of depression: Results from an individual patient data meta-analysis. *World Psychiatry*, 18(2), 183–191. <https://doi.org/10.1002/wps.20630>.
- Boschloo, L., Cuijpers, P., Karyotaki, E., Berger, T., Moritz, S., Meyer, B., & Klein, J. P. (2019b). Symptom-specific effectiveness of an internet-based intervention in the treatment of mild to moderate depressive symptomatology: The potential of network estimation techniques. *Behaviour Research and Therapy*, 122, 103440. <https://doi.org/10.1016/j.brat.2019.103440>.
- Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., ... Snippe, E. (2019). What do centrality measures measure in psychological networks? *Journal of Abnormal Psychology*, 128(8), 892–903. <https://doi.org/10.1037/abn0000446>.
- Buckman, J., Cohen, Z., O'Driscoll, C., Fried, E., Saunders, R., Ambler, G., ... Pilling, S. (2021). Predicting prognosis for adults with depression using individual symptom data: A comparison of modelling approaches. *Psychological Medicine*, 1–11. doi:10.1017/S0033291721001616.
- Carey, T. A., Griffiths, R., Dixon, J. E., & Hines, S. (2020). Identifying functional mechanisms in psychotherapy: A scoping systematic review. *Frontiers in Psychiatry*, 11, 291. <https://doi.org/10.3389/fpsy.2020.00291>.
- Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry*, 23(4), 318–327. <https://doi.org/10.3109/09540261.2011.606803>.
- Clark, D. M., Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): An analysis of public data. *The Lancet*, 391(10121), 679–686. [https://doi.org/10.1016/S0140-6736\(17\)32133-5](https://doi.org/10.1016/S0140-6736(17)32133-5).
- Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Hollon, S. D. (2019). Component studies of psychological treatments of adult depression: A systematic review and meta-analysis. *Psychotherapy Research*, 29(1), 15–29. <https://doi.org/10.1080/10503307.2017.1395922>.
- Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Huibers, M. J. (2016). How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *World Psychiatry*, 15(3), 245–258. <https://doi.org/10.1002/wps.20346>.
- Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychological Medicine*, 40(2), 211–223. doi:10.1002/wps.20701.
- Darden, M., Espie, C. A., Carl, J. R., Henry, A. L., Kanady, J. C., Krystal, A. D., & Miller, C. B. (2021). Cost-effectiveness of digital cognitive behavioral therapy (Sleepio) for insomnia: A Markov simulation model in the United States. *Sleep*, 44(4), zsa223. <https://doi.org/10.1093/sleep/zsa223>.
- Delgado, J., & Gonzalez Salas Duhne, P. (2020). Targeted prescription of cognitive-behavioral therapy versus person-centered counseling for depression using a machine learning approach. *Journal of Consulting and Clinical Psychology*, 88(1), 14. <https://doi.org/10.1037/ccp0000476>.
- Delgado, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Behaviour Research and Therapy*, 79, 15–22. doi:10.1016/j.brat.2016.02.003.
- Department of Health. (2019). *National curriculum for high intensity cognitive behavioural therapy courses* (3rd ed.). Department of Health. Retrieved from <https://healtheducationengland.sharepoint.com/Comms/Digital/Shared%20Documents/Forms/AllItems.aspx?id=%2FComms%2FDigital%2FShared%20Documents%2Fhee%2Enh%2Euk%20documents%2FWebsite%20files%2FIAPT%2F2%2E%20HIT%20Cognitive%20Behaviour%20Therapy%20Curriculum%20%5F%20depression%20and%20anxiety%20disorders%2Epdf&parent=%2FComms%2FDigital%2FShared%20Documents%2Fhee%2Enh%2Euk%20documents%2FWebsite%20files%2FIAPT&p=true> (Accessed 15 February 2022).
- Eronen, M. I. (2020). Causal discovery and the problem of psychological interventions. *New Ideas in Psychology*, 59, 100785. <https://doi.org/10.1016/j.newideapsych.2020.100785>.
- Fenn, K., & Byrne, M. (2013). The key principles of cognitive behavioural therapy. *InnovAiT*, 6(9), 579–585. <https://doi.org/10.1177/1755738012471029>.

- Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders*, 189, 314–320. <https://doi.org/10.1016/j.jad.2015.09.005>.
- Fried, E. I., & Nesse, R. M. (2014). The impact of individual depressive symptoms on impairment of psychosocial functioning. *PLoS One*, 9(2), e90311. <https://doi.org/10.1371/journal.pone.0090311>.
- Fried, E. I., & Nesse, R. M. (2015a). Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. *BMC Medicine*, 13(1), 1–11. <https://doi.org/10.1186/s12916-015-0325-4>.
- Fried, E. I., & Nesse, R. M. (2015b). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR* D study. *Journal of Affective Disorders*, 172, 96–102. <https://doi.org/10.1016/j.jad.2014.10.010>.
- Fried, E. I., Nesse, R. M., Zivin, K., Guille, C., & Sen, S. (2014). Depression is more than the sum-score of its parts: Individual DSM symptoms have different risk factors. *Psychological Medicine*, 44(10), 2067. doi:10.1017/S0033291713002900.
- GBD 2019 Mental Disorders Collaborators. (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry*, 11, 137–150. [https://doi.org/10.1016/S2215-0366\(21\)00395-3](https://doi.org/10.1016/S2215-0366(21)00395-3).
- Green, S. A., Honeybourne, E., Chalkley, S. R., Poots, A. J., Woodcock, T., Price, G., ... Green, J. (2015). A retrospective observational analysis to identify patient and treatment-related predictors of outcomes in a community mental health programme. *BMJ Open*, 5(5), e006103. <http://doi.org/10.1136/bmjopen-2014-006103>.
- Huibers, M. J., Lorenzo-Luaces, L., Cuijpers, P., & Kazantzis, N. (2021). On the road to personalized psychotherapy: A research agenda based on cognitive behavior therapy for depression. *Frontiers in Psychiatry*, 11, 1551. <https://doi.org/10.3389/fpsy.2020.607508>.
- Jang, K. L., Livesley, W. J., Taylor, S., Stein, M. B., & Moon, E. C. (2004). Heritability of individual depressive symptoms. *Journal of Affective Disorders*, 80(2–3), 125–133. [https://doi.org/10.1016/S0165-0327\(03\)00108-3](https://doi.org/10.1016/S0165-0327(03)00108-3).
- 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(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>.
- Martin, C., Iqbal, Z., Airey, N. D., & Marks, L. (2022). Improving Access to Psychological Therapies (IAPT) has potential but is not sufficient: How can it better meet the range of primary care mental health needs? *The British Journal of Clinical Psychology*, 61(1), 157–174. <https://doi.org/10.1111/bjc.12314>.
- McLennan, D., Noble, S., Noble, M., Plunkett, E., Wright, G., & Gutacker, N. (2019). *The English indices of deprivation 2019: Technical report*. Ministry of Housing, Communities and Local Government.
- McMillan, D., Gilbody, S., Beresford, E., & Neilly, L. (2007). Can we predict suicide and non-fatal self-harm with the Beck Hopelessness Scale? A meta-analysis. *Psychological Medicine*, 37(6), 769–778. doi:10.1017/S0033291706009664.
- 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(5), 461–464. doi:10.1192/bjp.180.5.461.
- National Institute for Clinical Excellence. (2009). *Depression in adults: Recognition and management – clinical guidelines [CG90]*. National Institute for Clinical Excellence. Retrieved from <https://www.nice.org.uk/guidance/cg90> (Accessed 29 January 2022).
- National Institute for Clinical Excellence. (2011). *Generalised anxiety disorder and panic disorder in adults: Management– clinical guidelines [CG113]*. National Institute for Clinical Excellence. Retrieved from <https://www.nice.org.uk/guidance/cg113> (Accessed 29 January 2022).
- Nemeroff, C. B. (2020). The state of our understanding of the pathophysiology and optimal treatment of depression: Glass half full or half empty? *American Journal of Psychiatry*, 177(8), 671–685. <https://doi.org/10.1176/appi.ajp.2020.20060845>.
- NHS Digital. (2021). *Psychological therapies, annual report on the use of IAPT services, 2020–21*. NHS Digital. Retrieved February 4, 2022, from <https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-annual-reports-on-the-use-of-iapt-services/annual-report-2020-21#>.
- O'Driscoll, C., Buckman, J., Fried, E. I., Saunders, R., Cohen, Z. D., Ambler, G., ... Pilling, S. (2021). The importance of transdiagnostic symptom level assessment to understanding prognosis for depressed adults: Analysis of data from six randomised control trials. *BMC Medicine*, 19(1), 109. <https://doi.org/10.1186/s12916-021-01971-0>.
- Panos, A., & Mavridis, D. (2020). TableOne: An online web application and R package for summarising and visualising data. *Evidence-Based Mental Health*, 23(3), 127–130. <http://doi.org/10.1136/ebmental-2020-300162>.
- Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, 31(2), 109. doi:10.1037/h0024436.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>.
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2(3), 169–188. <https://doi.org/10.1023/A:1020363010465>.
- Saunders, R., Buckman, J. E. J., Cape, J., Fearon, P., Leibowitz, J., & Pilling, S. (2019). Trajectories of depression and anxiety symptom change during psychological therapy. *Elsevier B.V. Journal of Affective Disorders*, 249, 327–335. <https://doi.org/10.1016/j.jad.2019.02.043>.
- Saunders, R., Buckman, J. E., & Pilling, S. (2020a). Latent variable mixture modelling and individual treatment prediction. *Behaviour Research and Therapy*, 124, 103505. <https://doi.org/10.1016/j.brat.2019.103505>.
- Saunders, R., Cape, J., Leibowitz, J., Aguirre, E., Jena, R., Cirkovic, M., ... Buckman, J. E. J. (2020b). Improvement in IAPT outcomes over time: Are they driven by changes in clinical practice? *The Cognitive Behaviour Therapist*, 13, 1–15. doi:10.1017/S1754470X20000173.
- Spitzer, R. L., Kroenke, K., & Williams, J. B. (2006). Generalized anxiety disorder 7-item (GAD-7) scale. *Archives of Internal Medicine*, 166, 1092–1097. doi:10.1001/archinte.166.10.1092.
- Stekhoven, D. J., & Bühlmann, P. (2012). MissForest – non-parametric missing value imputation for mixed-type data. *Bioinformatics (Oxford, England)*, 28(1), 112–118. <https://doi.org/10.1093/bioinformatics/btr597>.
- University College London. (n.d.). *Competency frameworks – cognitive and behavioural therapy*. University College London. Retrieved from https://www.ucl.ac.uk/pals/sites/pals/files/all_problem-specific_competences.pdf (Accessed 12 February 2022).
- Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., ... Higgins, P. D. (2013). Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open*, 3(8), e002847. <http://doi.org/10.1136/bmjopen-2013-002847>.
- Wichniak, A., Wierzbicka, A., Wałęcka, M., & Jernajczyk, W. (2017). Effects of antidepressants on sleep. *Current Psychiatry Reports*, 19(9), 1–7. <https://doi.org/10.1007/s11920-017-0816-4>.
- Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(2), 180–185. <https://doi.org/10.1002/wics.147>.
- Ye, Y. Y., Chen, N. K., Chen, J., Liu, J., Lin, L., Liu, Y. Z., ... Jiang, X. J. (2016). Internet-based cognitive-behavioural therapy for insomnia (ICBT-i): A meta-analysis of randomised controlled trials. *BMJ Open*, 6(11), e010707. <http://doi.org/10.1136/bmjopen-2015-010707>.