

## Original Article

**Cite this article:** Skelton M *et al* (2023). Trajectories of depression and anxiety symptom severity during psychological therapy for common mental health problems. *Psychological Medicine* **53**, 6183–6193. <https://doi.org/10.1017/S0033291722003403>

Received: 30 November 2021

Revised: 22 April 2022

Accepted: 10 October 2022

First published online: 13 December 2022

**Key words:**

cognitive-behavioural therapy; counselling; dose-response; internalising; person-centred; psychotherapy; structural equation modelling; symptom change; treatment response

**Author for correspondence:**

Thalia C. Eley,  
E-mail: [thalia.eley@kcl.ac.uk](mailto:thalia.eley@kcl.ac.uk)

# Trajectories of depression and anxiety symptom severity during psychological therapy for common mental health problems

Megan Skelton<sup>1,2</sup> , Ewan Carr<sup>1</sup> , Joshua E. J. Buckman<sup>3,4</sup> ,  
Molly R. Davies<sup>1</sup> , Kimberley A. Goldsmith<sup>1</sup> , Colette R. Hirsch<sup>1,5</sup> ,  
Alicia J. Peel<sup>1</sup> , Christopher Rayner<sup>1</sup> , Katharine A. Rimes<sup>1</sup> ,  
Rob Saunders<sup>3</sup> , Janet Wingrove<sup>6</sup>, Gerome Breen<sup>1,2</sup>  and Thalia C. Eley<sup>1,2</sup> 

<sup>1</sup>Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; <sup>2</sup>National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, London, UK; <sup>3</sup>Research Department of Clinical, Centre for Outcomes Research and Effectiveness (CORE), Educational and Health Psychology, University College London, London, UK; <sup>4</sup>iCope – Camden and Islington Psychological Therapies Services, Camden and Islington NHS Foundation Trust, London, UK; <sup>5</sup>South London and Maudsley NHS Foundation Trust, London, UK and <sup>6</sup>Talking Therapies Southwark, South London and Maudsley NHS Foundation Trust, London, UK

**Abstract**

**Background.** There is substantial variation in patient symptoms following psychological therapy for depression and anxiety. However, reliance on endpoint outcomes ignores additional interindividual variation during therapy. Knowing a patient's likely symptom trajectories could guide clinical decisions. We aimed to identify latent classes of patients with similar symptom trajectories over the course of psychological therapy and explore associations between baseline variables and trajectory class.

**Methods.** Patients received high-intensity psychological treatment for common mental health problems at National Health Service Improving Access to Psychological Therapies services in South London ( $N = 16\,258$ ). To identify trajectories, we performed growth mixture modelling of depression and anxiety symptoms over 11 sessions. We then ran multinomial regressions to identify baseline variables associated with trajectory class membership.

**Results.** Trajectories of depression and anxiety symptoms were highly similar and best modelled by four classes. Three classes started with moderate-severe symptoms and showed (1) no change, (2) gradual improvement, and (3) fast improvement. A final class (4) showed initially mild symptoms and minimal improvement. Within the moderate-severe baseline symptom classes, patients in the two showing improvement as opposed to no change tended not to be prescribed psychotropic medication or report a disability and were in employment. Patients showing fast improvement additionally reported lower baseline functional impairment on average.

**Conclusions.** Multiple trajectory classes of depression and anxiety symptoms were associated with baseline characteristics. Identifying the most likely trajectory for a patient at the start of treatment could inform decisions about the suitability and continuation of therapy, ultimately improving patient outcomes.

**Background**

Routinely collected patient information can be used to explain some of the variability in outcomes following psychological therapies for depression and anxiety (Delgadillo, Moreea, & Lutz, 2016; Goddard, Wingrove, & Moran, 2015; Kessler *et al.*, 2017; Saunders, Buckman, & Pilling, 2020a). Most studies to date have focused on outcomes measured at the end of treatment, such as remission or recovery. Although clinically informative, reliance on endpoint outcomes may obscure interindividual differences in trajectories of symptoms occurring *during* treatment. Subgroups of patients may follow distinct trajectories, such as initially slow responders who nevertheless show clinically significant improvement by the end of treatment. Without a good understanding of these trajectories, treatments might be ended or altered early because patients, or their clinicians, consider the treatment unsuitable.

Person-centered structural equation modelling techniques can be used to reveal unobserved classes (groups or clusters) of individuals who exhibit similar longitudinal trajectories, which differ to those of the other classes. It is then possible to explore how baseline variables are associated with trajectory class membership. Relevant variables could be inspected at the start of treatment to identify a patient's most likely pattern of symptoms throughout therapy, and therefore inform patient and clinician expectations and decisions. Such information could also help monitor whether patients are doing less well than expected and thereby increase

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

positive outcomes (de Jong et al., 2021; Delgadillo et al., 2018; Lambert, 2007). Studies that have employed these methods with psychological treatment data include analyses of symptoms of depression (Lutz, Stulz, & Köck, 2009), panic disorder (Lutz et al., 2014) and post-traumatic stress disorder (PTSD; Dewar, Paradis, & Fortin, 2020), which each reported between three and five trajectory classes and some associations with baseline variables such as comorbid anxiety and depression symptoms. However, most existing analyses are arguably underpowered to reliably identify trajectory classes and associations, with samples of fewer than 500 participants. One study investigated symptom trajectories at two North London National Health Service (NHS) Improving Access to Psychological Therapies (IAPT) sites, in a larger sample of 4394 patients (Saunders et al., 2019). As part of the NHS in England, the IAPT services provide psychological therapy for adults experiencing common mental health problems such as depression and anxiety. A stepped-care framework is followed, offering evidence based low- and high-intensity treatments that primarily differ in terms of therapist involvement and number of sessions. The North London study analysed data from patients who received high-intensity therapy between 2008 and 2013 (Saunders et al., 2019). Four classes of depression symptom trajectories were identified, and five classes of anxiety symptom trajectories. Compared with trajectory classes showing improvement, classes with initially moderate-to-severe symptoms and limited change over time were associated with higher baseline levels of functional impairment, depression, anxiety, and phobia symptoms. No associations were found with age, gender, ethnicity, or prescribed psychotropic medication. In the present study, we extended these findings in a larger and more ethnically diverse sample of IAPT patients and considered additional baseline variables to test for association. We also used more recent data which is likely of higher quality due to improvements in recording patient information in IAPT (HSCIC, 2014; Saunders et al., 2020b).

## Method

### Sample

The sample came from anonymised patient records of routinely collected data from the four IAPT services of the South London and Maudsley (SLaM) NHS Foundation Trust. Using the Clinical Record Interactive Search (Stewart et al., 2009), we extracted treatment records of patients who had started and ended psychological therapy for symptoms of depression or anxiety between July 2014 and September 2020, irrespective of the reason for ending. From the 115 304 patients covering 587 120 sessions extracted, we identified a high-intensity treatment sample of 16 258 patients and 110 773 sessions suitable for analysis (see Fig. 1 and online Supplementary Information S1). IAPT high-intensity therapies are formal psychological treatments consistent with National Institute for Health and Care Excellence guidelines (NICE, 2011) and include disorder-specific cognitive-behavioral therapy (CBT; Roth and Pilling, 2008), counselling for depression, and other types such as interpersonal therapy (see the IAPT manual for details of therapy protocols and techniques; NCCMH, 2021). We modelled up to 11 sessions, including a baseline assessment, to ensure sufficient complete data across time points for the analysis and as modelling more sessions than a substantial proportion of patients had data for (see online Supplementary Table S1) could make optimal solutions invalid and unstable

(Lutz et al., 2005). To be representative of patients receiving treatment, we did not exclude patients scoring below clinical thresholds on depression and anxiety symptom scales.

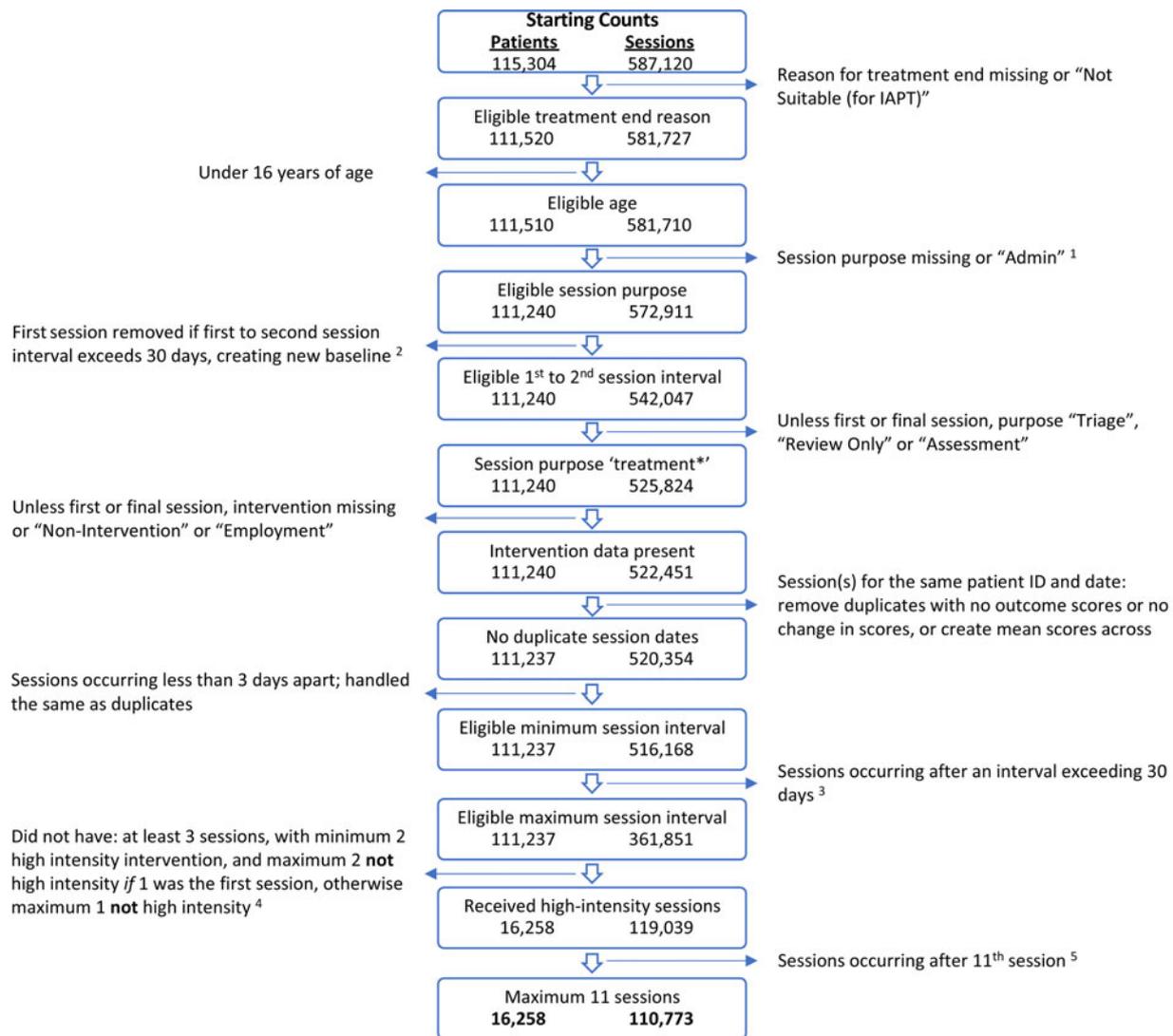
### Measures

Outcomes were self-report questionnaires administered at each session. Current symptoms of depression and anxiety were measured respectively by the Patient Health Questionnaire 9-item version (PHQ9) (Kroenke, Spitzer, & Williams, 2001; range 0-27; case threshold in IAPT  $\geq 10$ ) and the Generalised Anxiety Disorder 7-item scale (GAD7) (Spitzer, Kroenke, Williams, & Löwe, 2006; range 0-21; IAPT case threshold  $\geq 8$ ). Reliable improvement within trajectories was defined as a decrease of six or more points on the PHQ9, and at least four on the GAD7 (Gyani, Shafran, Layard, & Clark, 2013; NCCMH, 2021). We selected baseline self-reported patient characteristics to investigate for associations with trajectory class membership based on the available data and previously reported associations with treatment outcomes (Buckman et al., 2022; Delgadillo et al., 2016; Delgadillo, Dawson, Gilbody, & Böhnke, 2017a; Robinson, Kellett, & Delgadillo, 2020b; Saunders et al., 2019). These were: total scores from each of the PHQ9 and GAD7 (when not the outcome modelled), Work and Social Adjustment Scale (WSAS; functional impairment measure; range 0-40; Marks, 1986; Mundt, Marks, Shear, & Greist, 2002), age (in 10 year blocks), gender (female, male), ethnicity (White, Black, Asian, Mixed, other), employment (employed, unemployed, non-worker (e.g. retired, student)), psychotropic medication (not prescribed, prescribed) and disability (no, yes). In IAPT, 'problem descriptors' are used to indicate the disorder that is the agreed focus of treatment, in-line with NICE guidance (NICE, 2011). Problem descriptors are based on ICD-10 diagnostic codes, but their use does not mean that a patient necessarily met all diagnostic criteria for the disorder or that they would not meet criteria for other disorders. As such, comorbidity of disorders was unknown. Problem descriptor categories were depression, generalised anxiety disorder (GAD), PTSD, adjustment disorder, obsessive-compulsive disorder (OCD), mixed anxiety and depressive disorder, panic/phobia (panic disorder, agoraphobia, social phobia, specific phobia), and 'other' which included infrequent descriptors such as somatoform disorder. Despite showing strong associations in an existing study (Saunders et al., 2019), phobia scales were not included due to high, potentially non-random, missingness.

### Statistical analyses

#### Trajectory class models

We estimated separate growth mixture models (GMMs) for symptoms of depression and anxiety. For greater detail of the procedure, including all model selection criteria, see online Supplementary Information S2. To first determine the single trajectory form which best fit the observed data overall we compared linear (Bone, Delgadillo, & Barkham, 2021), quadratic (Phelps et al., 2018; Saunders et al., 2019) and negative log-linear (base 10) (Lutz et al., 2014) trajectories ('latent growth curves'). We then modelled up to six latent trajectory classes using GMM, each with the best-fitting form we had identified, to determine the optimal number of classes. This revealed whether the data was better explained by multiple latent growth curves than the single average one, indicating clusters of patients with similar trajectories rather than one group with homogeneous symptom



<sup>1</sup>Record of administrative activities; not treatment or assessment sessions with recorded outcomes.

<sup>2</sup>The first session in IAPT is often an assessment or triage, with treatment beginning formally at the second session. If the first to second session interval exceeded 30 days, the second session was used as the baseline. This was not repeated, as therapy with additional long intervals between the 'new' baseline and second session was considered too disrupted for stable conclusions.

<sup>3</sup>Growth mixture modelling assumes invariant time intervals across patients. High-intensity treatment is usually weekly but the data's naturalistic nature introduced substantial variance. To limit this, sessions after a 30-day interval were excluded. Online Supplementary Table S2 shows session interval descriptives with and without this filter.

<sup>4</sup>Limited to high-intensity therapies due to differences in the average number of sessions and level of structure between intensities. However, most IAPT patients receive at least one low-intensity session due to the stepped-care format, and the baseline assessment/triage session may also be labelled as low intensity.

<sup>5</sup>Including baseline assessment. Session number descriptives before and after this limitation, and the rationale for the limitation, are reported with online Supplementary Table S1.

**Fig. 1.** Flowchart detailing the exclusions applied to retain patient data suitable for analysis.

change. The upper limit of six classes was informed by past studies of trajectories during psychological therapy, which suggest between three and five classes (Dewar et al., 2020; Lutz et al., 2009; 2014; Saunders et al., 2019), thus the unlikely six class model acted as a ceiling to test against. We compared each model to a model with one fewer class using several fit indices, with a preference for the Bayesian information criterion (BIC; online Supplementary Information S2). The BIC was interpreted with the aid of an elbow plot to identify the point of diminishing gains from the addition of an extra class. We favoured models

that were interpretable in terms of clinical utility as well as past literature and theory. For statistical stability, we preferred models where each class contained more than 1% of the sample (Jung & Wickrama, 2008). We described patients in each class and the overlap between depression and anxiety classes.

We followed recommendations (Jung & Wickrama, 2008; Muthén & Muthén, 2000; van de Schoot, Sijbrandij, Winter, Depaoli, & Vermunt, 2017) to first perform a version of GMM with zero variance (latent class growth analysis; LCGA). This is a simpler, less computationally demanding model that makes

the often-unrealistic assumption that all individuals in a class follow exactly the same trajectory. This can result in trajectory classes that differ only in terms of starting score. We compared this simpler model to a GMM where the variance in each intercept was freed such that it could have a non-zero value but was constrained to be equal between classes, whilst the slope variance was fixed at zero. This allowed differences to be revealed in the patterns of change over time (slope of each class) rather than initial symptom levels (intercept of each class). This specification has successfully been used in previous symptom based GMMs (e.g. Lutz et al., 2014). Latent growth curve and LCGA results were similar between IAPT service-specific models (online Supplementary Figure S1) therefore analyses were performed across all four IAPT services. We used Mplus version 8.3 (Muthén & Muthén, 2017), alongside R version 3.6.3 (R Core Team, 2020) and the MplusAutomation package (Hallquist & Wiley, 2018).

### Missing data

Missing outcome data was handled using full information maximum likelihood with robust standard errors (Mplus option 'MLR') due to non-normal distributions of the outcomes (see online Supplementary Figure S2). If a patient had no observed data for the PHQ9 or GAD7 they were omitted from the model of that outcome. To determine support for generalisability of the results, patients included in the analytical sample were compared to excluded patients using suitable group difference tests. The large sample size could result in statistical significance of negligible differences, therefore we primarily considered effect sizes.

### Associations of baseline variables with trajectory class

Associations between variables measured at the start of treatment and trajectory class membership were tested using multinomial regressions. To handle missing data in the baseline variables, we ran multiple imputation using the 'mice' package in R, specifying 20 imputed datasets and 20 iterations (van Buuren & Groothuis-Oudshoorn, 2011). Explorations of missing and complete data between variables did not show any association patterns indicating non-random missingness. Patients were allocated to the trajectory class for which they had the highest probability of membership, and this 'most likely' class was the regression outcome. IAPT service was included as a covariate. We assessed statistical significance using a Bonferroni adjusted  $p$  value threshold of  $p < 0.025$  to account for the two independent models.

## Results

### Sample descriptives

Descriptives for the sample of 16 258 patients are in Table 1. The majority had received at least one session of CBT (50.1%) or counselling (49.1%); a small proportion received other treatments (6.1%) such as interpersonal therapy. The average number of sessions received, after limiting to a maximum of 11 including the baseline assessment, was 6.81 (s.d. = 2.67), with 11.6% receiving only the minimum 3 sessions and 18.4% receiving 11 (online Supplementary Table S1). At each session, over 99% of the patients remaining in treatment had complete PHQ9 and GAD7 scores (online Supplementary Table S3). Patients in the analytical sample were similar to excluded patients (online Supplementary Table S4), besides differences that reflected inclusion criteria (e.g. number of sessions).

### Growth mixture models

For both depression and anxiety symptoms, the best-fitting latent growth curve form was quadratic (online Supplementary Information S3). On average, patients' symptoms were initially moderate and improved steadily from baseline (session 0) over approximately the first five treatment sessions, showing reliable improvement by session eight. For LCGA, a four-class model was selected as the optimum solution for each outcome (online Supplementary Information S4). The trajectories were largely uninformative as they primarily differed in terms of baseline severity (intercept). We therefore estimated GMMs with free but equal variance in the intercepts; these are the focus of our results.

### Depression symptoms growth mixture model

For depression symptoms, a four-class GMM was selected based on fit indices alongside other considerations such as previous literature (online Supplementary Information S5). Figure 2a provides the mean trajectories of the four classes. Three classes showed moderate-severe symptoms at baseline. One of these showed no change (*moderate-severe plateau*; grey diamonds), the second had steady improvement (*moderate-severe, gradual improvement*; pink squares) and the third demonstrated fast improvement initially which plateaued after the sixth treatment session (*moderate-severe, fast improvement*; blue triangles). A fourth class showed mild symptoms and minimal improvement over the first four sessions which then plateaued (*mild, small improvement*; green circles). Approximately half the sample were assigned to this *mild* class and 14–18% in each of the other classes. Reliable improvement occurred on average by the seventh treatment session in the *gradual* class, and the second in the *fast* class, but was not observed in the *plateau* or *mild* class. At the final treatment session modelled, almost no (< 1%) *moderate-severe plateau* patients had 'recovered' to below PHQ9 and GAD7 case thresholds used in IAPT (< 10 and < 8, respectively). Recovery was also low in the *gradual improvement* class (18.7%), but higher in the *fast improvement* and *mild* classes (67.8% and 58.8%, respectively). Dropout was higher in the *gradual improvement* and *plateau* classes (20.3% and 18.9%) than *fast improvement* and *mild* classes (14.7% and 11.0%). The mean number of sessions was, however, similar across classes. The *plateau* class had the highest proportion referred to other services (11.5%). The full descriptives from the model are in online Supplementary Information S6.

### Anxiety symptoms growth mixture model

For anxiety symptoms, the GMM also suggested a four-class model (Fig. 2b; model selection and descriptives in online Supplementary Information S7 and S8). The trajectory classes for anxiety were very similar to that of depression and as such we used the same labels. However, more patients in the anxiety GMM were assigned to the *moderate-severe plateau* and *moderate-severe, fast improvement* classes, and fewer to the *moderate-severe, gradual improvement* and *mild, small improvement* classes. Reliable improvement occurred on average by the sixth treatment session in the *gradual* class and the second in the *fast* class but not in the *plateau* or *mild* class. Most patients who followed a *plateau* or *gradual improvement* trajectory had not recovered by the tenth treatment session (1.0% and 34.9% recovered, respectively), whilst most patients in the *fast improvement* and *mild* classes had (69.1% and 64.0%). The *plateau* class had the highest proportion of dropout (22.7%) and referral to other services (9.2%).

**Table 1.** Baseline descriptives of patients who received high-intensity psychological therapy for symptoms of depression and anxiety ( $N = 16\,258$ )

Variable	Mean (s.d.; range) or Count (Proportion %)
Age (years)	37.55 (13.36; 16–94)
Gender	
Female	10 979 (67.5%)
Male	5262 (32.4%)
Missing	17 (0.1%)
Depression symptoms (PHQ9)	13.98 (6.39; 0–27)
Missing	153 (0.9%)
Anxiety symptoms (GAD7)	12.53 (5.39; 0–21)
Missing	154 (1.0%)
Case on PHQ9 and/or GAD7 <sup>a</sup>	
Yes	13 735 (84.5%)
No	2368 (14.6%)
Missing	155 (1%)
Case on PHQ9 <sup>a</sup>	
Yes	11 873 (73.0%)
No	4232 (26.0%)
Missing	153 (0.9%)
Case on GAD7 <sup>a</sup>	
Yes	12 760 (78.5%)
No	3344 (20.6%)
Missing	154 (1.0%)
Functional impairment score (WSAS)	17.58 (9.31; 0–40)
Missing	5447 (33.5%)
Problem descriptor <sup>b</sup>	
Depression	6703 (41.2%)
Other	1423 (8.8%)
GAD	1393 (8.6%)
Adjustment disorder	1320 (8.1%)
PTSD	1132 (7.0%)
MADD	1129 (6.9%)
Panic/phobia	1003 (6.2%)
OCD	550 (3.4%)
Missing	1605 (9.9%)
Psychotropic medication	
Prescribed	5545 (34.1%)
Not prescribed	9977 (61.4%)
Missing	736 (4.5%)
Ethnicity	
White	9789 (60.2%)
Black	2964 (18.2%)
Mixed	1111 (6.8%)
Asian	961 (5.9%)
Other	557 (3.4%)

(Continued)

**Table 1.** (Continued.)

Variable	Mean (s.d.; range) or Count (Proportion %)
Missing	876 (5.4%)
Employment status	
Employed	10 033 (61.7%)
Unemployed	3572 (22.0%)
Non-worker <sup>c</sup>	2222 (13.7%)
Missing	431 (2.7%)
Disability <sup>d</sup>	
Yes	1575 (9.7%)
No	14 683 (90.3%)
Number of sessions (including baseline assessment)	6.81 (2.91; 3–11)
Recovered <sup>e</sup>	
Yes	5903 (36.3%)
No	7730 (47.5%)
Missing	2625 (16.1%)
Reason for end of treatment	
Discharge	13 138 (80.8%)
Dropout	2328 (14.3%)
Referral to another service	792 (4.9%)
Service	
0	7027 (43.2%)
1	3402 (20.9%)
2	3244 (20.0%)
3	2585 (15.9%)

Note: Only the baseline values are presented for variables that were measured at multiple time points. The 'Missing' row was omitted if there was no missing data.

<sup>a</sup>Case thresholds were PHQ9  $\geq 10$ , GAD7  $\geq 8$ .

<sup>b</sup>Indicates the disorder treated: GAD = generalised anxiety disorder; PTSD = post-traumatic stress disorder; MADD = mixed anxiety and depressive disorder; panic/phobia = panic disorder, agoraphobia, social phobia, specific phobia; OCD = obsessive-compulsive disorder; 'Other' included somatoform disorder, severe mental illness.

<sup>c</sup>'Non-worker' included homemaker, carer, retired, student.

<sup>d</sup>No negative responses were recorded; the absence of any value was counted as a negative response rather than missing.

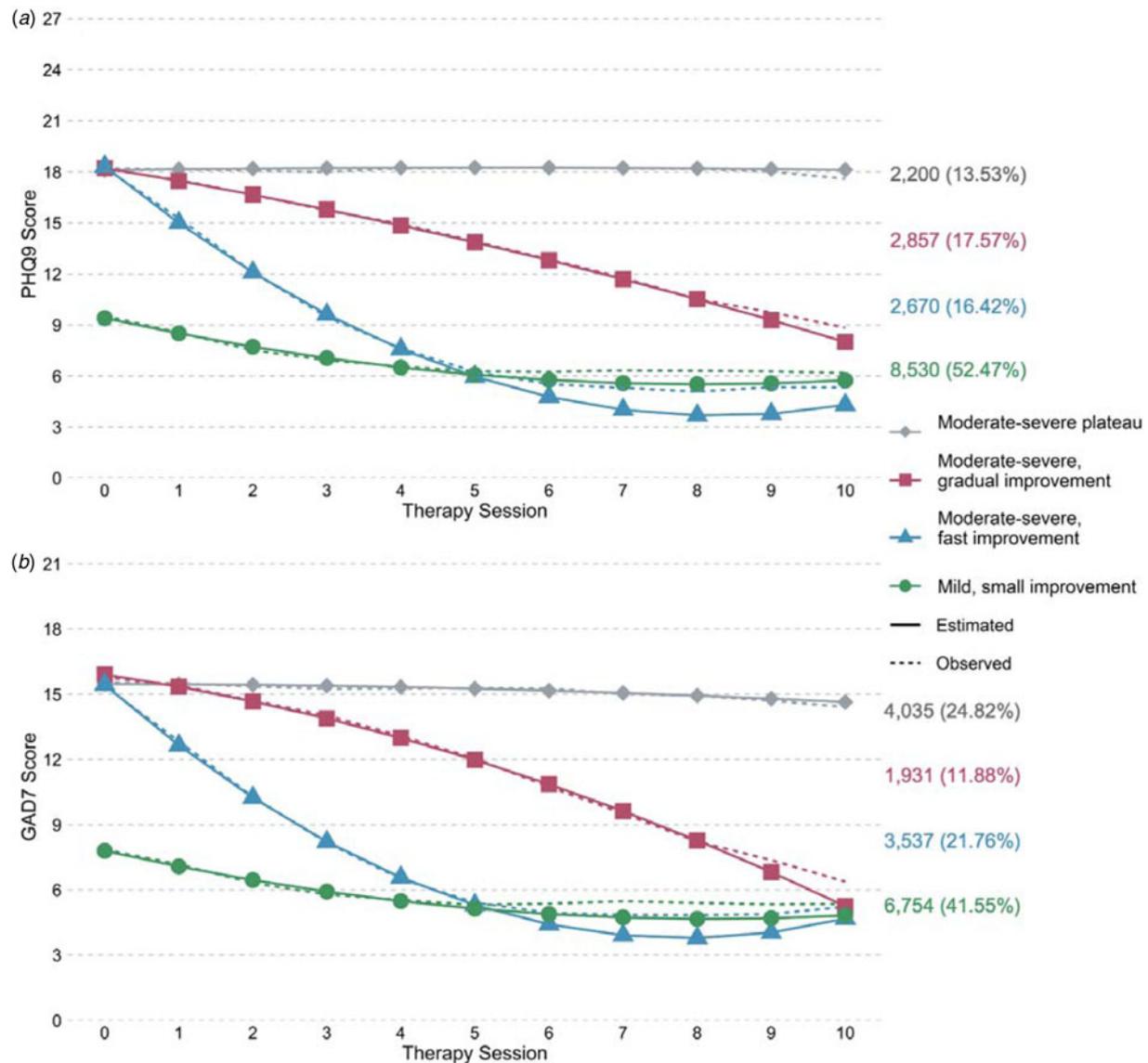
<sup>e</sup>Only calculated for patients above case thresholds on the PHQ9 or GAD7 at baseline and with observed scores for their final session, otherwise coded as missing. Represents whether the patient reached recovery within the 11 sessions modelled; if they received more sessions and then recovered, they would appear unrecovered here.

Online Supplementary Table S5 shows the overlap of trajectory classes between the depression and anxiety models. For example, 39.5% of patients in the depression *gradual improvement* class were in the anxiety *moderate-severe plateau* class, compared with 6.6% of the depression *fast improvement* class. In the anxiety model, 4.5% of the *gradual* and 0.8% of the *fast improvement* class were in the depression *plateau* class. Of the depression *plateau* class, 88.8% were in the anxiety *plateau* class, whilst 48.4% of the anxiety *plateau* class were in the depression *plateau* class.

#### Associations of baseline variables with trajectory class

##### Conditional growth mixture model of depression symptoms

Higher baseline anxiety was associated with lower odds of belonging to the *mild, small improvement* class, compared with



**Fig. 2.** Four-class growth mixture model of (a) depression symptoms (PHQ9) and (b) anxiety symptoms (GAD7), during high-intensity psychological therapy ( $N = 16\,258$ ).

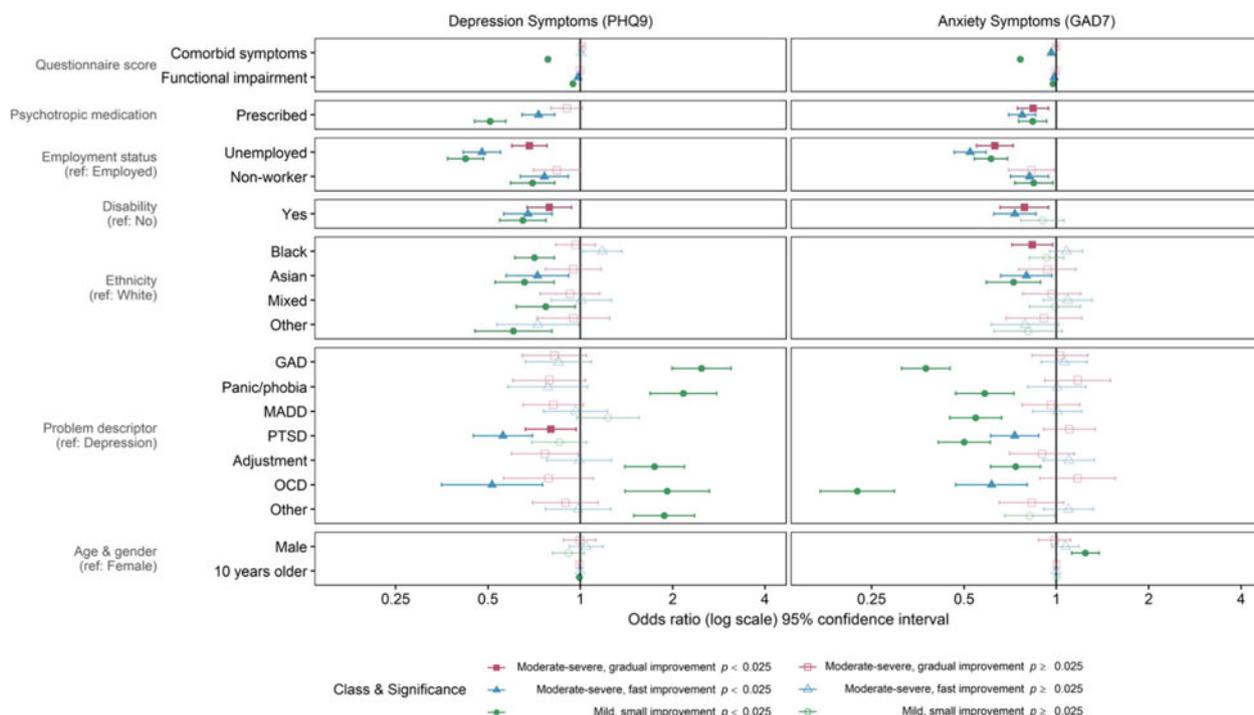
Note: This shows the model estimated mean and observed mean trajectories of each class. Patients had a likelihood of belonging to each trajectory class; counts and proportions (%) are based on their 'most likely' class membership.

the reference class *moderate-severe plateau* (Fig. 3, online Supplementary Table S6). Patients who had higher functional impairment, were prescribed psychotropic medication, or were not working (e.g. retired) compared with employed were less likely to be in the *mild* or the *moderate-severe, fast improvement* classes. Unemployed individuals or those reporting a disability had lower odds of being in the *fast* or *gradual improvement* or *mild* classes (i.e. the more favourable classes). Compared with White ethnicity patients, patients who identified as Black, Asian, multi-ethnic ('mixed'), or 'other' ethnicity were less likely to follow the *mild* trajectory compared with *moderate-severe plateau*. Patients identifying as Asian also had lower odds of belonging to the *fast improvement* class. Compared with patients with a depression problem descriptor, patients with GAD, panic/phobia, adjustment, OCD or 'other' disorder were more likely to be in the *mild* class. However, OCD and PTSD were associated, respectively, with decreased odds of the *fast improvement*, and

*fast* or *gradual improvement* trajectories. Age and gender had no notable associations.

#### Conditional growth mixture model of anxiety symptoms

Higher baseline depression or impairment scores and non-worker status were associated with decreased odds of being in the *moderate-severe, fast improvement* or *mild, small improvement* trajectory classes than the reference, *moderate-severe plateau* (Fig. 3, online Supplementary Table S7). Patients who were unemployed or prescribed psychotropic medication were less likely to belong to any favourable trajectory class. Reporting a disability was associated with decreased odds of the *fast* or *gradual improvement* classes. Patients identifying as Black, compared with White, were less likely to be in the *gradual improvement* class. Asian ethnicity was associated with lower odds of the *fast improvement* or *mild* class. Compared with a depression problem descriptor, all descriptors besides 'other' were associated with lower odds of the



**Fig. 3.** Multinomial regressions of trajectory class membership using baseline variables for a four-class growth mixture model of depression symptoms (PHQ9) and a four-class growth mixture model of anxiety symptoms (GAD7), during high-intensity psychological therapy ( $N = 16\,258$ ).

Note: The reference class in each model was moderate-severe plateau. Service, indicating where the patient received treatment, was included as a covariate with four categories. Error bars represent 95% confidence intervals. PHQ9 = Patient Health Questionnaire 9-item version; GAD7 = Generalised Anxiety Disorder 7-item scale. Comorbid symptoms refers to baseline GAD7 score in the depression model, and baseline PHQ9 score in the anxiety model. Functional impairment was measured by the Work and Social Adjustment Scale. Employment status ‘Non-worker’ included homemaker, carer, retired and student. GAD = generalised anxiety disorder; Panic/phobia = panic disorder, agoraphobia, social phobia, specific phobia; MADD = mixed anxiety and depressive disorder; PTSD = post-traumatic stress disorder; OCD = obsessive-compulsive disorder; ‘Other’ included somatoform disorder and severe mental illness.

*mild* class. Additionally, patients with PTSD or OCD were less likely to be in the *fast improvement* class. Patients reporting male gender were more likely to be in the *mild* class than females.

## Discussion

### Overview

We identified patterns of change in depression and anxiety symptoms during high-intensity psychological therapy among 16 258 patients. For both outcomes a quadratic slope was optimal and the data were best explained by four classes of trajectories whereby patients within a class exhibited similar trajectories. Baseline variables were associated with trajectory class membership.

### Growth mixture models

Both the depression and anxiety models had three classes with moderate-severe symptom scores at baseline. Using only intake symptom severity would therefore not allow a clinician to distinguish between three very different trajectories. One of these trajectories showed no change, labelled *moderate-severe plateau*. The other two classes reliably improved: *moderate-severe, gradual improvement* changing more steadily than *moderate-severe, fast improvement*. The fourth class, *mild, small improvement*, had average baseline symptoms below clinical thresholds, and thus less capacity for large improvements. These trajectory classes were broadly consistent in number and shape with a previous IAPT study (Saunders et al., 2019), despite several sample and

methodological differences in our study, including a greater proportion of patients receiving counselling and fewer receiving CBT.

The *gradual improvement* class may be particularly clinically relevant as the apparent lack of early response could lead to premature alteration or termination of treatment. Symptom change was initially indistinguishable from the *moderate-severe plateau* trajectory and only showed reliable improvement after six or seven sessions. ‘Recovery’ represents both depression and anxiety symptom scores falling below clinical thresholds by the final session. Recovery was lower in the *gradual* than the *fast improvement* class, especially for the depression model. Reflecting this, patients in the *gradual* class were less likely to experience improvement in the *other* outcome. This can be seen in the higher proportion of the depression *gradual* class than *fast improvement* class in the anxiety *plateau* class. The *gradual improvement* class might represent patients who require a high number of treatment sessions to show recovery. Supporting this, a study of high-intensity CBT found that most patients who significantly improved did so within 14 sessions (Robinson et al., 2020b) and a systematic review of therapy dose recommended up to 26 sessions for patients who have improved but not recovered before this point (Robinson, Delgado, & Kellett, 2020a).

The *moderate-severe plateau* class may also represent patients who would benefit from more treatment, although some ‘non-responders’ would probably not recover even after a very high number of sessions (Howard, Kopta, Krause, & Orlinsky, 1986). Notably, the frequency of onward referrals in this class suggested the presence of co-occurring problems that required specialist services. That said, despite limited symptom change, some patients

may have experienced other, unmeasured, benefits such as prevention of deterioration or hospitalisation. More patients belonged to the *plateau* class for the anxiety than the depression model. This may be driven by disorders such as OCD and PTSD, which on average require a greater dose of therapy to observe improvement than most depression and anxiety disorders (Robinson et al., 2020b), and are better detected by the GAD7 than the PHQ9 (Kroenke, Spitzer, Williams, Monahan, & Löwe, 2007). However, slightly over half of the anxiety *plateau* class showed mild or improving depression symptoms, whilst most patients in the depression *plateau* class showed no improvement on either measure as were also in the anxiety *plateau* class.

Patients in the *fast improvement* class generally appeared to require fewer than ten treatment sessions. They were below case thresholds after an average of four treatment sessions and had the highest recovery rate, in-line with meta-analytic evidence that reliable symptom improvement by the fourth session predicts recovery (Beard & Delgado, 2019). It is important to continue treatment for several sessions after remission to help prevent false positives or relapse (Robinson et al., 2020b), however, this finding warrants further investigation as it could have implications for service efficiency.

The mean initial symptom score estimated by the model for the *mild, small improvement* trajectory class, for both depression and anxiety models, was below case threshold, yet all patients had received high-intensity therapy. This was consistent with the observed baseline means. However, it would be inappropriate to conclude that high-intensity treatment was unwarranted for these patients. Scores varied around these means and many patients met the case threshold on the other symptom measure; the majority of *mild* class patients were an observed case on one or both measures. These patients may also have had additional symptoms not assessed by the PHQ9 or GAD7.

### Associations of baseline variables with trajectory class

The most consistent associations with trajectory class membership were prescribed medication, reporting a disability, and not being in employment, which were generally associated with lower odds of being in any of the more favourable classes compared with *moderate-severe plateau*. These associations are especially informative as they discriminated between individuals with similarly high baseline scores who improved or not. The association with employment status was consistent with findings from endpoint treatment outcome studies (Buckman et al., 2022; Delgado et al., 2016; Delgado, Huey, Bennett, & McMillan, 2017b). The prior IAPT trajectory study found no association with medication (Saunders et al., 2019) despite associations with treatment outcomes in endpoint studies (Robinson et al., 2020b). Neither employment or disability were included in the prior trajectory study, and the present findings highlight the importance of routinely recording them. These factors may negatively impact therapy response and acknowledging and supporting patients to navigate them might reduce the likelihood of a *moderate-severe plateau* trajectory. This supports the argument for psychological services working closely with employment advisors, as per the IAPT manual (NCCMH, 2021).

Higher functional impairment was associated with lower odds of the *fast improvement* or *mild* class than the *plateau* class, but not the *gradual improvement* class. This variable may therefore be useful for discriminating between *fast* and *gradual*

*improvement* for patients with similar baseline symptoms. The effect sizes for functional impairment were smaller than those for symptom scores, but still notable given they reflected the association for a one-unit difference on the WSAS, which has a wide scoring range. In the previous IAPT trajectory study, higher functional impairment and symptom scores similarly showed greater likelihood of the corresponding *moderate-severe plateau* class (Saunders et al., 2019). Our results were also consistent with a study reporting that patients with higher baseline scores on the PHQ9, GAD7, WSAS, prescribed antidepressants, a disability or unemployment were less likely to improve with an increasing number of sessions, similar to our *plateau* class (Robinson et al., 2020b).

The associations between *mild* trajectory class and problem descriptor in each model were ecologically valid. For example, patients were more likely to be in the *mild* trajectory class of the depression model if they had an anxiety descriptor compared with depression. Additionally, PTSD and OCD were associated with lower odds of a more favourable trajectory class. Improvement may have been observed after a greater number of sessions than modelled here, consistent with existing literature (Robinson et al., 2020b), or using measures more sensitive to symptoms of PTSD and OCD.

Patients who identified as one of the minoritised ethnic groups rather than White ethnicity were less likely to follow a favourable trajectory, especially the *mild* depression class, indicating that they had more severe symptoms at intake. These findings are consistent with lower recovery rates in national IAPT data and indicate that some patients from minoritised ethnic groups may benefit from culturally-adapted treatment (Arundell, Barnett, Buckman, Saunders, & Pilling, 2021; Beck, Naz, Brooks, & Jankowska, 2019). The ethnic diversity of the sample allowed us to study ethnicity in more detail than the binary variable used in similar studies and therefore identify that this pattern was most pronounced for Asian patients. That notwithstanding, there was insufficient power to analyse associations with ethnicity at an even greater level of granularity and this will require further research.

Overall, some variables were generally associated with greater likelihood of the *plateau* trajectory class than any more favourable class. These included reporting a disability, unemployment or prescribed medication. Other baseline characteristics differentiated between the more favourable classes, including non-worker status, functional impairment, specific problem descriptors and ethnicity.

### Strengths and limitations

A large sample size enabled us to produce more accurate estimates and include a broader range of ethnicity categories and baseline variables, such as employment and disability, compared with previous studies. These important and easily recorded factors could help identify likely symptom trajectories at the start of treatment. However, the results may be confounded by the presence of additional variables indicative of treatment prognosis, such as chronicity (Kessler et al., 2017; Lorenzo-Luaces, Rodriguez-Quintana, & Bailey, 2020). Whilst routinely collected data may generalise to clinical settings more readily than clinical trial data, the provision of different interventions, dependent on problem descriptor, makes IAPT data more complex to analyse. Investigations of treatment and descriptor-specific trajectories were beyond the scope of this study. The inclusion of patients with a depression

descriptor in the anxiety symptoms model, and vice versa, may therefore have contributed to the high proportion in the *mild* trajectory classes. Although comorbidity is common between depression and anxiety, the extent of comorbidity was unknown in the current sample as the IAPT problem descriptor only represents the condition being treated. That notwithstanding, the assessment of transdiagnostic symptoms appears particularly important for treatment outcomes (O'Driscoll et al., 2021) so we chose to include all patients in both models. Our findings therefore offer a broad representation of what we expect to see in routine clinical settings. The reliability and validity of IAPT problem descriptors has not been established. However, clinicians are trained to use ICD-10 to determine them, and in services with a higher frequency and accuracy of providing problem descriptors, patients are more likely to recover or reliably improve (Clark et al., 2018; Saunders et al., 2020b).

In terms of ethnicity, although generally representative of the local population that SLaM IAPT serves (Office for National Statistics, 2012), the patient sample was more ethnically diverse than IAPT nationally. The two most frequently reported ethnicities in the sample were White (60%) and Black (18%), and in IAPT national data are White (80%) and Asian (6%) (NHS Digital, 2021). It is critical that mental health and treatment outcome research is applicable for all patients, and oversampling from minoritised groups is one way of achieving this. In terms of ethnicity, this overrepresentation occurred naturally within our study, and we view the ethnic diversity of this sample as a strength. Additionally, although caution is recommended when using the IAPT measures across cultures (Beck et al., 2019), the PHQ9 and GAD7 have demonstrated good reliability and validity across a range of communities (see Beck et al., 2019 for a summary).

Another issue was dropout. Dropout is also observed in clinical trials and can occur due to different reasons, including dissatisfaction with treatment or recovery before the clinically recommended number of sessions. Regardless, and despite the use of missing data methods, scores at later sessions were based on remaining individuals. Replication in a sample receiving a fixed treatment length is therefore required.

Finally, the GAD7 does not cover core symptoms for disorders such as OCD and PTSD. However, it does show reasonably high specificity and sensitivity for several disorders, including these (Kroenke et al., 2007), and therefore provides some indication of likely trajectories for them. Future studies should aim to model trajectories of disorder specific symptoms, and also functional impairment, which is particularly important to patients and potentially improves later in treatment than symptoms (Howard, Lueger, Maling, & Martinovich, 1993). This was not possible in the present sample due to the extent of missingness on these variables. Modelling a higher number of sessions would also be beneficial, to reveal more about patients who require a greater dose of therapy.

### Implications and conclusions

Using a person-centered analysis allowed us to identify substantial heterogeneity in depression and anxiety symptom change during psychological therapy which could be explained by four classes of trajectories. The identified trajectory classes were differentially associated with baseline variables such as employment and ethnic group. To be implemented clinically, these findings need to be

extended into a predictive tool that combines baseline variables and outputs a patient's most likely symptom trajectory. Once validated in an independent dataset, this could be used to inform treatment plans at the start of therapy and also monitor whether patients are 'on-track' according to their predicted trajectory, which has been shown to improve patient outcomes (Delgado et al., 2018). This would be especially relevant for patients in the gradual improvement class who on average did not show reliable improvement until session six or seven. This study is therefore a crucial first step towards clinical use of trajectory classes to ultimately improve patient outcomes and service efficiency.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291722003403>

**Acknowledgments.** This study represents independent research funded by the National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre at SLaM NHS Foundation Trust and King's College London, and the Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or King's College London. The authors acknowledge the valuable work of the NIHR Maudsley Biomedical Research Centre Clinical Record Interactive Search team, especially Matthew Broadbent, who performed the data extraction.

**Author contribution.** All authors were involved in conceptualising the study. MS performed the data pre-processing and analysis with technical input from EC, JEJB, KAG, and RS. Clinical expertise was provided by JEJB, CRH, KAR, RS and JW. MS drafted the manuscript and all other authors provided critical feedback. All authors approved the final version prior to submission. TCE and GB supervised this work.

**Ethical standards.** We assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The use of anonymised data in this study was approved by the South Central - Oxford C Research Ethics Committee (15/SC/0586) and CRIS Oversight Committee (16010).

**Data availability.** This study used anonymised patient data from the SLaM NHS Foundation Trust that is not publicly available. The data can be accessed within a secure firewall via the Clinical Record Interactive Search (CRIS) tool by persons with the appropriate permissions. For more information please contact: [cris.administrator@slam.nhs.uk](mailto:cris.administrator@slam.nhs.uk).

**Grant numbers and disclosures.** MS was funded by an NIHR Maudsley Biomedical Research Centre studentship. JEJB was funded on a fellowship from the Wellcome Trust (201292/Z/16/Z). CR was supported by a grant from Fondation Peters to TCE and GB. AJP was supported by an ESRC studentship. GB has sat on advisory boards for Otsuka and Compass Pathways. The authors declare no conflict of interest.

### References

- Arundell, L. L., Barnett, P., Buckman, J. E., Saunders, R., & Pilling, S. (2021). The effectiveness of adapted psychological interventions for people from ethnic minority groups: A systematic review and conceptual typology. *Clinical Psychology Review*, 88, 102063. <https://doi.org/10.1016/j.cpr.2021.102063>.
- Beard, J. I. L., & Delgado, J. (2019). Early response to psychological therapy as a predictor of depression and anxiety treatment outcomes: A systematic review and meta-analysis. *Depression and Anxiety*, 36(9), 866–878. <https://doi.org/10.1002/da.22931>.
- Beck, A., Naz, S., Brooks, M., & Jankowska, M. (2019). *Improving access to psychological therapies (IAPT) Black, Asian and Minority Ethnic service user positive practice guide*. BABCP. Retrieved from <https://babcp.com/Portals/0/Files/About/BAME/IAPT-BAME-PPG-2019.pdf?ver=2020-06-16-004459-320>.

- Bone, C., Delgadillo, J., & Barkham, M. (2021). A systematic review and meta-analysis of the good-enough level (GEL) literature. *Journal of Counseling Psychology*, 68(2), 219–231. <https://doi.org/10.1037/cou0000521>.
- Buckman, J. E., Saunders, R., Stott, J., Cohen, Z. D., Arundell, L. L., Eley, T. C., ... Pilling, S. (2022). Socioeconomic indicators of treatment prognosis for adults with depression: A systematic review and individual patient data meta-analysis. *JAMA Psychiatry*, 79(5), 406–416. <https://doi.org/10.1001/jamapsychiatry.2022.0100>.
- 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).
- de Jong, K., Conijn, J. M., Gallagher, R. A. V., Reshetnikova, A. S., Heij, M., & Lutz, M. C. (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review*, 85, 102002. <https://doi.org/10.1016/j.cpr.2021.102002>.
- Delgadillo, J., Dawson, A., Gilbody, S., & Böhnke, J. R. (2017a). Impact of long-term medical conditions on the outcomes of psychological therapy for depression and anxiety. *The British Journal of Psychiatry*, 210(1), 47–53. <https://doi.org/10.1192/bjp.bp.116.189027>.
- Delgadillo, J., de Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., ... McMillan, D. (2018). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: A multisite, open-label, cluster randomised controlled trial. *The Lancet Psychiatry*, 5(7), 564–572. [https://doi.org/10.1016/S2215-0366\(18\)30162-7](https://doi.org/10.1016/S2215-0366(18)30162-7).
- Delgadillo, J., Huey, D., Bennett, H., & McMillan, D. (2017b). Case complexity as a guide for psychological treatment selection. *Journal of Consulting and Clinical Psychology*, 85(9), 835–853. <https://doi.org/10.1037/ccp0000231>.
- Delgadillo, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Behaviour Research and Therapy*, 79, 15–22. <https://doi.org/10.1016/j.brat.2016.02.003>.
- Dewar, M., Paradis, A., & Fortin, C. A. (2020). Identifying trajectories and predictors of response to psychotherapy for post-traumatic stress disorder in adults: A systematic review of literature. *The Canadian Journal of Psychiatry*, 65(2), 71–86. <https://doi.org/10.1177/0706743719875602>.
- Goddard, E., Wingrove, J., & Moran, P. (2015). The impact of comorbid personality difficulties on response to IAPT treatment for depression and anxiety. *Behaviour Research and Therapy*, 73, 1–7. <https://doi.org/10.1016/j.brat.2015.07.006>.
- Gyani, A., Shafraan, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: Lessons from year one of IAPT. *Behaviour Research and Therapy*, 51(9), 597–606. <https://doi.org/10.1016/j.brat.2013.06.004>.
- Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R package for facilitating large-scale latent variable analyses in mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4), 621–638. <https://doi.org/10.1080/10705511.2017.1402334>.
- Health and Social Care Information Centre (HSCIC). (2014). *Announcement of methodological change: Improving Access to Psychological Therapies (IAPT) Reports*. Community and Mental Health Team, Health and Social Care Information Centre.
- Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *The American Psychologist*, 41(2), 159–164. <https://www.ncbi.nlm.nih.gov/pubmed/3516036>.
- Howard, K. I., Lueger, R. J., Maling, M. S., & Martinovich, Z. (1993). A phase model of psychotherapy outcome: Causal mediation of change. *Journal of Consulting and Clinical Psychology*, 61(4), 678–685. <https://doi.org/10.1037//0022-006x.61.4.678>.
- Jung, T., & Wickrama, K. A. S. (2008). An Introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, 2, 302–317. <https://doi.org/10.1111/j.1751-9004.2007.00054.x>.
- Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Ebert, D. D., ... Zaslavsky, A. M. (2017). Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiology and Psychiatric Sciences*, 26(1), 22–36. <https://doi.org/10.1017/S2045796016000020>.
- 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>.
- Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine*, 146(5), 317–326. <https://doi.org/10.7326/0003-4819-146-5-200703060-00004>.
- Lambert, M. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 17(1), 1–14. <https://doi.org/10.1080/10503300601032506>.
- Lorenzo-Luaces, L., Rodriguez-Quintana, N., & Bailey, A. J. (2020). Double trouble: Do symptom severity and duration interact to predicting treatment outcomes in adolescent depression? *Behaviour Research and Therapy*, 131, 103637. <https://doi.org/10.1016/j.brat.2020.103637>.
- Lutz, W., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M. K., Gorman, J. M., ... Barlow, D. H. (2014). Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. *Journal of Consulting and Clinical Psychology*, 82(2), 287–297. <https://doi.org/10.1037/a0035535>.
- Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., ... Iveson, S. (2005). Predicting change for individual psychotherapy clients on the basis of their nearest neighbors. *Journal of Consulting and Clinical Psychology*, 73(5), 904–913. <https://doi.org/10.1037/0022-006X.73.5.904>.
- Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders*, 118(1-3), 60–68. <https://doi.org/10.1016/j.jad.2009.01.019>.
- Marks, I. M. (1986). *Behavioural psychotherapy: Maudsley pocket book of clinical management*. Bristol: Wright/Institute of Psychiatry Publishing.
- 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. <https://doi.org/10.1192/bjp.180.5.461>.
- Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, 24(6), 882–891. <https://doi.org/10.1111/j.1530-0277.2000.tb02070.x>.
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- National Collaborating Centre for Mental Health (NCCMH). (2021). *The improving access to psychological therapies manual* (version 5). National Collaborating Centre for Mental Health. Retrieved from <https://www.england.nhs.uk/wp-content/uploads/2018/06/the-iapt-manual-v5.pdf>.
- National Institute for Health and Care Excellence (NICE). (2011). *Common mental health disorders: Identification and pathways to care CG123*. London: NICE. <http://guidance.nice.org.uk/CG123>.
- NHS Digital. (2021). Psychological Therapies, Annual report on the use of IAPT services, 2020-21. Retrieved 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. E., 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), 1–14. <https://doi.org/10.1186/s12916-021-01971-0>.
- Office for National Statistics. (2012). *2011 Census: Key statistics for local authorities in England and Wales*. Retrieved from <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/2011censuskeystatisticsforlocalauthoritiesinenglandandwales>.
- Phelps, A. J., Steel, Z., Metcalf, O., Alkemade, N., Kerr, K., O'Donnell, M., ... Forbes, D. (2018). Key patterns and predictors of response to treatment for military veterans with post-traumatic stress disorder: A growth mixture modelling approach. *Psychological Medicine*, 48(1), 95–103. <https://doi.org/10.1017/S0033291717001404>.
- R Core Team. (2020). *R: A language and environment for statistical computing* (Version 3.6.3). Vienna, Austria: The R Foundation for Statistical Computing Platform. <https://www.R-project.org/>.

- Robinson, L., Delgadillo, J., & Kellett, S. (2020a). The dose-response effect in routinely delivered psychological therapies: A systematic review. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 30(1), 79–96. <https://doi.org/10.1080/10503307.2019.1566676>.
- Robinson, L., Kellett, S., & Delgadillo, J. (2020b). Dose-response patterns in low and high intensity cognitive behavioral therapy for common mental health problems. *Depression and Anxiety*, 37(3), 285–294. <https://doi.org/10.1002/da.22999>.
- Roth, A. D., & Pilling, S. (2008). Using an evidence-based methodology to identify the competences required to deliver effective cognitive and behavioural therapy for depression and anxiety disorders. *Behavioural and Cognitive Psychotherapy*, 36(2), 129–147. <https://doi.org/10.1017/S1352465808004141>.
- 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. *Journal of Affective Disorders*, 249, 327–335. <https://doi.org/10.1016/j.jad.2019.02.043>.
- Saunders, R., Buckman, J. E. J., & 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? *Cognitive Behaviour Therapy*, 13, e16. <https://doi.org/10.1017/S1754470X20000173>.
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, 166(10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>.
- Stewart, R., Soremekun, M., Perera, G., Broadbent, M., Callard, F., Denis, M., ... Lovestone, S. (2009). The South London and Maudsley NHS foundation trust biomedical research centre (SLAM BRC) case register: Development and descriptive data. *BMC Psychiatry*, 9, 51. <https://doi.org/10.1186/1471-244X-9-51>.
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>.
- van de Schoot, R., Sijbrandij, M., Winter, S. D., Depaoli, S., & Vermunt, J. K. (2017). The GRoLTS-checklist: Guidelines for reporting on latent trajectory studies. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(3), 451–467. <https://doi.org/10.1080/10705511.2016.1247646>.