nature mental health

Supplementary information

https://doi.org/10.1038/s44220-025-00567-4

Symptom change in depression and anxiety during psychological therapy for autistic adults

In the format provided by the authors and unedited

Supplement: Symptom Change in Depression and Anxiety During Psychological Therapy for Autistic Adults

Contents

Supplementary Table 1: The stepped-care model in NHS Talking Therapies for Anxiety and Depression (TTad), previously named Improving Access to Psychological Therapies (IAPT)

Intervention step within stepped-care model	**Condition**	**Intervention**
Step 1: Assessment		Recognition of problem Assessment / watchful waiting
Step 2: Low-intensity interventions	Mild-moderate: 1. Depression 2. Panic disorder 3. Generalised Anxiety Disorder (GAD) 4. Obsessive Compulsive Disorder (OCD)	1. Computerised Cognitive Behavioural Therapy (cCBT), guided self-help, Behavioural Activation (BA) and exercise 2. cCBT, guided self-help, pure self-help 3. cCBT, guided self-help, psycho-education groups, pure self-help 4. Guided self-help
Step 3: High-intensity interventions	Moderate-severe (and mild-moderate if no improvement at Step 2): 1. Depression 2. Panic Disorder 3. GAD 4. Social phobia (mild-severe) 5. Post-traumatic stress disorder (PTSD) (mild-severe) 6. OCD	1. CBT, Interpersonal Therapy (IPT), BA, Dynamic Interpersonal Therapy (DIT), couples' therapy and counselling for depression 2. CBT 3. CBT 4. CBT 5. CBT, Eye Movement Desensitisation and Reprocessing (EMDR) 6. CBT

Information about the TTad stepped-care model is presented above in Supplementary Table 1. Detailed information about the TTad treatment procedure is available in the IAPT manual (34). TTad follows a stepped-care model.

Service users included in the dataset were initially seen for an assessment to ascertain the primary presenting problem. Following this, individuals with mild to moderate levels of difficulty were commonly offered low-intensity (short-term, guided self-help) intervention in individual or group format. Where clients did not respond to this treatment, or where initial presenting difficulties showed greater severity or complexity, they were offered a high-intensity treatment (16-20 session individual evidence-based therapy).

Supplementary Table 2: Information on variables and measures

Measure	Source	Information
Primary outcome measures		
Patient Health Questionnaire-9 (PHQ-9) (17)	TTad	The PHQ-9 is a 9-item depression screening tool and questionnaire, rated using a Likert scale (0-3). The PHQ-9 was routinely collected at each appointment. The PHQ-9 has been used extensively in general population research and in clinical settings, with good validity and reliability (17).
Generalized Anxiety Disorder Assessment-7 (GAD-7) (18)	TTad	The GAD-7 is a 7-item screening tool and questionnaire for Generalised Anxiety Disorder (GAD), rated using a Likert scale (0-3), and was routinely collected at each appointment. The GAD-7 has shown good reliability and validity, excellent internal consistency, and good sensitivity and specificity in detecting clinical anxiety (18).
Secondary outcome measures		
Work and Social Adjustment Scale (WSAS) (22)	TTad	The WSAS is a 5-item questionnaire assessing impact on i) work, ii) home management, iii) social leisure activities, iv) private leisure activities and v) close relationships. It is rated on a Likert scale (0-8), and was routinely collected at each appointment. It has shown good test-retest reliability and internal consistency, and is sensitive to disorder severity and treatment-related change (22). In TTad, clients are instructed to answer "N/A" (not applicable) to question i) (work) if they are retired or unemployed. Hence in the present study, this subscale was excluded and total WSAS scores were computed from questions 2-5 (more below).
Demographic and clinical variables		
Age	TTad, HES, MHSDS	Calendar age in years, at the point at which the referral was received.
Sex	TTad, HES, MHSDS	A binary variable, identified by the service user (male/female). Sex (with a wider range of options) was not available in this dataset.
Presenting problem	TTad	A categorical variable (anxiety, depression, or 'mixed anxiety and depression'), reflecting the primary focus for treatment assigned to the service user
Ethnicity	TTad, HES, MHSDS	A binary variable (White, ethnically minoritised), created from a range of categories identified by the service user due to low cell counts in many categories.
Intellectual disability	MHSDS	A binary variable, identified from relevant MHSDS ICD-10 codes (yes/no).
Employment	TTad, HES, MHSDS	A binary variable (in employment, not in employment), created from a range of categories identified by the service user
Psychotropic medication	TTad, HES, MHSDS	A binary variable (taking psychotropic medication, not taking psychotropic medication), identified by the service user
Long-term health condition	TTad, HES, MHSDS	A binary variable (has a long-term health condition, does not have a long-term health condition), identified by the service user and healthcare records
Recovery metrics		
Reliable improvement	TTad	A binary variable, calculated to indicate whether reduction in PHQ-9 or GAD-7 scores between first and last measurement points exceeds the error threshold for the measurement scale (PHQ-9 change > 5; GAD-7 change > 3).
Reliable recovery	TTad	A binary variable, calculated to indicate the presence of i) reliable improvement and ii) the final measurement score is below the "caseness" threshold for the measure.
Reliable deterioration	TTad	A binary variable, calculated to indicate the inverse of reliable improvement, i.e., increase in scores between first and last measurement points that exceeds the error threshold for the measure.

Measurement information for variables from MODIFY used in the present study are shown in Supplementary Table 2. The TTad dataset includes routinely collected data for all service users seen in IAPT services across England between 2012 and 2019. For each person, it includes demographic information, and information on the intervention received (e.g., dates seen, treatment delivered and measures collected). At each treatment appointment, service users were asked by clinical staff to complete a set of standardised measures that includes

the Generalized Anxiety Disorder Assessment-7 (GAD-7) (18), the Patient Health Questionnaire-9 (PHQ-9) (17) and the Work and Social Adjustment Scale (WSAS)(22).

The Hospital Episode Statistics (HES) dataset includes individual-level data from care records for inpatient, outpatient and Accident and Emergency (A&E) treatment from all National Health Service (NHS) hospitals across England. It includes demographic, geographical, administrative and clinical information (35).

The Mental Health Services Dataset (MHSDS) includes individual-level data from secondary care services (i.e., outpatient and community specialist services), and includes diagnostic codes relating to mental health conditions and neurodevelopmental conditions (including autism and Intellectual Disability) (37)

The Hospital Episode Statistics-Office of National Statistics (HES-ONS) dataset linked information from the HES and Office of National Statistics (ONS) mortality data. It includes individual-level information about the cause, date and place of death (36).

Data from HES, MHSDS and HES-ONS were only retrieved for individuals with TTad data available. All data were fully anonymised, and linkage was achieved using anonymised subject identifiers provided by NHS Digital.

Supplementary Table 3: Ethnicity data

Office for National Statistics Ethnicity Category	Frequency (N)	Percent (%)
White	6,107	93.97
Mixed	145	2.23
Asian	135	2.08
Black	85	1.31
Chinese	5	.08
Other	22	.34

Cell counts for Office for National Statistics (ONS) ethnicity categories are shown in Supplementary Table 3. We needed to treat ethnicity as a binary variable to include it in the R3Step analysis and this is acknowledged in the Limitations section of the report.

Supplementary Table 4: Days between baseline and timepoints 2-8

	Mean (SD)	Median (IQR)	Range
Days between sessions	20.48 (13.94)	17 (11.86 – 24.61)	2 – 133.5
Days between baseline and timepoint 8	163.84 (111.52)	136 (94.86 – 196.89)	16 - 1,796

The mean, median and range of days between psychological therapy sessions are presented above in Supplementary Table 4. PHQ-9 and GAD-7 measurements are attempted at each session in a TTad setting. The TTad dataset contains data for up to 30 measurements. In this study, 8 timepoints were modelled.

The date of the first and last appointments were available, and were used to generate the number of days in treatment for each participant. Dividing this by the number of sessions received, created an estimate of the average time between sessions for each participant.

Supplementary Table 5: Work and Social Adjustment Scale (WSAS)(22) missing and not applicable (N/A) responses (N = 7,175)

WSAS subscale	Missing	Coded N/A	N (baseline)
Employment	2,985	1,219	2,971
Home management	2,945	6	4,224
Social leisure activities	2,955	6	4,214
Private leisure activities	2,949	6	4,226
Close relationships	2,947	7	4,221

Missing and "not applicable" (N/A) responses on the WSAS are presented in Supplementary Table 5 above. The WSAS total score constitutes a measure of functional impairment (difficulties in daily living), comprising employment, home management, social leisure activities, private leisure activities and close relationships (22). In TTad settings, clients who are retired or not in employment are instructed to mark the employment subscale as "N/A" (not applicable), coded "9" in the TTad dataset (34). In the present sample, N>1,000 participants coded the employment question as "N/A" For this reason and in keeping with prior research using TTad data (15, 16, 39), the WSAS data used in this study excluded the employment subscale. Total scores were computed by summing home management, social leisure, private leisure and close relationship subscales. Sensitivity analyses that included the employment subscale were conducted and reported in Supplementary Table 16 (depression) and Supplementary Table 17 (anxiety). To aid comparability, both age and total WSAS scores were standardised before being entered into the R3Step regression analysis (42),

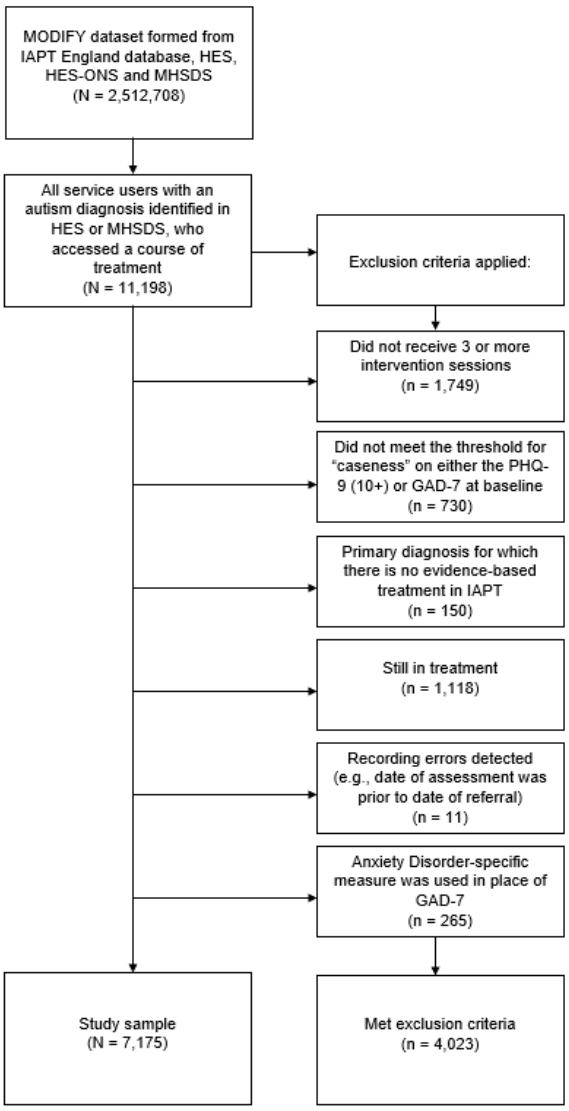
Supplementary Table 6: Fit indices used to assess growth model fit

Index	Step	Information
Root Mean Square Error of Approximation (RMSEA)	LGCM	An absolute fit index, which assesses how far a hypothesised model is from a perfect model. RMSEA values <.05 indicate a good fit, <.08 a reasonable fit and <.10 a mediocre fit (45).
Comparative Fit Index (CFI), Tucker-Lewis Index (TLI)	LGCM	The CFI and TLI are incremental fit indices, which compare the fit of a hypothesised model with that of a baseline model. CFI and TLI values >.95 indicates a good fit and >.90 an acceptable fit (46).
Standardised Root Mean Square Residual (SRMR)	LGCM	A measure of the mean absolute correlation residual. SRMR values <.05 indicates a good fit, and <.08 a reasonable fit (46).
Bayesian Inference Criteria (BIC), Sample Size-Adjusted BIC (SA-BIC)	LGCM, LCGA, GMM	The BIC and SABIC use the log likelihood value, and penalise based on the number of parameters. The SABIC also penalises based on the sample size. BIC and SABIC are commonly used to compare growth mixture models with increasing class numbers, and the model with the lowest value would typically be deemed optimal (47).
Vuong-Lo-Mendell-Rubin Likelihood Ratio Test (VLMR-LRT), Lo-Mendell-Rubin Adjusted Likelihood Ratio Test (LMR-A LRT), Bootstrap Likelihood Ratio Test (BLRT)	LCGA, GMM	The VLMR-LRT, LMR-A-LRT and BLRT compare the ratio of log-likelihoods for a k-class model with the k-1 class model, and report a significance test to clarify whether the addition of the *k*th class made an improvement (48).

Key: LGCM = Latent Growth Curve Model; LCGA = Latent Class Growth Analysis; GMM = Growth Mixture Model

The range of fit indices designed to test models are presented above in Supplementary Table 6, along with their thresholds and the selection process applied.

Supplementary Figure 1: Study flow diagram

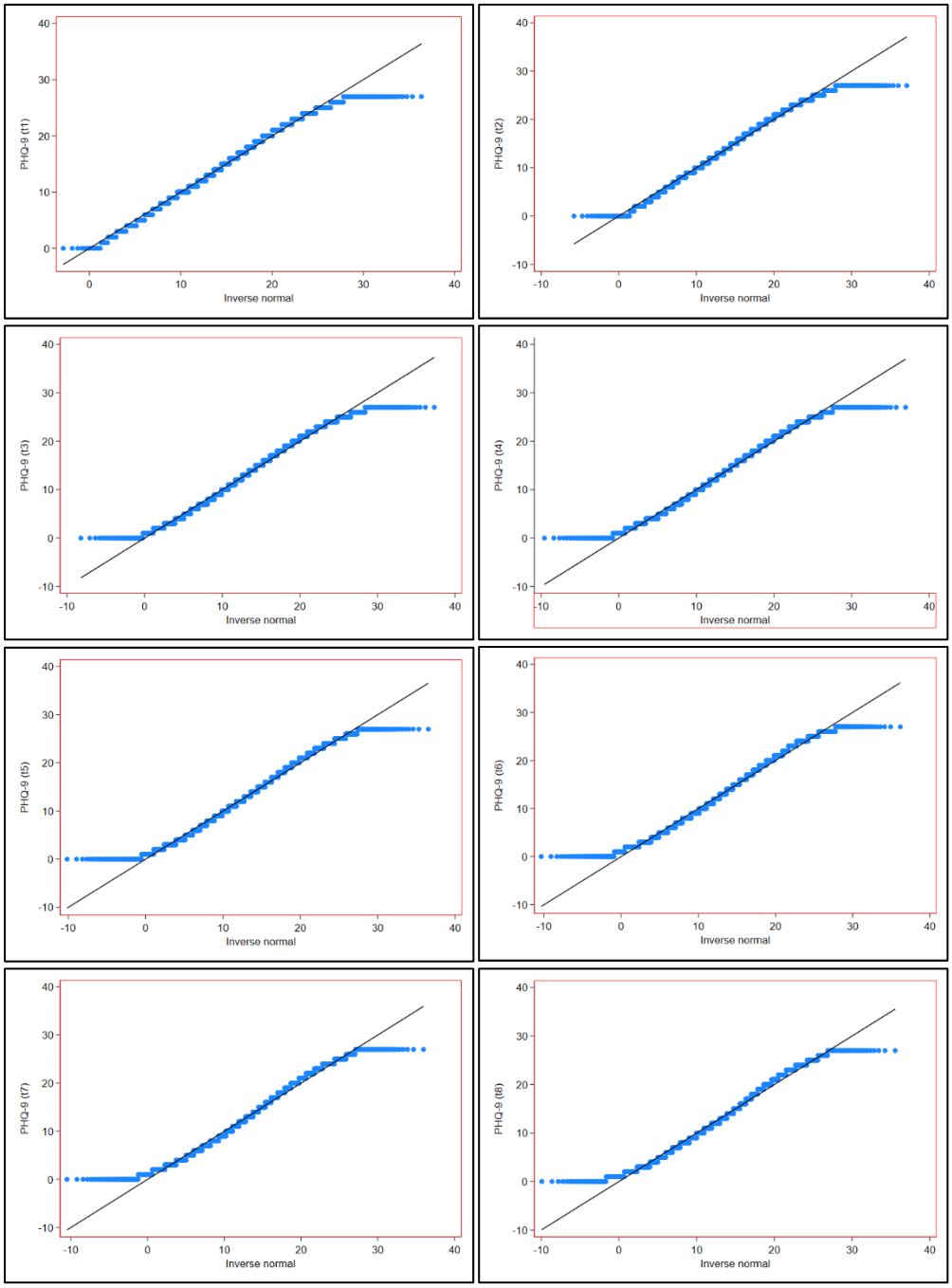


Supplementary Table 7: Distributions of depression and anxiety data

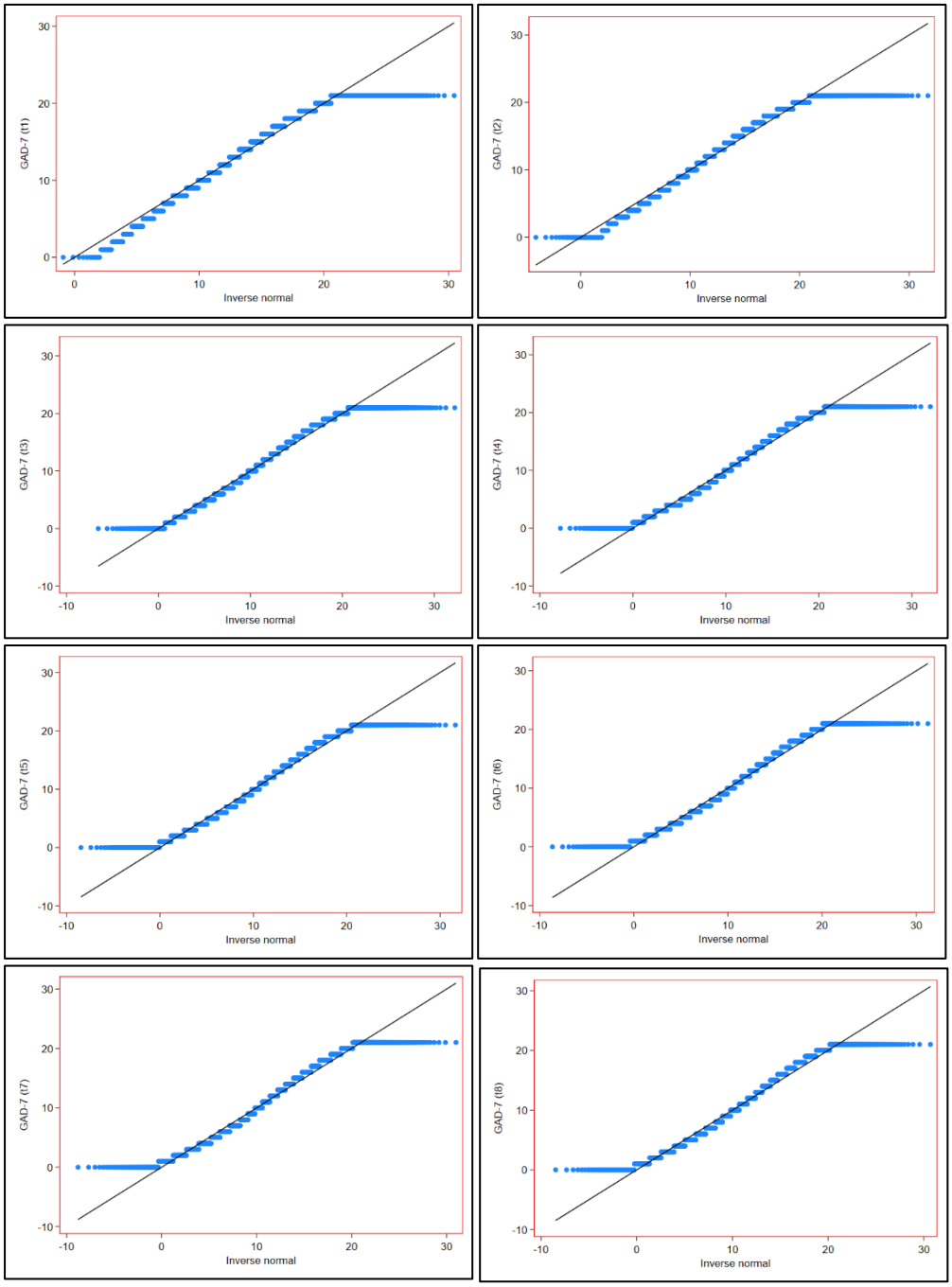
Variable	N	Mean (SD)	Median (IQR)	Skewness	Kurtosis	Doornik-Hansen χ^2 (*p*)
PHQ-9:						
t1	6,969	16.75 (5.41)	17 (13-21)	-.28	2.58	237.11 (<.01)
t2	6,946	15.66 (5.91)	16 (12-20)	-.26	2.53	227.65 (<.01)
t3	6,919	14.55 (6.28)	15 (10-19)	-.13	2.31	230.14 (<.01)
t4	5,998	13.69 (6.48)	14 (9-18)	.00	2.25	190.19 (<.01)
t5	5,157	13.20 (6.58)	13 (8-18)	.05	2.26	161.24 (<.01)
t6	4,308	12.90 (6.65)	13 (8-18)	.11	2.22	173.30 (<.01)
t7	3,485	12.72 (6.75)	12 (7-18)	.14	2.17	168.27 (<.01)
t8	2,756	12.81 (6.73)	12 (8-18)	.15	2.22	121.74 (<.01)
GAD-7:						
t1	6,966	14.77 (4.32)	15 (12-18)	-.53	2.74	682.27 (<.01)
t2	6,944	13.81 (4.94)	14 (10-18)	-.48	2.53	723.48 (<.01)
t3	6,919	12.84 (5.35)	13 (9-17)	-.31	2.23	542.44 (<.01)
t4	6,002	12.10 (5.54)	12 (8-17)	-.16	2.06	407.05 (<.01)
t5	5,162	11.59 (5.65)	12 (7-16)	-.09	2.05	313.32 (<.01)
t6	4,311	11.30 (5.70)	11 (7-16)	-.02	2.03	252.65 (<.01)
t7	3,490	11.09 (5.78)	11 (7-16)	-.01	2.01	218.54 (<.01)
t8	2,758	11.10 (5.79)	11 (7-16)	.00	2.02	168.05 (<.01)

Supplementary Table 7 above presents descriptive statistics for the PHQ-9 and GAD-7 by each timepoint. Data were negatively-skewed and leptokurtic, but skewness and kurtosis were deemed to fall within acceptable bounds. Q-Q plots are presented in Supplementary Figures 2 and 3 below.

Supplementary Figure 2: Q-Q plots representing distribution of depression data across timepoints



Supplementary Figure 3: Q-Q plots representing distribution of anxiety data across timepoints



Supplementary Table 8: Logistic regression of missing depression and anxiety data, with covariates

Variable	Baseline OR (95% CI)	t8 OR (95% CI)
PHQ-9:		
Sex	.78 (.59 – 1.03)	1.08 (.98 – 1.19)
Ethnicity	1.84 (.81 – 4.19)	1.29 (1.05 – 1.58)
Employment status	.75 (.53 – 1.04)	.82 (.74 - .90)
Presenting problem (depression, anxiety, mixed)	1.00 (.98 – 1.03)	.99 (.98 – 1.00)
Long-term condition	.71 (.50 – 1.02)	1.07 (.96 – 1.20)
Intellectual Disability	1.27 (.59 – 2.72)	.82 (.65 – 1.05)
GAD-7:		
Sex	.83 (.63 – 1.10)	1.08 (.98 – 1.19)
Ethnicity	1.88 (.83 – 4.26)	1.30 (1.06 – 1.60)
Employment status	.75 (.53 – 1.04)	.82 (.74 - .90)
Presenting problem (depression, anxiety, mixed)	1.00 (.97 – 1.02)	.99 (.98 – 1.00)
Long-term condition	.76 (.53 – 1.09)	1.07 (.96 – 1.20)
Intellectual Disability	1.29 (.61 – 2.76)	.82 (.65 – 1.05)

Supplementary Table 8 above reports the association between missingness of depression and anxiety data and covariates (sex, ethnicity, employment status, primary diagnosis, long-term condition and Intellectual Disability). Odds Ratios revealed that no covariates were significantly related to data missingness at baseline, but ethnically-minoritised ethnicity was significantly related to greater odds of missing data by the final timepoint. Data were considered to be Missing at Random (MAR), and the increased likelihood of dropout for individuals from ethnically-minoritised backgrounds was accounted for in subsequent interpretation.

Supplementary Figure 4: Model specification (latent growth curve model, time-fixed)

```
VARIABLE:    Names = d1 d2 d3 d4 d5 d6 d7 d8 a1 a2 a3 a4 a5 a6 a7 a8 id;
             Usevariables = d1 d2 d3 d4 d5 d6 d7 d8;
             Idvariable = id;
             Missing = all(-9999);
MODEL:       Int Slo | d1@0 d2@1 d3@2 d4@3 d5@4 d6@5 d7@6 d8@7;
```

Supplementary Figure 5: Model specification (latent growth curve model, time-freed)

```
VARIABLE:    Names = d1 d2 d3 d4 d5 d6 d7 d8 a1 a2 a3 a4 a5 a6 a7 a8 id;
             Usevariables = d1 d2 d3 d4 d5 d6 d7 d8;
             Idvariable = id;
             Missing = all(-9999);
MODEL:       Int Slo | d1@0 d2@1 d3* d4* d5* d6* d7* d8*;
```

Two model specifications were designed for LGCM, and the syntax is shown above in Supplementary Figure 4 (time intervals between session modelled as fixed) and Supplementary Figure 5 (time intervals between sessions allowed to be freely estimated).

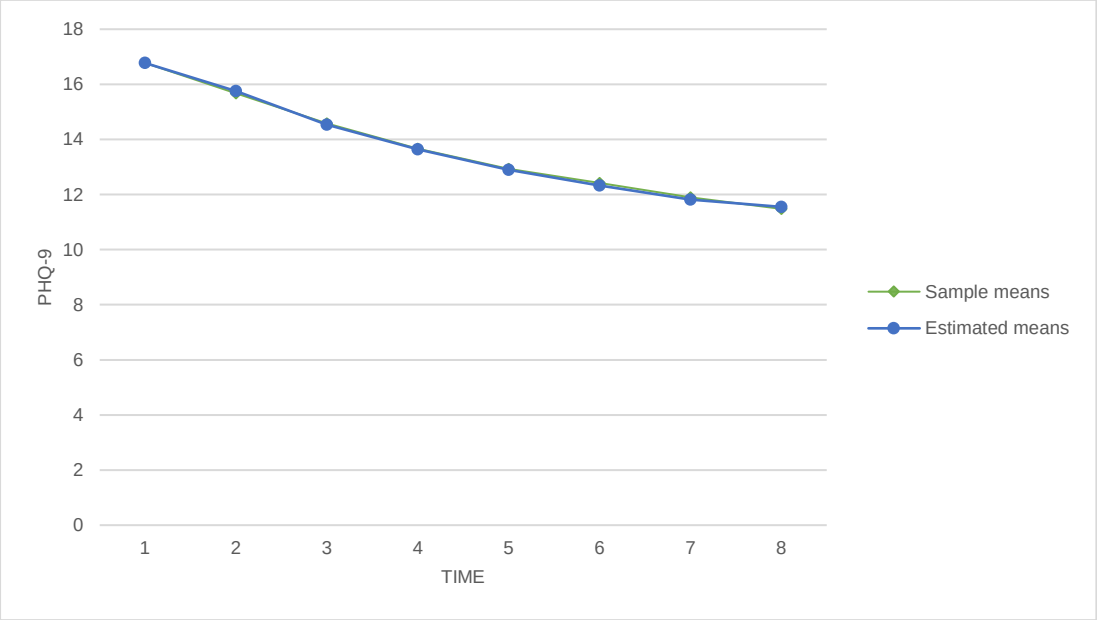
Supplementary Table 9: Model fit indices (latent growth curve model)

Measure	Model	AIC	BIC	SA-BIC	RMSEA (95% CI)	CFI / TLI	SRMR
PHQ 9	Time fixed	246504	246552	246552	.088 (.085 - .092)	.945 / .951	.109
	Time freed	**246050**	**246180**	**246180**	**.084 (.080 - .088)**	**.960 / .955**	**.063**
GAD 7	Time fixed	234797	234886	234845	.092 (.088 - .095)	.936 / .942	.124
	Time freed	**234321**	**234452**	**234391**	**.088 (.084 - .092)**	**.952 / .947**	**.070**

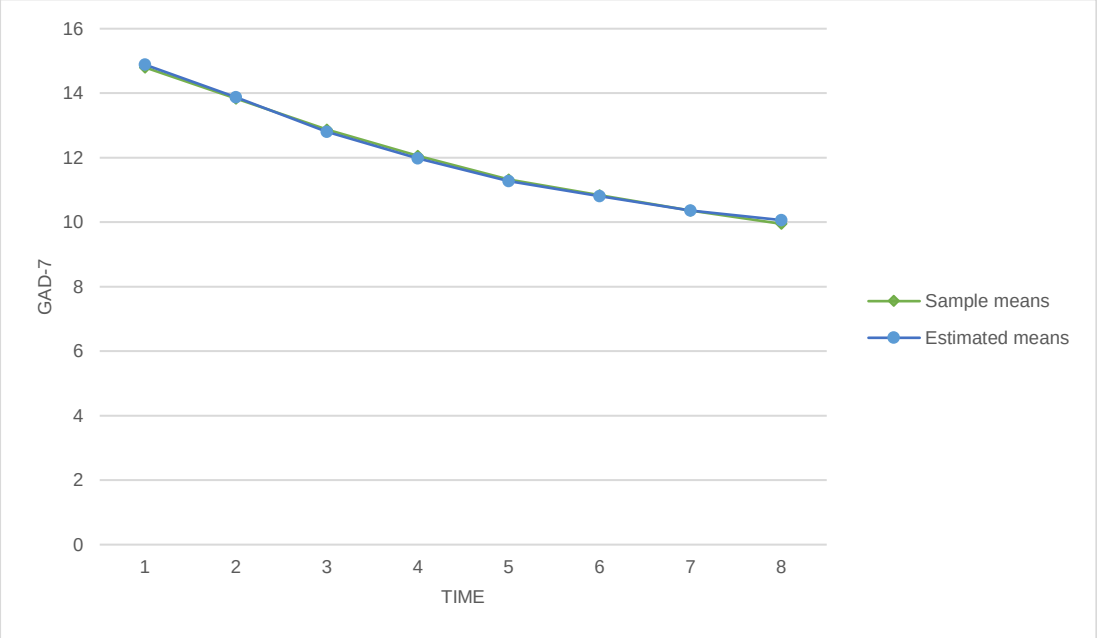
Key: Models selected on the basis of superior fit indices are highlighted bold.

Supplementary Table 9 above reports the model fit indices for the LGCM. Fit indices are explained in Supplementary Table 6. Model specifications are reported in Supplementary Figures 4 and 5. All fit indices were superior for “time-freed” specifications (i.e., times between sessions allowed to be freely estimated by the model). The time-freed models are represented visually in Supplementary Figure 6 and Supplementary Figure 7.

Supplementary Figure 6: Latent Growth Curve Model (LGCM; N = 7,175), depression



Supplementary Figure 7: Latent Growth Curve Model (LGCM; N = 7,175), anxiety



Sample and estimated means for the time-freed LGCM are represented in Supplementary Figure 6 (depression) and Supplementary Figure 7 (anxiety) above.

Supplementary Figure 8: Model specification (latent class growth analysis, depression)

```
VARIABLE:    Names = d1 d2 d3 d4 d5 d6 d7 d8 a1 a2 a3 a4 a5 a6 a7 a8 id;
             Usevariables = d1 d2 d3 d4 d5 d6 d7 d8;
             Idvariable = id;
             Missing = all(-9999);
             Classes = c(8);
ANALYSIS:    Type = Mixture;
             Estimator = mlr;
             Starts = 100 50;
             Stiterations = 10;
MODEL:       %OVERALL%
             Int Slo | d1@0 d2@1 d3* d4* d5* d6* d7* d8*;
             Int-Slo@0;
```

Syntax for the LCGA depression model is shown in Supplementary Figure 8 above. LCGA is recommended as an intermediate model-fitting step between LGCM and GMM, due to the more restricted nature of this approach and the improved likelihood of model convergence (13). The model was specified to have 100 random initial starts and 50 final starts, with 10 starting iterations using maximum likelihood estimation with robust standard errors. The model was specified to freely estimate time intervals after session 2, building on the superior fit of this specification in the LGCM. Fit indices for the LCGA models are presented in Supplementary Table 10.

Supplementary Table 10: Model fit indices (latent class growth analyses)

Measure	Model	AIC	BIC	SA-BIC	VLMR-LRT (*p*)	LMR-A-LRT (*p*)	BLRT (*p*)
PHQ-9	1-class	276379	276489	276438	-	-	-
	2-class	258029	258160	258099	<.0001	<.0001	<.0001
	3-class	251579	251730	251660	<.0001	<.0001	<.0001
	4-class	249327	249499	249419	<.0001	<.0001	<.0001
	5-class	248375	248568	248479	.0153	.0171	<.0001
	6-class	247644	247857	247759	.0028	.0033	<.0001
	7-class	247114	247348	247240	.0022	.0027	<.0001
	8-class	**246712**	**246967**	**246849**	**.0165**	**.0185**	**<.0001**
GAD-7	1-class	261937	262047	261996	-	-	-
	2-class	243857	243988	243928	<.0001	<.0001	<.0001
	3-class	238488	238640	238570	<.0001	<.0001	<.0001
	4-class	236455	236627	236548	<.0001	<.0001	<.0001
	5-class	235609	235802	235713	.0003	.0004	<.0001
	6-class	234848	235061	234963	<.0001	<.0001	<.0001
	7-class	**234315**	**234549**	**234441**	**.0019**	**.0023**	**<.0001**
	8-class	233981	234235	234118	.1010	.1064	<.0001

Key: Models selected on the basis of superior fit indices are highlighted bold

Supplementary Table 10 above reports the model fit indices for the LCGA. Fit indices are explained in Supplementary Table 6. The model specification is reported in Supplementary Figure 8. VLMR-LRT and LMR-A-LRT tests were non-significant for the 8-class anxiety model, indicating the 7-class model would be a better fit. Models continued to improve up to 8 classes for the depression model.

Supplementary Figure 9: Model specification (growth mixture model, depression)

```
VARIABLE:    Names = d1 d2 d3 d4 d5 d6 d7 d8 a1 a2 a3 a4 a5 a6 a7 a8 id;
             Usevariables = d1 d2 d3 d4 d5 d6 d7 d8;
             Idvariable = id;
             Missing = all(-9999);
             Classes = c(5);
ANALYSIS:    Type = Mixture;
             Estimator = mlr;
             Starts = 100 50;
             Stiterations = 10;
MODEL:       %OVERALL%
             Int Slo | d1@0 d2@1 d3* d4* d5* d6* d7* d8*;
```

Syntax for the LCGA depression model is shown in Supplementary Figure 9 above. Models were specified using the same number of random starts and iterations as the LCGA (Supplementary Figure 8). Fit indices are reported in Supplementary Table 11.

Supplementary Table 11: Model fit indices (growth mixture models)

Measure	Model	AIC	BIC	SA-BIC	VLMR-LRT (*p*)	LMR-A-LRT (*p*)	BLRT (*p*)	Entropy
PHQ-9	1-class	246050	246180	246120	-	-	-	1.00
	2-class	245897	246049	245979	<.0001	<.0001	<.0001	.48
	3-class	245726	245898	245819	<.0001	<.0001	<.0001	.51
	4-class	245674	245867	245778	.0001	.0002	<.0001	.59
	5-class	**245635**	**245849**	**245750**	**.0107**	**.0125**	**<.0001**	.56
	6-class	245613	245847	245739	.1837	.1930	<.0001	.54
GAD-7	1-class	234321	234452	234391	-	-	-	1.00
	2-class	233986	234137	234068	<.0001	<.0001	<.0001	.59
	3-class	233681	233853	233774	<.0001	<.0001	<.0001	.56
	4-class	233614	233806	233717	.0002	.0003	<.0001	.62
	5-class	233510	233724	233625	.0120	.0136	<.0001	.60
	6-class	233467	233701	233593	.0120	.1039	<.0001	.62
	7-class	**233409**	**233664**	**233546**	**.0163**	**.0190**	**<.0001**	.61
	8-class	233395	233670	233542	.4288	.4391	<.0001	.66

Key: Models selected on the basis of superior fit indices are highlighted bold

Fit indices for the GMM are reported above in Supplementary Table 11. These showed improved fit over LCGA (AIC, BIC, SA-BIC) and classes were more clearly identified for the anxiety model. On this basis, GMM models were selected as an improved specification over LCGA.

The 5-class depression model was selected on the basis that VLMR-LRT and LMR-A-LRT tests showed significant results (indicating no significant improvement in fit) for the 6-class model. While AIC, BIC and SA-BIC continued to improve, inspection of scree plots confirmed that improvements were small and the 5-class model reflected an “elbow” in the plot line.

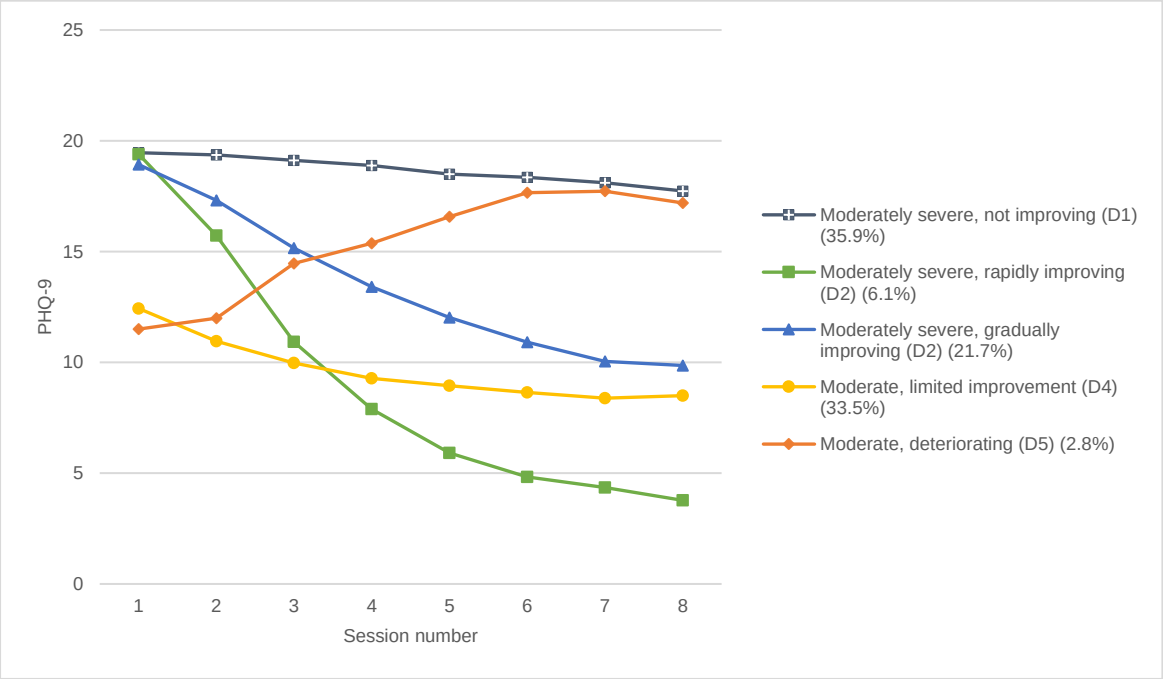
The 7-class anxiety model was selected on the basis that VLMR-LRT and LMR-A-LRT tests showed significant results for the 8-class model, and in addition the BIC value increased (indicating poorer fit). Reductions in AIC and SA-BIC values were small, and inspection of scree plots again confirmed that improvements were small and the 7-class model reflected an “elbow” in the plot line.

BLRT tests failed to reach significance for any of the models, indicating that according to this index models continued to improve with the addition of further classes. Fit indices are commonly not in agreement, and hence researchers have been urged to be transparent about the basis for model selection (43, 48).

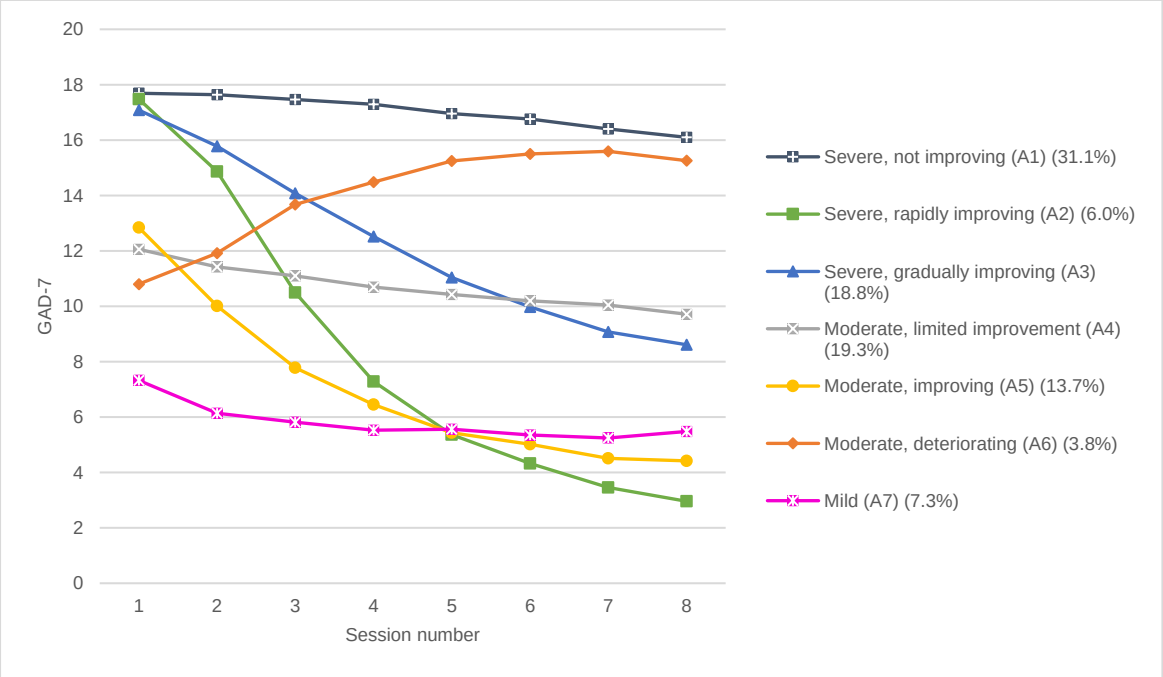
Sample means by most likely class assignment are reported in the body of the main paper. Sample means for the GMM are presented when weighted by class probabilities (Tech7 output in MPlus) in Supplementary Figure 10 (depression) and Supplementary Figure 11 (anxiety). Probability-weighted sample means and standard deviations by timepoint are reported in Supplementary Table 12. Plots for the model-estimated means are presented in Supplementary Figure 12 (depression) and Supplementary Figure 13 (anxiety). Each of these alternative specifications serves as a complement to the plot of the model sample means by most likely class in the Results section of the paper. Alternative GMM specifications were subsequently attempted. Model syntax is presented in Supplementary

Figure 14, These alternative specifications led to errors in the variance-covariance matrices and so the original specification was retained.

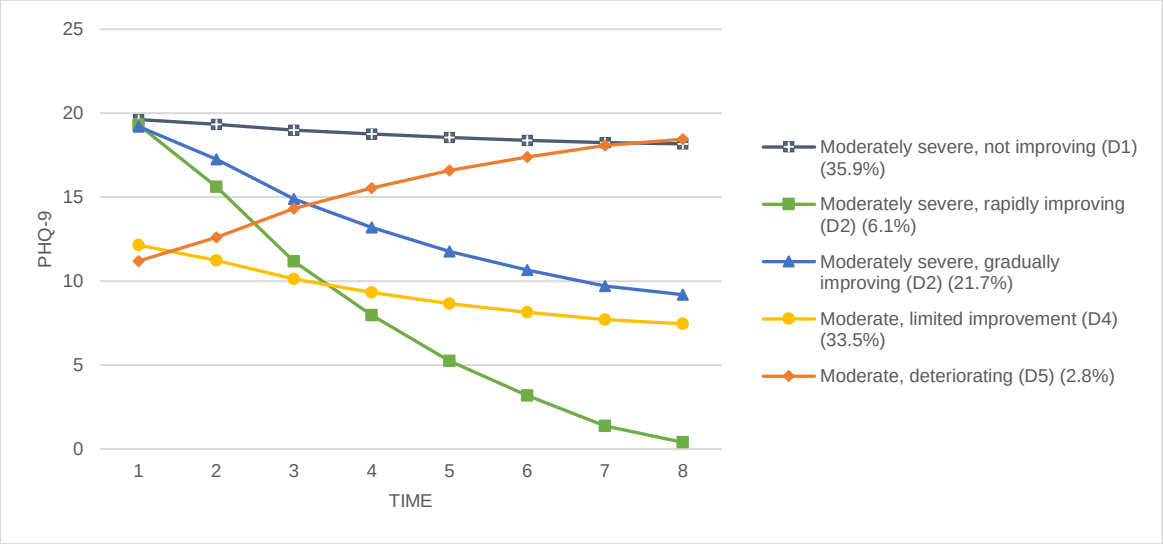
Supplementary Figure 10: Depression growth mixture model sample means, weighted by estimated probabilities (N = 7,175)



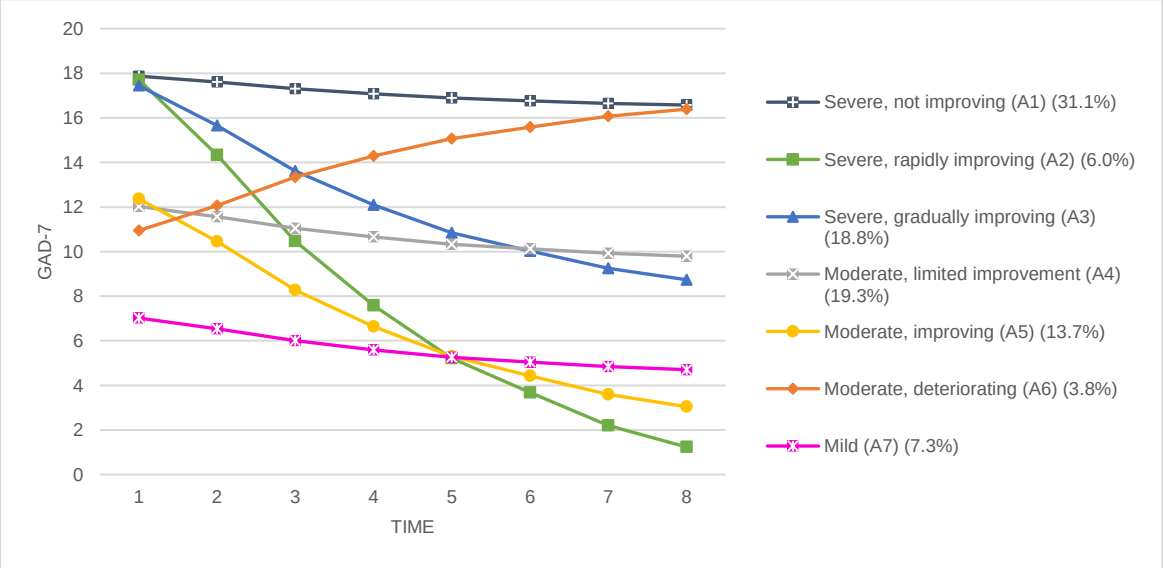
Supplementary Figure 11: Anxiety growth mixture model sample means, weighted by estimated probabilities (N = 7,175)



Supplementary Figure 12: Model-estimated trajectories (growth mixture model, depression; N = 7,175)



Supplementary Figure 13: Model-estimated trajectories (growth mixture model, anxiety; N = 7,175)



Supplementary Table 12: Sample means and standard deviations for probability-weighted models

Measure	Class	t1 (SD)	t2 (SD)	t3 (SD)	t4 (SD)	t5 (SD)	t6 (SD)	t7 (SD)	t8 (SD)
PHQ-9	D1	19.46 (4.23)	19.37 (4.36)	19.13 (4.52)	18.82 (4.69)	18.49 (4.81)	18.36 (4.82)	18.11 (4.93)	17.74 (5.20)
	D2	19.40 (3.96)	15.73 (5.61)	10.92 (6.04)	7.90 (5.34)	5.91 (4.62)	4.84 (3.81)	4.36 (3.54)	3.77 (3.59)
	D3	18.93 (4.11)	17.32 (4.69)	15.16 (5.15)	13.41 (5.23)	12.03 (5.10)	10.91 (4.90)	10.05 (4.86)	9.86 (4.88)
	D4	12.44 (4.41)	10.95 (4.67)	9.97 (4.84)	9.28 (4.80)	8.94 (4.80)	8.64 (4.72)	8.34 (4.78)	8.50 (4.77)
	D5	11.50 (4.43)	12.00 (5.19)	14.46 (5.52)	15.37 (5.15)	16.58 (5.10)	17.65 (4.96)	17.72 (5.06)	17.20 (5.33)
GAD-7	A1	17.69 (2.68)	17.64 (2.79)	17.47 (3.08)	17.30 (3.12)	16.96 (3.30)	16.76 (3.36)	16.41 (3.55)	16.10 (3.83)
	A2	17.48 (2.50)	14.86 (4.09)	10.50 (5.08)	7.28 (4.45)	5.36 (3.56)	4.33 (3.19)	3.45 (2.64)	2.96 (2.78)
	A3	17.08 (2.62)	15.78 (3.37)	14.08 (3.98)	12.52 (4.08)	11.03 (4.02)	9.98 (3.82)	9.08 (2.48)	8.61 (3.89)
	A4	12.06 (3.02)	11.42 (3.56)	11.09 (3.66)	10.70 (3.63)	10.42 (3.65)	10.19 (3.70)	10.05 (3.68)	9.71 (3.84)
	A5	12.84 (3.00)	10.01 (3.89)	7.78 (3.94)	6.45 (3.60)	5.43 (3.33)	5.01 (3.18)	4.51 (3.08)	4.42 (3.25)
	A6	10.80 (3.15)	11.91 (4.08)	13.68 (3.85)	14.48 (3.74)	15.25 (3.49)	15.51 (3.49)	15.59 (3.42)	15.26 (3.95)
	A7	7.32 (3.10)	6.14 (3.19)	5.82 (3.36)	5.53 (3.32)	5.56 (3.39)	5.35 (3.25)	5.24 (3.47)	5.48 (3.40)

Key	
D1	Moderately severe, not improving
D2	Moderately severe, rapidly improving
D3	Moderately severe, gradually improving
D4	Moderate, limited improvement
D5	Moderate, deteriorating
A1	Severe, not improving
A2	Severe, rapidly improving
A3	Severe, gradually improving
A4	Moderate, limited improvement
A5	Moderate, improving
A6	Moderate, deteriorating
A7	Mild

Supplementary Figure 14: Alternative model specifications (growth mixture model; depression)

	Fully variant	Mixed model
MODEL:		
VARIABLE:	Names = d1 d2 d3 d4 d5 d6 d7 d8 a1 a2 a3 a4 a5 a6 a7 a8 id;	Names = d1 d2 d3 d4 d5 d6 d7 d8 a1 a2 a3 a4 a5 a6 a7 a8 id;
	Usevariables = d1 d2 d3 d4 d5 d6 d7 d8;	Usevariables = d1 d2 d3 d4 d5 d6 d7 d8;
	Idvariable = id;	Idvariable = id;
	Missing = all(-9999);	Missing = all(-9999);
	Classes = c(2);	Classes = c(2);
ANALYSIS:	Type = Mixture;	Type = Mixture;
	Estimator = mlr;	Estimator = mlr;
	Starts = 100 50;	Starts = 100 50;
	Stiterations = 10;	Stiterations = 10;
MODEL:	%OVERALL%	%OVERALL%
	Int Slo | d1@0 d2@1 d3* d4* d5* d6* d7* d8*;	Int Slo | d1@0 d2@1 d3* d4* d5* d6* d7* d8*;
	%c#1%	%c#1%
	[Int Slo]	[Int Slo]
	Int (c1_intvar);	Int (c1_intvar);
	Slo(c1_slovar);	Slo(c1_slovar);
	Int with Slo (c1_covintslo);	Int with Slo (c1_covintslo);
	d1 d2 d3 d4 d5 d6 d7 d8 (c1_resfix);	d1 d2 d3 d4 d5 d6 d7 d8 (resfix);
	%c#2%	%c#2%
	[Int Slo]	[Int Slo]
	Int (c2_intvar);	Int (c2_intvar);
	Slo(c2_slovar);	Slo(c2_slovar);
	Int with Slo (c2_covintslo);	Int with Slo (c2_covintslo);
	d1 d2 d3 d4 d5 d6 d7 d8 (c2_resfix);	d1 d2 d3 d4 d5 d6 d7 d8 (resfix);

Two additional model designs were investigated after the selected model reported in this study, in case these produced superior results – the syntax is presented in Supplementary Figure 14 above. One “fully variant” (FV) model was designed, in which both variances-covariances and residual variances were allowed to vary between latent classes. One “mixed” model was designed, in which variance-covariances were allowed to vary between latent classes, but residual variances were held constant across classes.

The FV anxiety model returned a problem with the covariance matrix at the point the 4th class was added (Supplementary Table 13). The mixed model returned psi matrix errors when the 3rd anxiety class and 4th depression class respectively were added (Supplementary Table 14). By contrast there were no psi matrix or covariance matrix problems encountered with the more restricted model that held both variances-covariances and residual variances equal across classes (Supplementary Figure 9, Supplementary Table 11). Therefore although these alternative specifications showed improved entropy values and fit indices (AIC, BIC, SA-BIC), the original model was retained.

Supplementary Table 13: Model fit indices (fully variant growth mixture models)

Measure	Model	AIC	BIC	SA-BIC	VLMR-LRT (*p*)	LMR-A-LRT (*p*)	BLRT (*p*)	Entropy	Errors
PHQ-9	1-class	246078	246160	246122	-	-	-	1.00	
	2-class	242397	242520	242463	<.0001	<.0001	<.0001	.49	
	3-class	241693	241858	241782	<.0001	<.0001	<.0001	.56	
	4-class	241243	241449	241354	<.0001	<.0001	<.0001	.63	
	5-class	241044	241291	241177	.0250	.0264	<.0001	.67	
	6-class	240907	241196	241063	.0073	.0079	<.0001	.63	
	7-class	240810	241140	240988	.1140	.1171	<.0001	.63	
GAD-7	1-class	234396	234478	234440	-	-	-	1.00	
	2-class	230225	230349	230291	<.0001	<.0001	<.0001	.60	
	3-class	229178	229343	229267	.0236	.0249	<.0001	.60	
	4-class	233614	233806	233717	.0002	.0003	<.0001	.62	x
	5-class	233510	233724	233625	.0120	.0136	<.0001	.60	
	6-class	233467	233701	233593	.0120	.1039	<.0001	.62	
	7-class	233409	233664	233546	.0163	.0190	<.0001	.61	
	8-class	233395	233670	233542	.4288	.4391	<.0001	.66	

Supplementary Table 14: Model fit indices (mixed growth mixture models)

Measure	Model	AIC	BIC	SA-BIC	VLMR-LRT (*p*)	LMR-A-LRT (*p*)	BLRT (*p*)	Entropy	Errors
PHQ-9	1-class	246050	246180	246120	-	-	-	1.00	
	2-class	245897	246049	245979	<.0001	<.0001	<.0001	.48	
	3-class	245726	245898	245819	<.0001	<.0001	<.0001	.51	
	4-class	245674	245867	245778	.0001	.0002	<.0001	.59	x
	5-class	245635	245849	245750	.0107	.0125	<.0001	.56	
	6-class	245613	245847	245739	.1837	.1930	<.0001	.54	
GAD-7	1-class	234321	234452	234391	-	-	-	1.00	
	2-class	233986	234137	234068	<.0001	<.0001	<.0001	.59	
	3-class	233681	233853	233774	<.0001	<.0001	<.0001	.56	x
	4-class	233614	233806	233717	.0002	.0003	<.0001	.62	
	5-class	233510	233724	233625	.0120	.0136	<.0001	.60	
	6-class	233467	233701	233593	.0120	.1039	<.0001	.62	
	7-class	233409	233664	233546	.0163	.0190	<.0001	.61	
	8-class	233395	233670	233542	.4288	.4391	<.0001	.66	

Supplementary Table 15: Reliable Change Indices by growth mixture model classes

Measure	Class	Reliable improvement (%)	Reliable recovery (%)	Reliable deterioration (%)
PHQ-9	D1	24.94	9.70	7.17
	D2	99.67	97.39	.00
	D3	88.81	57.46	.38
	D4	44.98	39.06	2.36
	D5	6.54	6.54	62.62
GAD-7	A1	30.80	9.40	9.22
	A2	99.42	94.83	.29
	A3	91.72	48.56	.38
	A4	38.51	37.66	7.29
	A5	96.49	89.38	.40
	A6	5.41	5.41	74.59
	A7	37.92	26.25	7.29

Key	
D1	Moderately severe, not improving
D2	Moderately severe, rapidly improving
D3	Moderately severe, gradually improving
D4	Moderate, limited improvement
D5	Moderate, deteriorating
A1	Severe, not improving
A2	Severe, rapidly improving
A3	Severe, gradually improving
A4	Moderate, limited improvement
A5	Moderate, improving
A6	Moderate, deteriorating
A7	Mild

Information on Reliable Change Indices is presented in the Results. The proportion of each trajectory class that experienced reliable improvement, reliable recovery and reliable deterioration by the end of treatment is presented above in Supplementary Table 15.

Supplementary Table 16: Depression model sensitivity analysis - multinomial logistic regression with full Work and Social Adjustment Scale data (employment subscale included)

Comparison group: D1	**D2**		**D3**		**D4**		**D5**	
	OR	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Model 1:								
Sex (female vs male)	.69	.40 – 1.17	1.17	.82 – 1.67	.76	.56 – 1.02	1.37	.61 – 3.06
Age	1.17	.89 – 1.54	**.72**	**.56 - .92**	1.08	.93 – 1.25	**1.48**	**1.05 – 2.08**
Ethnicity (ethnically minoritised group vs White)	.19	.01 – 3.26	.72	.32 – 1.61	.84	.47 – 1.51	2.01	.52 – 7.78
Employment	.63	.37 – 1.06	.88	.61 – 1.27	.70	.47 – 1.51	.93	.37 – 2.35
Psychotropic medication	.71	.41 – 1.23	**.62**	**.43 - .91**	**.39**	**.29 - .52**	.85	.36 – 2.35
Long-term health condition	1.14	.68 – 1.89	1.25	.87 – 1.81	1.08	.81 – 1.45	.53	.19 – 1.45
Intellectual Disability (has diagnosis vs does not have diagnosis)	1.78	.82 – 3.85	.86	.28 – 1.91	1.41	.86 – 2.30	1.77	.55 – 5.74
Presenting problem (anxiety vs depression)	.89	.47 – 1.68	.70	.41 – 1.18	**3.13**	**2.18 – 4.50**	*	*
Presenting problem ('mixed anxiety and depression' vs depression)	.44	.19 – 1.02	.88	.56 – 1.39	1.21	.79 – 1.84	*	*
Presenting problem (missing vs depression)	.92	.49 – 1.73	.81	.48 – 1.34	**1.70**	**1.14 – 2.53**	*	*
WSAS Total Score	**.55**	**.34 - .88**	.76	.56 – 1.03	**.16**	**.12 - .20**	**.18**	**.09 - .34**
Model 2:								
WSAS: Employment	.90	.78 – 1.04	.93	.84 - 1.03	**.84**	**.77 - .91**	.84	.66 – 1.05
WSAS: Home	.94	.80 – 1.09	.95	.85 – 1.06	**.83**	**.76 - .90**	.82	.67 – 1.01
WSAS: Social leisure	.88	.74 – 1.05	.93	.83 – 1.05	**.82**	**.75 - .90**	.90	.73 – 1.10
WSAS: Private leisure	.89	.76 – 1.06	1.01	.91 – 1.11	**.74**	**.68 - .80**	**.80**	**.65 - .98**
WSAS: Close relationships	.99	.84 – 1.15	1.03	.92 – 1.15	**.84**	**.77 - .91**	**.79**	**.63 - .98**
Comparison group: D4	**D2**		**D3**		**D4**		**D5**	
	OR	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Model 1:								
Sex (female vs male)	.90	.50 – 1.63	**1.54**	**1.02 – 2.32**	-	-	1.80	.83 – 3.91
Age	1.09	.81 – 1.46	**.67**	**.50 - .89**	**-**	**-**	1.37	.99 – 1.90
Ethnicity (ethnically minoritised group vs White)	.23	.01 – 3.71	.86	.32 – 2.35	-	-	2.39	.66 – 8.64
Employment (in employment vs not in employment)	.90	.51 – 1.59	1.27	.83 – 1.93	-	-	1.33	.55 – 3.20
Psychotropic medication (taking vs not taking)	**1.81**	**1.02 – 3.24**	**1.59**	**1.05 – 2.41**	-	-	2.17	.96 – 4.90
Long-term health condition (has diagnosed condition vs does not have diagnosed condition)	1.05	.60 – 1.83	1.16	.77 – 1.76	-	-	.49	.18 – 1.32
Intellectual Disability (has diagnosis vs does not have diagnosis)	1.26	.53 – 2.98	.61	.27 – 1.38	-	-	1.26	.42 – 3.79
Presenting problem (anxiety vs depression)	**.28**	**.14 - .56**	**.22**	**.13 - .39**			*	*
Presenting problem ('mixed anxiety and depression' vs depression)	**.36**	**.14 - .91**	.73	.43 – 1.26			*	*
Presenting problem (missing vs depression)	.55	.27 – 1.09	**.48**	**.26 - .86**			*	*
WSAS Total Score	**3.44**	**2.02 – 5.86**	**4.81**	**3.55 – 6.53**	**-**	**-**	1.13	.61 – 2.11
Model 2:								
WSAS: Employment	1.08	.92 – 1.26	1.11	.99 – 1.25	**-**	**-**	1.00	.79 – 1.25
WSAS: Home	1.13	.96 – 1.32	**1.14**	**1.02 – 1.29**	-	-	.99	.81 – 1.21

WSAS: Social leisure	1.07	.88 – 1.30		1.13	1.00 – 1.28		-	-		1.01	.89 – 1.32
WSAS: Private leisure	**1.21**	**1.01 – 1.45**		**1.36**	**1.22 – 1.53**		**-**	**-**		1.01	.88 – 1.33
WSAS: Close relationships	1.18	1.00 – 1.39		**1.23**	**1.10 – 1.28**		**-**	**-**		.94	.77 – 1.15

Key

ORs are highlighted in bold when the 95% confidence interval does not include 1, indicating statistical significance.

D1	Moderately severe, not improving
D2	Moderately severe, rapidly improving
D3	Moderately severe, gradually improving
D4	Moderate, limited improvement
D5	Moderate, deteriorating

Supplementary Table 16 shows a sensitivity analysis, whereby the multinomial logistic regression analysis reported in the main paper were re-analysed in relation to the depression model using full WSAS scores (i.e., the employment question was excluded in the main analysis and was included in this sensitivity analysis. See Supplementary Table 5 for further information.

Supplementary Table 17: Anxiety model sensitivity analysis - multinomial logistic regression with full Work and Social Adjustment Scale data (employment subscale included)

Comparison group: A1	**A2**		**A3**		**A4**		**A5**		**A6**		**A7**	
	OR	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Model 1:												
Sex (female vs male)	.86	.56 – 1.30	.84	.62 – 1.15	.76	.56 – 1.03	**.52**	**.35 - .78**	.72	.37 – 1.43	**.54**	**.36 - .80**
Age	.77	.58 – 1.01	.90	.77 – 1.06	.85	.72 – 1.01	1.07	.86 – 1.34	1.08	.81 – 1.43	1.13	.93 – 1.38
Ethnicity (ethnically minoritised group vs White)	*	*	.94	.53 – 1.67	.55	.23 – 1.32	1.01	.51 – 2.00	**2.71**	**1.23 – 5.96**	1.49	.79 – 2.80
Employment (in employment vs not in employment)	.85	.57 – 1.28	.75	.55 – 1.02	.95	.70 – 1.28	.83	.56 – 1.23	1.34	.71 – 2.53	1.06	.73 – 1.54
Psychotropic medication (taking vs not taking)	.75	.49 – 1.15	**.64**	**.47 - .87**	.76	.57 – 1.03	**.47**	**.32 - .68**	.86	.47 – 1.55	**.67**	**.46 - .98**
Long-term health condition (has diagnosed condition vs does not have diagnosed condition)	.76	.49 – 1.16	1.04	.76 – 1.41	.78	.57 – 1.07	.99	.69 – 1.42	.85	.43 – 1.67	.96	.66 – 1.38
Intellectual Disability (has diagnosis vs does not have diagnosis)	1.44	.78 – 2.65	.93	.52 – 1.66	.88	.50 – 1.56	**1.86**	**1.13 – 3.07**	.90	.26 – 3.09	.57	.26 – 1.23
Presenting problem (anxiety vs depression)	.99	.58 – 1.68	.82	.55 – 1.23	**.44**	**.30 - .64**	.90	.58 – 1.40	.87	.35 – 2.14	**.17**	**.09 - .33**
Presenting problem ('mixed anxiety and depression' vs depression)	.85	.44 – 1.66	1.08	.71 – 1.65	**.55**	**.36 - .84**	.59	.32 – 1.08	1.17	.50 – 2.72	**.33**	**.18 - .57**
Presenting problem (missing vs depression)	1.32	.76 – 2.28	.77	.49 – 1.20	**.55**	**.36 - .83**	.83	.49 – 1.42	.94	.37 – 2.37	**.57**	**.37 - .88**
WSAS Total Score	**.60**	**.44 - .82**	**.69**	**.54 - .87**	**.36**	**.29 - .46**	**.24**	**.19 - .31**	**.33**	**.23 - .48**	**.23**	**.19 - .29**
Model 2:												
WSAS: Employment	.98	.87 – 1.11	.98	.90 – 1.06	.92	.84 – 1.00	**.88**	**.79 - .98**	.92	.77 – 1.09	**.87**	**.78 - .97**
WSAS: Home	.98	.86 – 1.11	.95	.87 – 1.04	.95	.88 – 1.04	.90	.81 – 1.00	.95	.81 – 1.11	.93	.83 – 1.04
WSAS: Social leisure	**.88**	**.78 – .99**	**.86**	**.78 - .95**	**.87**	**.80 - .95**	**.77**	**.69 - .85**	.85	.73 – 1.00	**.84**	**.76 - .94**
WSAS: Private leisure	.90	.79 – 1.02	1.01	.93 – 1.10	**.86**	**.80 - .93**	**.80**	**.71 - .90**	.86	.74 – 1.00	**.78**	**.70 - .87**
WSAS: Close relationships	.97	.86 – 1.11	1.00	.91 – 1.10	**.91**	**.84 - .98**	.97	.88 – 1.08	.87	.74 – 1.01	**.87**	**.79 - .96**
Comparison group: A4	**A2**		**A3**		**A4**		**A5**		**A6**		**A7**	
	OR	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Model 1:					-	-						
Sex (female vs male)	1.13	.71 – 1.78	1.11	.74 – 1.65	-	-	.69	.41 – 1.15	.95	.43 – 2.09	.71	.44 – 1.16
Age	.90	.67 – 1.20	1.06	.84 – 1.32	-	-	1.26	.94 – 1.69	1.26	.88 – 1.80	1.33	1.02 – 1.72
Ethnicity (ethnically minoritised group vs White)	*	*	1.71	.64 – 4.54	-	-	1.82	.59 – 5.62	**4.91**	**1.43 – 16.89**	2.70	.99 – 7.39
Employment (in employment vs not in employment)	.90	.58 – 1.41	.79	.54 – 1.17	-	-	.88	.53 – 1.44	1.42	.68 – 2.94	1.12	.71 – 1.77
Psychotropic medication (taking vs not taking)	.98	.61 – 1.56	.83	.57 – 1.22	-	-	**.61**	**.38 - .98**	1.12	.56 – 2.21	.88	.56 – 1.38
Long-term health condition (has diagnosed condition vs does not have diagnosed condition)	.97	.60 – 1.56	1.33	.89 – 1.99	-	-	1.27	.78 – 2.05	1.09	.50 – 2.40	1.23	.78 – 1.94
Intellectual Disability (has diagnosis vs does not have diagnosis)	1.63	.79 – 3.36	1.05	.49 – 2.26	-	-	**2.12**	**1.03 – 4.35**	1.02	.24 – 4.39	.64	.26 – 1.59
Presenting problem (anxiety vs depression)	**2.26**	**4.07**	**1.88**	**1.15 – 3.10**			**2.06**	**1.16 – 3.65**	1.99	.72 – 5.50	.38	.18 - .83
Presenting problem ('mixed anxiety and depression' vs depression)	1.55	.76 – 3.14	**1.96**	**1.16 – 3.33**			1.07	.51 – 2.25	2.12	.81 – 5.54	.59	.30 – 1.17
Presenting problem (missing vs depression)	**2.40**	**1.30 – 4.43**	1.40	.80 – 2.47			1.52	.77 – 3.01	1.71	.58 – 5.04	1.03	.58 – 1.83

WSAS Total Score	**1.64**	**1.17 – 2.31**		**1.89**	**1.47 – 2.42**		**-**	**-**		**.67**	**.50 - .91**		.90	.59 – 1.39		**.64**	**.50 - .82**
Model 2:																	
WSAS: Employment	1.07	.94 – 1.22		1.07	.96 – 1.19		-	-		.96	.84 – 1.10		1.00	.82 – 1.23		.95	.84 – 1.08
WSAS: Home	1.03	.90 – 1.18		1.00	.90 – 1.12		-	-		.94	.82 – 1.08		1.00	.83 – 1.20		.97	.85 – 1.11
WSAS: Social leisure	1.01	.89 – 1.15		.99	.89 – 1.10		-	-		.88	.78 – 1.00		.98	.83 – 1.16		.97	.86 – 1.09
WSAS: Private leisure	1.09	.90 – 1.19		**1.18**	**1.06 – 1.30**		**-**	**-**		.93	.80 – 1.08		1.00	.84 – 1.19		.90	.79 – 1.03
WSAS: Close relationships	1.07	.94 – 1.22		1.10	.99 – 1.23		-	-		1.07	.95 – 1.21		.96	.80 – 1.13		.96	.86 – 1.08

Key

ORs are highlighted in bold when the 95% confidence interval does not include 1, indicating statistical significance.

A1	Severe, not improving
A2	Severe, rapidly improving
A3	Severe, gradually improving
A4	Moderate, limited improvement
A5	Moderate, improving
A6	Moderate, deteriorating
A7	Mild

Supplementary Table 17 shows a sensitivity analysis, whereby the multinomial logistic regression analysis reported in the main paper were re-analysed in relation to the anxiety model using full WSAS scores (i.e., the employment question was excluded in the main analysis and was included in this sensitivity analysis. See Supplementary Table 5 for further information.

Supplementary Table 18: Guidelines for Reporting on Latent Trajectory Studies (GRoLTS) Checklist (Van de Schoot et al., 2017)

Item number	Checklist item	Yes/No	Location in manuscript	Comments
1	Is the metric of time used in the statistical model reported?	Yes	Methods: Procedure	
2	Is information presented about the mean and variance of time within a wave?	Yes	Supplementary Table 4	
3a	Is the missing data mechanism reported?	Yes	Results; Methods: Procedure	
3b	Is a description provided of what variables are related to attrition / missing data?	Yes	Supplementary Table 8	
3c	Is a description provided of how missing data in the analyses were dealt with?	Yes	Methods: Procedure	
4	Is information about the distribution of the observed variables included?	Yes	Supplementary Table 7; Supplementary Figure 2; Supplementary Figure 3	
5	Is the software mentioned?	Yes	Methods: Procedure	
6a	Are alternative specifications of within-class heterogeneity considered (e.g., LCGA vs LGMM) and clearly documented? If not, was sufficient justification provided as to eliminate certain specifications from consideration?	Yes	Methods: Procedure; Supplementary Figure 4; Supplementary Figure 14	
6b	Are alternative specifications of the between-class differences in variance-covariance matrix structure considered and clearly documented? If not, was sufficient justification provided as to eliminate certain specifications from consideration?	Yes		We attempted models in which variances and residual variances were estimated separately between classes, but these encountered errors with psi-matrix. These are reported in Figure 14.
7	Are alternative shape/functional forms of the trajectories described?	Yes	Method; Supplementary Figure 4	We compared linear (intercept and slope) with a model that makes no assumptions about the shape of growth (free time scores)
8	If covariates have been used, can analyses still be replicated?	N.A		Covariates were not included in the model; instead, the R3Step procedure was followed (see Methods: Procedure)
9	Is information reported about the number of random start values and final iterations included?	Yes	Supplementary Figure 8; Supplementary Figure 9	
10	Are the model comparison (and selection) tools described from a statistical perspective?	Yes	Supplementary Table 6	
11	Are the total number of fitted models reported, including a one-class solution?	Yes	Supplementary Table 9; Supplementary	

			Table 10; Supplementary Table 11	
12	Are the number of cases per class reported for each model (absolute sample size, or proportion)?	Yes	Results; Supplementary Table 11	
13	If classification of cases in a trajectory is the goal, is entropy reported?	Yes	Supplementary Table 11	
14a	Is a plot included with the mean trajectories of the final solution?	Yes	Results	
14b	Are plots included with the estimated mean trajectories for each model?	Yes	Supplementary Figure 12; Supplementary Figure 13	Sample mean trajectories are also reported in Results
14c	Is a plot included of the combination of estimated means of the final model and the observed individual trajectories split out for each latent class?		Supplementary Figure 10; Supplementary Figure 11	Plots were included instead weighted by estimated probabilities
15	Are characteristics of the final class solution numerically described (i.e., means, SD/SE, *n*, C1 etc.)?	Yes	Results	
16	Are the syntax files available (either in the appendix, supplementary materials, or from the authors)?	Yes	Supplementary Figure 4; Supplementary Figure 5;Supplementary Figure 8; Supplementary Figure 9	

Supplementary Table 19: The REporting of studies Conducted using Observational Routinely collected Data (RECORD) checklist (Benchimol et al., 2015)

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	a) Abstract b) Abstract	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	1.1 Abstract 1.2 Abstract 1.3 Abstract
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	Introduction		
Objectives	3	State specific objectives, including any prespecified hypotheses	Introduction		
Methods					
Study Design	4	Present key elements of study design early in the paper	Introduction Methods: Study design		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Methods: Study design; Participants		

Participants	6	*(a)* *Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b)* *Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case	A) Methods: Participants; Results B) N.A	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	6.1 Methods: Participants 6.2 Methods: Participants 6.3 Supplementary Figure 1
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Participants: Measures; Supplementary Table 1; Supplementary Table 2	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Method: Participants; Method: Procedure
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Participants: Measures; Supplementary Table 2; Supplementary Table 5		
Bias	9	Describe any efforts to address potential sources of bias	Methods: Procedure; Supplementary Table 8		
Study size	10	Explain how the study size was arrived at	Methods: Participants; Supplementary Figure 1		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Methods: Procedure; Supplementary Tables 7 – 14; Supplementary Figures 2 - 14		

Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	a) Methods: Procedure b) Methods: Procedure c) Methods: Procedure; Supplementary Table 5; Supplementary Table 8 d) Methods: Procedure; Supplementary Table 5; Supplementary Table 8 e) Results; Supplementary Table 16; Supplementary Table 17		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	Methods: Study Design

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Methods: Participants; Methods: Procedure; Supplementary Table 7; Supplementary Figures 2-3; Supplementary Table 8
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Methods: Study design
Results					

Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for nonparticipation at each stage. (c) Consider use of a flow diagram	a) Results Supplementary Figure 1 b) Results Supplementary Figure 1 c) Results Supplementary Figure 1	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Methods: Participants; Supplementary Figure 1
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	a) Results b) Supplementary Figure 5; Supplementary Figure 8 c) Supplementary Table 4		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure	Methods: Procedure; Results; Supplementary Table 6; Supplementary Tables 11-14; Supplementary Figures 9-14		

		category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounderadjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	a) Results; Supplement b) N.A c) N.A		

Other analyses	17	Report other analyses done— e.g., analyses of subgroups and interactions, and sensitivity analyses	Supplementary Table 13; Supplementary Table 14; Supplementary Table 16; Supplementary Table 17; Supplementary Figure 14		
Discussion					
Key results	18	Summarise key results with reference to study objectives	Discussion		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Discussion: Limitations	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Discussion: Limitations
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Discussion		
Generalisability	21	Discuss the generalisability (external validity) of the study results	Discussion		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Acknowledgements; Competing Interests Statement		
Accessibility of protocol, raw data, and programming code		..	Data Availability; Code Availability	RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	

Supplementary Table 19 above presents the RECORD statement – a checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.