PRedictors Of Psychological trEatment outcomes for common mentaL health problems in NHS Talking Therapy service users (Propel)

Alexandra Schmidt[1], Clara Strauss[1,2], Nick Grey[1,2], Darya Gaysina[1]

[1] School of Psychology, University of Sussex, Brighton, BN1 9QH, UK
[2] Sussex Partnership NHS Foundation Trust, Sussex Education Centre, Nevill Avenue, Hove, BN3 7HZ, UK

Background

Common mental health problems (CMHP), such as depression and anxiety disorders, are highly prevalent in adults, with 1 in 6 adults being affected (Mueller et al., 2017; NHS Digital, 2016) with some studies suggesting the prevalence in older adults to be as high as 1 in 4 (Andreas et al., 2017) and even higher for older people living in care homes, with current estimates indicating that 40% of individuals are affected (Royal College of Psychiatrists, 2018). CMHP are associated with physical health problems, cognitive decline, and mortality (Hare et al., 2014; Rodda et al., 2011). This is particularly relevant to the population in the present time as we live in an ageing population: it is estimated that by 2030, 1 out of 5 people (21.8%) in the UK will be aged 65 or over, 6.8% will be aged 75 or over and 3.2% will be 85 or over (Office for National Statistics, 2017), whilst time spent in poor health is increasing (ONS, 2017).

The National Health Service (NHS) in England established a new primary care psychological therapy services in 2008, NHS Talking Therapy Service (formerly known as Improving Access to Psychological Therapies (IAPT)) for delivering evidence-based treatments as recommended by the National Institute for Health and Care Excellence (NICE). Treatments include self-help support, cognitive behavioural therapy and counselling in a stepped-care model. NHS Talking Therapy Services now cover all of England and was accessed by 1.76 million people in 2022/23, with 46.8% of patients moving into reliable recovery and 66.5% reporting ‘reliable improvement’ in symptom severity (NHS Digital, 2024). Whilst the reported reliable recovery rates in the NHS Talking Therapies are encouraging, they also demonstrate that there is room for improvement as half of the patients are not meeting the definition of reliable recovery by the end of their treatment.

Previous research suggested that individuals may benefit more or less from psychological therapy based on their unique context and characteristics (Delgadillo et al.,

2016). A recent systematic literature review of psychological treatment outcome predictors for older adults identified predictors across different domains, including psychosocial (e.g., coping styles), socio-demographic (e.g., level of education), clinical (e.g., baseline symptom severity) and treatment-related (e.g., number of sessions attended) (Schmidt et al., 2024). However, the review highlighted some limitations, such as high heterogeneity across studies, small sample sizes as well as the exploratory nature of this research field resulting in some conditions and treatment being understudied. In contrast to our review, previous research identifying predictors for psychological treatment outcomes for common mental health problems across all ages identified different predictors of treatment outcomes, such as gender and social support (Amati et al., 2017), which suggests potential differences in predictors between different age groups. Few studies to date have explored how predictors may affect psychological treatment outcomes differentially for working age compared to older adults, for an example see Saunders et al. (2021).

Machine learning models can be advantageous in eliminating potentially unimportant features and assessing possible non-linear associations and interactions (Chekroud et al., 2021). Previous prediction modelling research using machine learning techniques reported some promising results, however, these typically included routinely collected data and did not consider other factors previously reported to be associated with treatment outcomes, e.g. from the psychosocial domain, such as coping style (Bone et al., 2021; Hilbert et al., 2020).

The proposed study aims to address these limitations by collecting a wide set of variables from patients awaiting treatment at the NHS Talking Therapies services and develop a parsimonious prediction model for psychological treatment outcomes.

Research questions and aims

In this proposed study we aim to investigate the factors that may predict psychological treatment outcomes in the NHS Talking Therapy services.

Data collected from study participants at baseline will be combined with data routinely collected in the NHS Talking Therapy Services. The study will be investigating the following research questions:

1. Are there significant differences between older (65+ years) and working age adults (18-64 years) in baseline measures and post-treatment outcomes?
2. How accurately can machine learning predict treatment outcome(s) reliable recovery (recovery, reliable improvement, reliable deterioration) following psychological treatment in the NHS Talking Therapies service?
3. What are the individual factors (sociodemographic, clinical, treatment-related, psychosocial) that have the highest importance in the prediction model for 'reliable recovery' (recovery, reliable improvement, and reliable deterioration)?
    a. Are there important differences in these factors between older (65+ years) and working age adults (18-64 years)? (Sensitivity analyses)

Methods

Data collection process

In stage 1, participants were recruited from 21 participating NHS Trusts across England between March 2022 and March 2023. At the time of recruitment, these participants had been assessed and accepted into the NHS Talking Therapy service and were awaiting the start of

their treatment. At baseline, participants completed a questionnaire with potential factors impacting psychological treatment outcomes. The questionnaire measures were informed by a previously conducted systematic review of existing literature (Schmidt et al., 2024, under review) and comprised 18 measures across different domains (sociodemographic, clinical, psychosocial), for full details see Table 5.

In stage 2, participating NHS Talking Therapy Services provided data extracts from their respective systems which includes routinely collected data such as gender, condition/problem descriptor, Step 2 / Step 3, type of treatment, number of sessions attended, whether the patient was referred on to the next step and pre- and post-treatment symptom severity for each condition (e.g., PHQ-9, GAD-7, OCI, SPIN, MI, PCL-5, PDSS, WSAS).

Sample size calculations

For the sample size calculation for the prediction model (research question 2), pmsampsize (Ensor, 2022) was used to establish the required samples size for different numbers of parameters as shown in Table 1. These calculations assume an outcome prevalence of 50% (based on the reported statistics on 'reliable recovery' in the NHS Talking Therapy Services, a C-statistic of 0.7 and 90 shrinkage.

Table 1. Calculation of required sample sizes for different number of predictors for prediction model

Parameters	Expected outcome	C-statistic	Shrinkage	Sample required
10	0.5	0.7	0.9	695
11				764
12				834

13				903
14				973
15				1024
16				1112
17				1181
18				1251

Ethical approval

The study was approved by South Central-Hampshire B Research Ethics Committee (Health Research Authority, 2 Redman Place, Stratford, London, E20 1JQ, UK; +44 20 7104 8289; hampshireb.rec@hra.nhs.uk), ref: 21/SC/0387, approved 24/12/2021.

Measures

Outcomes

The Patient Health Questionnaire (PHQ-9)(Kroenke et al., 2001) is a 9-item self-report measure of depression symptom severity used in all NHS Talking Therapy Services. Items are rated on a four-point scale. The 'caseness' threshold for this measure is ≥ 10.

Generalised Anxiety Disorder Assessment (GAD-7) (Spitzer et al., 2006) is a 7-item measure of generalised anxiety used in the NHS Talking Therapy Services. Items are rated on a 4-point scale and the measure has excellent psychometric properties. The 'caseness' threshold for this measure is ≥ 8.

Obsessive-Compulsive Inventory (OCI) (Foa et al., 1998) is a 42-item measure of obsessive-compulsive disorder used in the NHS Talking Therapy Services. Items are rated on a 5-point scale. The caseness threshold for this measure is ≥ 40.

Social Phobia Inventory (SPIN) (Connor et al., 2000) is a 17-item measure of social phobia used in the NHS Talking Therapy Services. Items are rated on a 5-point scale. The caseness threshold for this measure is ≥ 19.

Agoraphobia-Mobility Inventory (MI) (Chambless et al., 1985) is a 27-item measure of agoraphobia used in the NHS Talking Therapy Services. Items are rated on a 5-point scale, with each situation being rated twice to reflect the degree that the situation is avoided. The caseness threshold for this measure is 2.3 per item average.

PTSD Checklist for DSM-5 (PCL-5) (Blevins et al., 2015) is a 20-item measure for PTSD used in the NHS Talking Therapy Services. Items are rated on a 5-point scale. The caseness threshold for this measure is ≥ 32.

Panic Disorder Severity Scale (PDSS) (Shear et al., 2001) is a 7-item measure to grade panic disorder severity used in the NHS Talking Therapy Services. Items are rated on a 5-point scale. The caseness threshold for this measure is ≥ 8.

Functioning (WSAS): The Work and Social Adjustment Scale (WSAS) (Mundt et al., 2002) is a 5-item measure of daily occupational and social functioning that is used routinely in the NHS Talking Therapy Services.

Binary outcomes

For the calculation of the binary outcome of 'reliable recovery' we will analyze the change between the first and last session score on the instrument for the condition being treated as outlined in Table 2. We may also be investigating 'recovery' 'reliable improvement', and 'reliable deterioration'.

Table 2. NHS Talking Therapies definition of outcome measures

Outcome	Measures used	Definition	Coded
Reliable recovery	PHQ-9	Scoring above 'caseness' threshold on	0 = not reliably recovered,

	GAD-7	the PHQ-9 or GAD-7 sum scores at the start of treatment, below 'caseness' threshold on both measures at follow-up ***and*** showing reliable improvement on either PHQ or GAD.	1 = reliably recovered
Reliable improvement	PHQ-9 GAD-7	To account for measurement error, a reliable change threshold has been defined corresponding to ≥ 6 for PHQ and ≥ 4 for GAD. Reliable improvement is defined as experiencing an improvement in PHQ or GAD sum score corresponding to the reliable change threshold.	0 = not reliably improved, 1 = reliably improved
Reliable recovery	PHQ-9 GAD-7	Scoring above 'caseness' threshold on the PHQ-9 or GAD-7 sum scores at the start of treatment and below 'caseness' threshold on both measures at follow-up (NHS digital, 2019). The caseness thresholds that will be used are ≥ 10 for PHQ-9 and/or ≥ 8 for GAD-7.	0 = not recovered, 1 = recovered
Reliable deterioration	PHQ-9 GAD-7	Reliable change threshold has been defined corresponding to ≥ 6 for PHQ and ≥ 4 for GAD; reliable deterioration as experiencing a worsening in PHQ or GAD sum score corresponding to the	0 = not reliably deteriorated, 1 = reliably deteriorated

		reliable change threshold.	

Selection of predictors

Potential predictors that will be considered in these analyses are outlined in Table 3 and have been selected based on existing research, e.g., systematic reviews into factors relating to psychological treatment outcomes (Amati et al., 2017; Schmidt et al., 2024) and mental health outcomes, such as self-compassion (Athanasakou et al., 2020; MacBeth & Gumley, 2012) , resilience (Hu et al., 2015) and loneliness (Erzen & Çikrikci, 2018).

Table 3. Overview of predictors

Domain	Measure	Continuous/ Categorical	Timepoint collected	NHS Talking Therapy Service (NHS TTS) or baseline questionnaire (BQ))
Sociodemographic	Age (Working age adult /Older adult)	Categorical	Stage 1	BQ
	Age – years	Continuous		
	Gender (Male/Female/Non-binary)	Categorical	Stage 2	NHS TTS
	Ethnicity (White/Mixed/Asian/Black/Other)	Categorical	Stage 1	BQ
	Education (No qualification/GCSE/Alevel/BTEC/Bachelor's degree, Master's degree, Doctorate)	Categorical	Stage 1	BQ
	Employment status (Employed/Unemployed/ Full-time student/Retired/Other)	Categorical	Stage 1	BQ
	Benefit receipt (No/yes)	Categorical	Stage 1	BQ

Clinical	Baseline symptom severity PHQ-9, GAD-7, WSAS	Continuous	Stage 2	NHS TTS
	Problem descriptor (Depression/Generalised anxiety disorder/Post-traumatic stress disorder/Obsessive-compulsive disorder/Panic disorder/Agoraphobia/Social anxiety disorder/Specific phobia/Health anxiety/Adjustment disorders/Mixed anxiety and depression/Body dysmorphic disorder)	Categorical	Stage 2	NHS TTS
	Duration of current mental health problems (12 months or less/More than 12 months)	Categorical	Stage 1	BQ
	Recurrence of mental health problems (No/yes)	Categorical	Stage 1	BQ
	Age at onset (Adulthood/Childhood or adolescence)	Categorical	Stage 1	BQ
	MH duration of illness total	Continuous	Stage 1	BQ
	Psychotropic medication use (No/yes)	Categorical	Stage 1	BQ
	Other mental health problems (No/yes)	Categorical	Stage 1	BQ
	Type of other mental health problems*	Categorical	Stage 1	BQ
	Physical health problems (No/yes)	Categorical	Stage 1	BQ
	Type of physical health problems*	Categorical	Stage 1	BQ
Treatment-related	Type of treatment[1] (CBT/Counselling/Group/EMDR/LTC Pathway/IPT/CAT/ERP/CfT/ACT/ Other high intensity/GSH/PSH/CCBT/BA/Psychoeducation/Employment support/Workshops or webinar/other low intensity/)	Categorical	Stage 2	NHS TTS

	Intensity of treatment (Step 2/Step 3)	Categorical	Stage 2	NHS TTS
	Number of sessions attended	Continuous	Stage 2	NHS TTS
Psychosocial	Perceived health (total and 4 subscales)**	Continuous	Stage 1	BQ
	Loneliness	Continuous	Stage 1	BQ
	Locus of control – internal	Continuous	Stage 1	BQ
	Locus of control – external	Continuous	Stage 1	BQ
	Personality (5 dimensions: extraversion, conscientiousness, agreeableness, openness to experience, neuroticism)	Continuous	Stage 1	BQ
	Self-compassion (total & 4 subscales)**	Continuous	Stage 1	BQ
	Resilience	Continuous	Stage 1	BQ
	Stressful life events (total)	Continuous	Stage 1	BQ
	Current impact of stressful life events (total & 3 subscales)**	Continuous	Stage 1	BQ
	Coping style (14 coping styles)	Continuous	Stage 1	BQ

[1] CBT = Cognitive behavioural therapy, EMDR = Eye Movement Desensitisation and Reprocessing, LTC = Long term condition, IPT = Interpersonal Therapy, CAT = Cognitive Analytic Therapy, ERP = Exposure and response prevention, CfT = Compassion-focused therapy, ACT = Acceptance and commitment therapy, GSH = Guided self-help, CCBT = Computerised cognitive behavioural therapy, PSH = Pure self help, BA = Behavioural activation

*Binary variables for other mental health problems or physical health problems will be used in main model. Types of other mental health or physical health problems may be explored further in sensitivity analyses

**Total measures will be used for model initially. Subscales may be explored further if measure identified as important factor

Planned statistical analyses

Research question 1: Descriptive statistics and age-group differences in baseline and post-treatment measures

A descriptive summary of all variables will be created for the baseline and post-treatment measures, for each age group separately (means and standard deviation for continuous variables, frequencies, and percentages for categorical variables). Normality will be assessed by inspecting histograms and diagnostic plots. Chi-square or t-tests will be performed to ascertain any statistically significant differences in the baseline variables between the two age groups and effect sizes (Cramer's V or Cohen's d) will be computed.

Research question 2 & 3: Prediction model for psychological treatment outcome of reliable recovery

1) **Exploratory data analysis**

Data visualisation

A series of multivariate visualisation plots will be created to understand patterns in the data and potential associations between predictors and outcome of reliable recovery, e.g. scatterplot matrices by class for continuous variables/features and density plots by class for categorical variables/features. For categorical variables, the distribution of responses will be assessed and levels reduced/collapsed for infrequent levels and where this makes sense theoretically.

Univariate associations

Univariate logistic regression tests will be performed to identify statistically significant associations ($\alpha = .05$) between individual factors and the outcome. Additionally,

interaction terms with age group (working age adult/older adult) will be tested to identify any age specific associations between features and the outcome.

2) Data cleaning

Missing data

Missing data for predictors will be assessed and analysis performed on patterns of missingness. Missing data will be imputed using the 'missRanger' package in R Studio (Mayer, 2024). MissRanger offers the option of combining random forest imputation with predictive mean matching for both continuous and categorical variables in a single dataset which can be used for all analyses. Data will be separately imputed for training and test datasets. The imputation will be done for variables with missingness of less than 20%.

Data management

All continuous variables will be centered and scaled. Continuous variables will be dummy-coded and reference levels will be identified.

3) Building the model

Next, a multivariate prediction model for reliable recovery will be developed using a logistic regression model ('classification model') (Figure 1). The data will be split into train (80%) and test (20%) data. The models will be trained using 10-fold cross validation (repeated 10 times), which iteratively divides our participants into training and test. Lasso, ridge and elastic net regularisation algorithm(s) will be applied to reduce overfitting and to reduce or remove any redundant features in the prediction model. Lasso regularises model parameters by shrinking redundant model parameters, reducing some of them zero, thereby simplifying the model (L1 penalties). Ridge regularisation does not remove any parameters but shrinks less important features towards zero (L2 penalties). In lasso and ridge regressions the hyperparameter lambda λ controls the strength of regularisation applied to the model and

a larger value of lambda results in stronger regularisation. The lambda range that will be tested is = 0, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10.

Elastic net combines the L1 and L2 penalties and the amount of each penalty is controlled by the hyperparameter α. When $\alpha = 0$ the elastic net is equivalent to a ridge regression, whereas when $\alpha = 1$, it is equivalent to a lasso regression. We will run different models testing the following α values: 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.

The performance of each of these will be evaluated to ensure the most appropriate model for the dataset is identified, using different fit indicators. Confusion matrices will be reviewed to obtain measures of accuracy, sensitivity, specificity and positive and negative predictive value. Discrimination measures indicate the ability of a model to correctly discriminate between events and non-events and are typically quantified via AUC (area under the curve) of the receiver operating characteristic (ROC) curve (C-statistic).

The best performing model will be used as the final tuned model, which will then be tested in the hold-out dataset.

Feature importance will be assessed by examining the coefficients of the final model. Additionally, SHapley Additive ExPlanations (SHAP) values will be computed and plotted. The SHAP method is based on cooperative game theory and used to show the contribution or the importance of each feature on the prediction of the model.

All analyses (data exploration, data cleaning and model building) will be carried out using R, version 4.4.1 (2024-06-14), including the following R packages: tidyverse (Wickham et al., 2019), caret (Kuhn, 2008), missRanger (Mayer, 2024) and shapR (Sellereite & Jullum, 2020)

4) Sensitivity and exploratory analyses

Whilst in the primary analyses a prediction model will be built on the total sample, we will perform subgroup analyses by age group (working age v older adults) to identify if there are any age specific differences in prediction accuracy and feature importance.

Further sensitivity analyses may be carried out on high level variables that group aspects of a variable in the main model, such as binary variables for 'other mental health problems' or 'physical health problems. If found an important feature in the prediction model, types of mental health and physical health problems will be explored to assess effects and importance for specific conditions. Similarly, for variables where total and subscale scores exist, such as self-compassion, perceived health and impact of stressful life events, if total score found to be an important feature, subscales will be explored further to assess effect and importance for these.

We may further explore if the addition of variables collected at baseline in this study enhance predictive ability of the model compared to a model that contains routinely collected data only.

References

Amati, F., Banks, C., Greenfield, G., & Green, J. (2017). Predictors of outcomes for patients with common mental health disorders receiving psychological therapies in community settings: a systematic review. *Journal of Public Health*, *40*(3), e375-e387. https://doi.org/10.1093/pubmed/fdx168

Andreas, S., Schulz, H., Volkert, J., Dehoust, M., Sehner, S., Suling, A., Ausín, B., Canuto, A., Crawford, M., Da Ronch, C., Grassi, L., Hershkovitz, Y., Muñoz, M., Quirk, A., Rotenstein, O., Santos-Olmo, A. B., Shalev, A., Strehle, J., Weber, K., . . . Härter, M. (2017). Prevalence of mental disorders in elderly people: The European MentDis_ICF65+ study. *The British Journal of Psychiatry*, *210*(2), 125-131. https://doi.org/10.1192/bjp.bp.115.180463

Athanasakou, D., Karakasidou, E., Pezirkianidis, C., Lakioti, A., & Stalikas, A. (2020). Self-Compassion in Clinical Samples: A Systematic Literature Review. *Psychology*, *Vol.11No.02*, 26, Article 98184. https://doi.org/10.4236/psych.2020.112015

Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Development and Initial Psychometric Evaluation. *J Trauma Stress*, *28*(6), 489-498. https://doi.org/10.1002/jts.22059

Bone, C., Simmonds-Buckley, M., Thwaites, R., Sandford, D., Merzhvynska, M., Rubel, J., Deisenhofer, A. K., Lutz, W., & Delgadillo, J. (2021). Dynamic prediction of psychological treatment outcomes: development and validation of a prediction model using routinely collected symptom data. *Lancet Digit Health*, *3*(4), e231-e240. https://doi.org/10.1016/s2589-7500(21)00018-2

Chambless, D. L., Caputo, G. C., Jasin, S. E., Gracely, E. J., & Williams, C. (1985). The Mobility Inventory for Agoraphobia. *Behaviour Research and Therapy*, *23*(1), 35-44. https://doi.org/https://doi.org/10.1016/0005-7967(85)90140-8

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., Dwyer, D., & Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, *20*(2), 154-170. https://doi.org/10.1002/wps.20882

Connor, K. M., Davidson, J. R., Churchill, L. E., Sherwood, A., Foa, E., & Weisler, R. H. (2000). Psychometric properties of the Social Phobia Inventory (SPIN). New self-rating scale. *Br J Psychiatry*, *176*, 379-386. https://doi.org/10.1192/bjp.176.4.379

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/https://doi.org/10.1016/j.brat.2016.02.003

Erzen, E., & Çikrikci, Ö. (2018). The effect of loneliness on depression: A meta-analysis. *Int J Soc Psychiatry*, *64*(5), 427-435. https://doi.org/10.1177/0020764018776349

Foa, E. B., Kozak, M. J., Salkovskis, P. M., Coles, M. E., & Amir, N. (1998). The validation of a new obsessive–compulsive disorder scale: The Obsessive–Compulsive Inventory. *Psychological Assessment*, *10*(3), 206-214. https://doi.org/10.1037/1040-3590.10.3.206

Hare, D. L., Toukhsati, S. R., Johansson, P., & Jaarsma, T. (2014). Depression and cardiovascular disease: a clinical review. *Eur Heart J*, *35*(21), 1365-1372. https://doi.org/10.1093/eurheartj/eht462

Hilbert, K., Kunas, S. L., Lueken, U., Kathmann, N., Fydrich, T., & Fehm, L. (2020). Predicting cognitive behavioral therapy outcome in the outpatient sector based on clinical routine data: A machine learning approach. *Behaviour Research and Therapy*, *124*, 103530. https://doi.org/https://doi.org/10.1016/j.brat.2019.103530

Hu, T., Zhang, D., & Wang, J. (2015). A meta-analysis of the trait resilience and mental health. *Personality and Individual Differences*, *76*, 18-27. https://doi.org/https://doi.org/10.1016/j.paid.2014.11.039

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*, *16*(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

MacBeth, A., & Gumley, A. (2012). Exploring compassion: A meta-analysis of the association between self-compassion and psychopathology. *Clinical Psychology Review*, *32*(6), 545-552. https://doi.org/https://doi.org/10.1016/j.cpr.2012.06.003

Mueller, C., Thompsell, A., Harwood, D., Bagshaw, P., & Burns, A. (2017). *Mental Health in Older People A Practice Primer*. NHS.

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *Br J Psychiatry*, *180*, 461-464. https://doi.org/10.1192/bjp.180.5.461

NHS Digital. (2016). *Adult Psychiatric Morbidity Survey: Survey of Mental Health and Wellbeing*.

NHS digital. (2019). *Psychological Therapies A guide to IAPT data and publications*. file:///Users/makalani/Downloads/Guide+to+IAPT+data+and+publications.pdf

NHS Digital. (2024). *NHS Talking Therapies, for anxiety and depression, Annual reports, 2022-23*.

Office for National Statistics. (2017). *National population projections: 2016-based*. Office for National Statistics. Retrieved 12th June from https://www.ons.gov.uk/releases/nationalpopulationprojections2016basedstatisticalbulletin

Rodda, J., Walker, Z., & Carter, J. (2011). Depression in older adults. *Bmj*, *343*, d5219. https://doi.org/10.1136/bmj.d5219

Royal College of Psychiatrists. (2018). *Suffering in silence: age inequality in older people's mental health care*. R. C. o. Psychiatrists. https://www.rcpsych.ac.uk/docs/default-source/improving-care/better-mh-policy/college-reports/college-report-cr221.pdf?sfvrsn=bef8f65d_2

Saunders, R., Buckman, J. E. J., Stott, J., Leibowitz, J., Aguirre, E., John, A., Lewis, G., Cape, J., & Pilling, S. (2021). Older adults respond better to psychological therapy than working-age adults: evidence from a large sample of mental health service attendees. *J Affect Disord*, *294*, 85-93. https://doi.org/10.1016/j.jad.2021.06.084

Schmidt, A., Grey, N., Strauss, C., & Gaysina, D. (2024). Predictors of treatment outcome of psychological therapies for common mental health problems (CMHP) in older adults: A systematic literature review. *Clin Psychol Rev*, *112*, 102463. https://doi.org/10.1016/j.cpr.2024.102463

Shear, M. K., Rucci, P., Williams, J., Frank, E., Grochocinski, V., Vander Bilt, J., Houck, P., & Wang, T. (2001). Reliability and validity of the Panic Disorder Severity Scale: replication and extension. *J Psychiatr Res*, *35*(5), 293-296. https://doi.org/10.1016/s0022-3956(01)00028-0

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