Understanding Predictors of Outcome for Guided Self-Help Interventions for Anxiety

Caroline Wojnarowski

A thesis submitted in partial fulfilment of the requirements for the degree of Doctorate in Clinical Psychology

Clinical Psychology Unit, Department of Psychology

The University of Sheffield

May 2024

Declaration

I, the author, confirm this thesis is my own work and that I am aware of the University of Sheffield guidance on unfair means (www.sheffield.ac.uk/new-students/unfair-means). This thesis is submitted in partial fulfilment of the requirements for the degree of Doctorate in Clinical Psychology. This work has not been submitted for any other degree or to any other institution. No funding has been received for this thesis. No conflicts of interests stated. For any enquiries about data or code sharing, please contact cwojnarowski2@sheffield.ac.uk or m.simmonds-buckley@sheffield.ac.uk.

Structure and Word Count

Section One: Literature Review

Excluding references and tables:	7,549
Including references and tables:	12,897

Section Two: Empirical Study

Excluding references and tables:	7,953
Including references and tables:	11,007

Total

Excluding references and tables:	15,502
Including references and tables:	23,904

Lay Summary

Demand for treatment of anxiety is growing and so services need to provide effective and efficient psychological interventions. This is driving the development of more low intensity (LI) interventions, but current evidence suggests efficacy is mixed. Guided self-help (GSH) interventions are designed to help patients manage symptoms via psychoeducational methods and brief contact time with interpersonal support of a facilitator. Given the uptake of LI interventions, it is important to understand what predicts outcome and how this information can be utilised to improve services.

The first chapter explores which psychosocial and treatment characteristics influence how people with anxiety respond to low intensity cognitive behavioural therapy (CBT-GSH). Identification of baseline characteristics supports treatment matching and identification of in-treatment variables supports understanding of how to adapt LI treatment. A search of the existing literature for published studies in this area was completed. Twenty-four studies were found which examined a total of one-hundred-and-sixteen predictor variables. There were no baseline characteristics that were individually consistently associated with CBT-GSH outcomes across studies, and so further research is needed to determine the most suitable candidate for CBT-GSH. Between session engagement and facilitator's experience delivering GSH consistently predicted outcome. This fits with the wider literature on psychotherapy outcomes and suggests future research should focus on adapting CBT-GSH to increase engagement.

With a newly developed version of GSH informed by cognitive analytic therapy (CAT-GSH), the second chapter aims to understand common and differential predictors of treatment outcome following CBT-GSH or CAT-GSH for anxiety and develop a treatment matching algorithm. Considering the impact of patient preference on outcomes was a further aim, alongside understanding the influence of receiving the indicated-optimal treatment on outcomes. Pre-existing data from a patient preference trial completed in a National Health Service (NHS) Talking Therapies services was retrospectively analysed. Separate predictive models were developed for CAT-GSH and CBT-GSH using baseline sociodemographic, clinical and treatment preference variables. Patients were grouped into having optimal GSH vs not having optimal GSH using the patient advantage index (PAI). Receiving optimal GSH improved outcomes within a subgroup, but the PAI-recommended treatment was not influenced by patient preference. This study is the first to apply the PAI approach to recommend GSH interventions, indicating that treatment matching algorithms have the potential to improve outcomes via improved treatment allocation.

Acknowledgements

Firstly, I would like to thank my research supervisor Dr Melanie Simmonds-Buckley for teaching me so many important skills and helping me to feel able to say I've done some machine learning! Thank you also to Dr Steve Kellett for your guidance and expertise in such an important area of research. I have valued working with you both and have learnt so much.

Thank you to everyone that participated in the patient preference trial of CAT-GSH and CBT-GSH in Oldham NHS Talking Therapies. Without you, this project wouldn't exist.

To my best friends for always believing in me and supporting me for the last 11 years through my journey to be a qualified Clinical Psychologist. And to the coolest trainees I know, thank you Katie, Emily & Laura for your endless support and listening ears.

I couldn't have done this without the unwavering support from my family and my partner, Kyle. Thank you to my Mum for always being at the end of the phone when I needed someone to talk to and my Dad, for learning what a Clinical Psychologist is so he could tell everyone at home. But especially, thank you, Kyle, for riding the DClinPsy emotional rollercoaster with me and never giving up on me. Thank you for your continuous encouragement, patience and being my biggest fan over the years. Thank you for putting a smile on my face when I needed it the most.

List of Contents

Declaration ii

Structure and Word Count iii

Lay Summary iv

Acknowledgements vi

Part One: Literature Review **2**

Abstract 3

Introduction 5

Methods 9

Results 14

Discussion 36

Conclusion 44

References 45

Appendices 60

Part Two: Empirical Study **78**

Abstract 79

Introduction 82

Methods 87

Results 101

Discussion 111

Conclusion 119

References 120

Appendices 133

Part One: Literature Review

Predictors of Treatment Outcome for Cognitive Behavioural Therapy Guided Self-Help Interventions for Anxiety Related Disorders: A Systematic Review.

Abstract

Background: Increasing demand for talking therapies for anxiety is driving the development and delivery of effective and efficient treatments. Cognitive behavioural therapy guided self-help (CBT-GSH) is a brief psychoeducational low-intensity psychological intervention for mild-moderate anxiety supported by a practitioner. CBT-GSH has a mixed evidence base, suggesting that it is not beneficial for all patients, and little is known about what influences responsivity to CBT-GSH. Exploring predictors of outcome can help improve patient outcomes by matching patients to treatments or adapting interventions. Therefore, this review aimed to identify those factors that predict CBT-GSH outcomes and consider the implications of these.

Methods: A systematic review (PROSPERO registration number CRD42023418755) was conducted by searching PubMed, PsycInfo and Scopus up to the 27th November 2023. Included studies had to have an adult sample, delivered 1:1 CBT-GSH (not group therapy) for anxiety and studied at least one predictor of outcome. The Critical Appraisal Skills Programme (CASP) checklists were used for risk of bias assessments. Predictors of response were summarised using narrative synthesis.

Results: A total of 24 eligible studies were identified and these investigated 116 predictors of outcome, which were categorized into 10 domains (demographics, baseline severity, comorbidity, within session engagement, between session engagement, engagement with practitioner, other treatments, experience, process, verbal fluency). Forty-eight percent of variables were significant predictors of outcome; however, most were inconclusive across studies or lacked consensus in the direction

of effect. Between session engagement and experience delivering treatment consistently predicted CBT-GSH outcome. Within session engagement and process variables are considered partial predictors of outcome.

Conclusion: Due to moderate levels of risk of bias within the included studies and a lack of reported data, caution should be taken when interpreting the results. This review suggests there are no pre-treatment variables that consistently predict those who might benefit from CBT-GSH. In terms of in-treatment variables, future research should focus on methods to increase between session engagement and understanding how experienced practitioners create better outcomes.

Keywords: *Anxiety, guided self-help, cognitive behavioural therapy, predictors*

Practitioner Points

- GSH practitioners should try to maximise engagement in homework tasks to increase the likelihood of an effective intervention.
- Currently there are no reliable pre-treatment predictors of outcome following CBT-GSH for anxiety, therefore there is insufficient evidence to support the development of a prediction model for CBT-GSH.
- Further research is required to establish robust predictors of response to CBT-GSH for anxiety.

Introduction

Prevalence of Anxiety

Anxiety can be defined as a feeling of “unease” such as feeling worried or experiencing fear. Anxiety can present in various ways, for example a fear of social situations (social anxiety disorder); a specific fear (phobia); a fear about one’s own health (health anxiety) or a general fear (generalised anxiety disorder).

Fifty-nine percent of adults in the UK report experiencing “low” levels of anxiety (ONS, 2024). This prevalence of anxiety disorders has created an increased demand on mental health services around the world. Therefore, services are seeking effective and efficient interventions to treat this increasing clinical population. This is driving the development of more low intensity psychological interventions as they are brief and require less clinician training to deliver, however, current evidence suggests efficacy is mixed.

Low Intensity Psychological Interventions

A low intensity (LI) psychological intervention (i) uses self-help materials (ii) is ≤ 6 hours of contact time with each contact around 30 minutes or less, and (iii) trained practitioners/supporters can provide input (Shafran et al, 2021).

In England, Talking Therapies (previously Improving Access to Psychological Therapies [IAPT]) services provided by the National Health Service (NHS) are designed on the stepped-care principle (NICE, 2011) and therefore deliver LI interventions for individuals with mild-moderate anxiety and depression early in the stepped-care process. LI interventions can be delivered one to one or in groups and include computerised cognitive behavioural therapy (CBT), guided and/or pure self-help, psychoeducation groups and workshop style LI groups. The most common

theory underpinning LI interventions is cognitive-behavioural, but there are examples of other theories such as cognitive-analytic therapy guided self-help (CAT-GSH; Kellett et al., 2022). CBT has been adapted more than other therapies because the concepts translate easily into the psychoeducational context (Gaudiano, 2013).

Guided self-help interventions are defined as those designed to help patients manage symptoms, primarily using a health technology such as self-help books, instructional videos, or interactive interventions using information technology. Interventions are conducted predominantly independent of professional contact. This approach is argued to be more accessible to patients and less demanding on clinician time (Gega et al., 2004). Within the NHS, cognitive behavioural therapy guided self-help (CBT-GSH; Fairburn, 2013) is delivered as a brief intervention (6-8 sessions) for individuals experiencing mild-moderate anxiety. CBT-GSH uses psychoeducational workbooks that contain information, homework and change exercises that are supported by supervised practitioners. The LI practitioner has been likened to the role of a ‘coach’ rather than a traditional ‘therapist’ (Turpin, 2010). Trials of CBT-GSH compared to pure self-help have suggested larger effects when a practitioner is supporting the patient rather than accessing self-help material alone (Furmak et al., 2009; Pleva & Wade, 2007). This suggests that there is a relationship between practitioner involvement and outcomes.

Efficacy of LI Interventions

The evidence for the effectiveness of LI CBT for anxiety disorders is mixed, particularly between post-treatment and follow-up. A recent meta-analysis of LI CBT for generalised anxiety found that all studies favoured LI CBT over controls, suggesting that LI CBT has shown potential as an effective and efficient treatment for mild to

moderate anxiety (Powell et al., 2024). Similarly, Coull & Morris' (2011) meta-analysis found that CBT-GSH was effective at post-treatment, however it had limited effectiveness at follow-up or with clinical samples.

Some clinical trials of CBT-GSH have shown significant effects at reducing anxiety levels at post-treatment (Amin et al., 2020; Delgadillo et al., 2014). However, some have found the effects of CBT-GSH to be non-significant at post-treatment, but significant at 18-months follow-up (Andersson et al., 2012). Yet a longitudinal cohort study found that relapse/recurrence rates were at 65.8% after 2 years post-treatment, suggesting limited durability of low intensity interventions (Delgadillo et al., 2018).

This discrepancy within the evidence base highlights that LI interventions are not a panacea and it is important to better understand who they do work for and who they do not work for. Partly as LI CBT such as CBT-GSH is now widely adopted as an accessible psychological intervention and an alternative to high intensity, face-to-face CBT. During 2021-22, 72.7% of patients accessing NHS Talking Therapies received CBT-GSH (Nicholls, 2023). It should be noted however, that currently in the NHS, CBT-GSH is the only low intensity therapy model available for 1:1 treatment. There is currently no alternative therapy model available for individuals who do not respond to CBT-GSH. Therefore, implications should focus on how to improve outcomes for the populations that research suggests CBT-GSH is not effective for, rather than offering them a different treatment.

Personalised Approaches to Mental Health Care

Precision medicine is an approach to treating health conditions that is tailored to the individual, according to the evidence of effectiveness for each patient adopting a more personalised approach to care (König et al., 2017). In terms of psychological therapies,

this drives the want to provide the best and earliest intervention, matched to the needs of the patient, moving beyond a “one size fits all” approach. Knowing what treatment works for whom could improve outcomes, improve cost-effectiveness of services and maximise the quality of health care (Kosorok & Laber, 2019).

One way that this can be done is exploring patient predictors of outcome for psychological interventions. A predictor is a measured variable that may have a general prognostic relationship with end of treatment outcomes. Variables that could be predictors include clinical features (e.g. baseline severity), sociodemographic factors (e.g. gender) or treatment factors (e.g. number of sessions attended). Current evidence presents these variables as predictors of outcome after psychological treatment for anxiety, for example anxiety severity (Mululo et al., 2012), coping skills (Kunas et al., 2021), early response to treatment (Beard & Delgadillo, 2019) and employment (Schat et al., 2013).

If there are significant relationships between individual characteristics or features and outcome, it is possible to assume others with similar characteristics would benefit from the same intervention. Baseline predictors could enable pre-treatment variables to support treatment matching, whereas predictors based on in-treatment variables can support understanding of how treatment can be adapted/optimised to improve outcomes.

Predictors of anxiety outcomes after traditional ‘high intensity’ CBT have been investigated and most commonly found that baseline symptom severity impacts outcomes (Haby et al., 2006; Kampman et al., 2007; Wergeland et al., 2016). There is some evidence that therapist competence (Fauskanger Bjaastad et al., 2018), full time employment (Hedman et al., 2012), comorbidity (Brandenburg, 2017) and high levels

of safety behaviours (Butler et al., 2021) also predict outcomes following CBT. There is much less understanding of how these factors relate to outcomes after guided self-help CBT for anxiety.

Current Study

The main objective of this review was to synthesise the available evidence from the LI evidence base (i.e., randomised controlled trials and longitudinal cohort studies) that have investigated variables associated with treatment outcome following CBT-GSH for anxiety disorders. Identification of replicated predictors of treatment response that can be measured at baseline would provide initial understanding of the characteristics or features that might support future treatment matching if the GSH versions based on different therapeutic models are developed. Whereas treatment features associated with outcomes could increase understanding of ways to maximise CBT-GSH effectiveness. Therefore, a secondary aim was to consider the implications of knowing these predictors and the ways in which CBT-GSH can be usefully adapted.

Method

Study Protocol

The review has been reported according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA; Page et al., 2021). The PRISMA Reporting Guideline can be found in Appendix A. The systematic review protocol was registered and published in the International Prospective Register of Systematic Reviews (PROSPERO) ahead of conducting the review (protocol ID: PROSPERO 2023: CRD42023418755).

Eligibility Criteria

Using the Population, Intervention, Comparison, Outcomes and Study (PICOS) framework, studies were eligible for inclusion if they met criteria in Table 1 (Eriksen & Frandsen, 2018). Although within NHS Talking Therapies, CBT-GSH is defined as 6-8 sessions, the review aimed to include studies from around the world. Previous reviews have capped the number of GSH sessions at 12 (Cuijpers et al., 2010) and due to the variety of CBT-GSH delivery methods between the studies, this criterion was also adopted for the current review.

Table 1

Inclusion and Exclusion Criteria

	Inclusion Criteria	**Exclusion Criteria**
Population	Patients (aged ≥ 18 years) with a self-reported or clinician assessed anxiety disorder. Diagnosis of an anxiety disorder identified via diagnostic interview, diagnostic criteria manual or by scoring above clinical threshold on a validated screening measure. Diagnosis meets the review criteria for an anxiety disorder i.e., generalised anxiety disorder, phobia, social anxiety disorder, panic disorder.	Patients aged < 18 years. No formal diagnosis of an anxiety disorder/ not scoring above clinical threshold on a validated screening measure.
Intervention	Psychological intervention must be based on CBT principles and be delivered one-to-one.	Intervention is not based on CBT principles and/or is delivered in a group setting.

	Inclusion Criteria	Exclusion Criteria
	Patients must have at least one contact with a practitioner. Number of CBT-GSH sessions/modules must not exceed 12.	Patients have no contact with a practitioner. Number of CBT-GSH sessions/modules exceed 12.
Comparator	Not applicable	Not applicable
Outcome	Standardised measure of anxiety symptoms (e.g., GAD-7, SPIN, BAI), administered at least at baseline and post-intervention. Impact of at least one variable on post-intervention anxiety symptoms is statistically analysed.	Anxiety symptoms not measured using a standardised measure. Anxiety symptoms not measured at baseline and post-intervention. No statistical analysis of at least one variable on post-intervention anxiety symptoms.
Setting	Any outpatient or inpatient setting.	None.
Study Design	Randomised controlled trials or cohort studies. Published in English language.	Qualitative research, grey literature, conference proceedings, presentations or media articles. Not published in English Language,

Note. Beck Anxiety Inventory (BAI), Generalised Anxiety Disorder Scale (GAD-7), Social Phobia Inventory (SPIN)

Search Strategy

The search strategy was developed using best practice guidelines (Centre for Reviews and Dissemination, 2009). Three databases (PubMed, PsycNet, Scopus) were searched on the 9thJune 2023 for relevant terms using variations of the key words ‘anxiety’, ‘cognitive behaviour therapy’, ‘self-help’ and ‘low intensity’ as defined by Coull & Morris (2011). The full search strategy can be found in Appendix B. Terms

related to 'predictor' were not included in the search terms as this increased the risk of missing studies as predictor analyses were not always the primary focus of a study outlined in the title/abstract (i.e., were reported as secondary analyses). An updated search was completed in November 2023 which found 119 new studies, however, none of these matched the inclusion criteria. This review focussed on published peer reviewed studies only (excluding grey literature), due to the high number of retrieved studies and to ensure quality.

Selection of Articles

Database results were combined, and any duplicates were removed. Titles and abstracts of articles identified by the search strategy were screened for eligibility by the primary reviewer. A second reviewer screened 10% of these titles and abstracts to check inter-rater agreement and any disagreements were discussed and resolved by consensus (98% agreement). Full texts of the remaining articles were sourced and again assessed for eligibility by the primary reviewer with 10% screened by a 2nd reviewer (87% agreement). Forward and backward citation searching was performed for all included studies to identify any additional articles missed by the database search. Study selection was outlined in a PRISMA flow diagram (Moher et al., 2009) using software developed by Haddaway et al. (2022).

Quality Assessment

Eligible papers were independently assessed for risk of bias by two reviewers, using the Critical Appraisal Skills Programme (CASP) Checklists (CASP, 2022; see Appendices C to D) for randomised controlled trials (RCTs) and longitudinal cohort studies. The CASP checklists assess validity of study design, sound methodology, clarity of results and the impact of results locally.

As there is no guidance to categorise risk of bias from the CASP Checklists, this study used the number of “yes” responses in the CASP Checklist to categorise studies. For both checklists, a “yes” response would indicate something of good quality and thus the more “yes” responses, the lower the risk of bias. For RCTs: 0-3 high risk; 4-6 medium – high risk; 7-8 medium risk; 9-11 low to medium risk and 12-13 low risk. For cohort studies: 0-3 high risk; 5-4 medium to high risk; 6-7 medium; 8-10 low to medium risk and 11-12 low risk. A second reviewer assessed 20% of the included studies for risk of bias. Assessment of inter-rater reliability was conducted using Cohen’s Kappa (Cohen, 1960) and this was categorised as “substantial” agreement, k = .79.

Data Extraction and Synthesis

Data was extracted and tabulated using the Cochrane Collaboration Data Collection form (Higgins & Green, 2011). The following data was extracted from the papers – number of participants, treatment name, treatment delivery, comparator treatment, practitioner credentials, treatment duration, practitioner input, outcome measure, diagnosis, diagnostic criteria, variables assessed as predictors, number of predictors analysed, results and significance. Three studies had missing p-values and effect sizes (Andersson et al., 2008; Berger et al., 2013; Haug et al., 2015). The authors were contacted on the 9th of October 2023 and given 3 weeks to respond. Unfortunately, they did not respond, and the missing information remains.

It was not possible to apply meta-analytic methods for predictor variables due to the under-reporting of statistical information or lack of replication in sufficient studies. Therefore, a detailed narrative synthesis of the results across studies is presented. Predictors used across included studies were categorised into relevant groups.

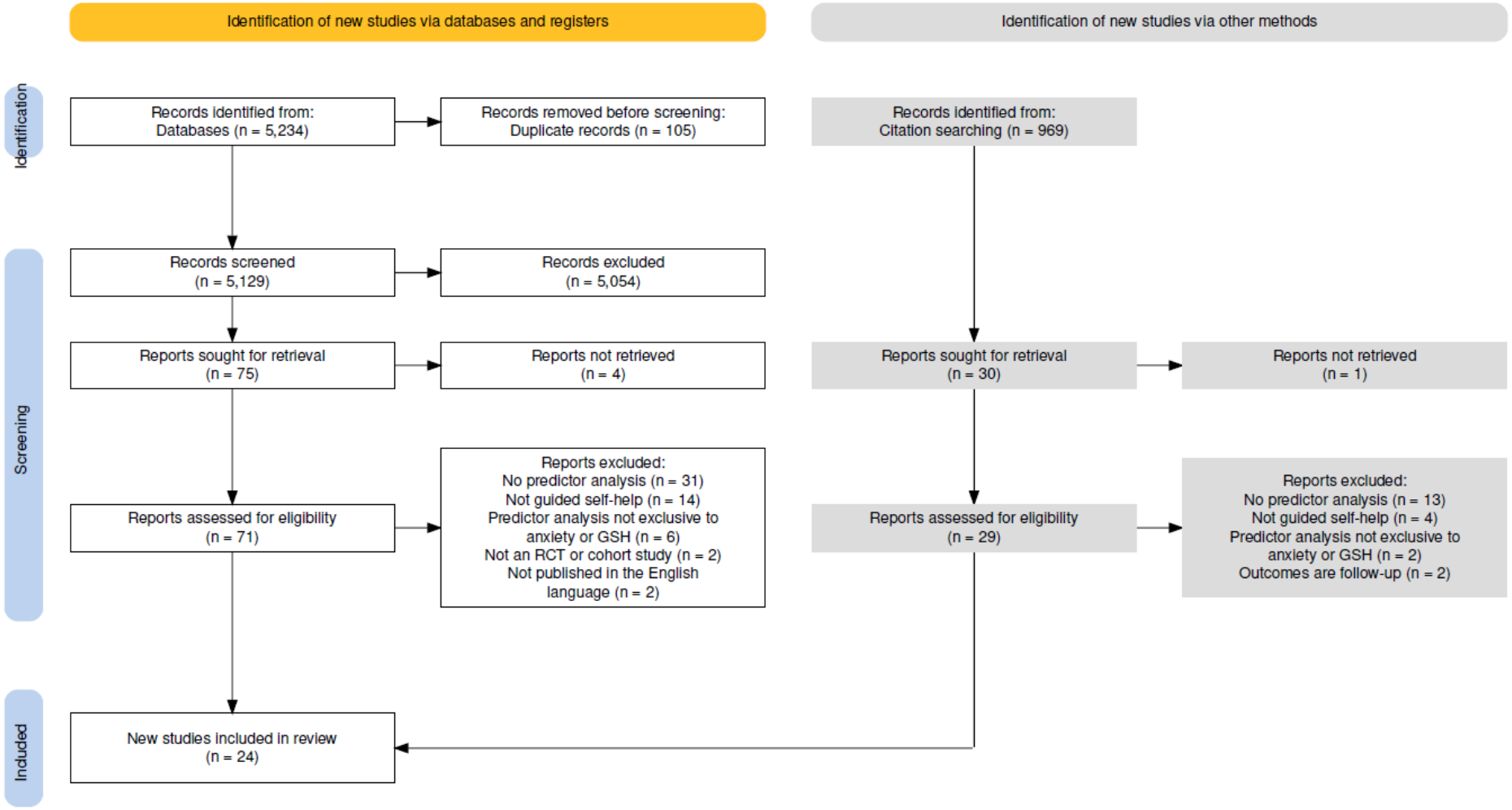
Pearson's *r* was calculated for as many predictors as possible to present a standardised metric for each outcome, these are then categorised into small, medium and large effect sizes according to Cohen (2013). This was completed using online software (Lenhard & Lenhard, 2016). This was not possible for all predictors due to missing information, such as the study reported predictors as "non-significant" but with no p-values or due to restrictions from the formulas. An example of this is for beta coefficients, the calculator could only compute an *r* value if the coefficient was between -0.5 and 0.5. The software was also unable to convert Spearman's Rho (Spearman, 1904) to Pearson's *r* and so effect size was categorised using Rea & Parker's (1992) criteria. Significance values are presented in tables for each predictor category along with effect size and quality assessment.

Results

The search strategy returned 5138 unique titles and abstracts. The most common reasons for exclusion during screening were populations of adolescents/children; treatments other than guided self-help CBT; no predictors of outcomes analysed and disorders other than anxiety. Eighty-four full-text articles were assessed for eligibility. Common reasons for exclusion included predicting long term effectiveness, not end of treatment outcome; self-help treatment was unguided; no predictors of outcome were analysed. Sixteen articles were identified as eligible for inclusion in the review. Forward and backward searches generated an additional 969 articles, of which 8 met the above eligibility criteria for review. As a result, k = 24 studies were included in the systematic review. Figure 1 shows the PRISMA diagram summarising study selection and exclusion reasons.

Figure 1

PRISMA Diagram for Selection of Articles



Study Characteristics

There were (k = 11) longitudinal cohort studies and (k = 13) RCTs. Studies were conducted across a range of countries – Australia (k = 2), Canada (k = 1), China (k = 1), Denmark (k = 1), India (k = 1), Israel (k = 1), Norway (k = 4), Spain (k = 1), Sweden (k = 6), Switzerland (k = 2), United Kingdom (k = 1) and the United States of America (k = 3). Most studies used screening tools and/or diagnostic interviews to determine a diagnosis of an anxiety disorder, however three studies used only a screening tool and one study did not require a formal diagnosis. Most of the studies used participants with a diagnosis of social anxiety (k = 12), with studies of panic disorder, generalised anxiety disorder and anxiety each having 4 studies each. Due to differences in diagnosis between the studies, measures of outcome ranged considerably (see Table 2). The total pooled sample size (n) across the included studies was 10,828.

Treatments ranged from 1-12 sessions and were delivered mostly via the internet, except for one study of bibliotherapy which was a workbook with assistance from a facilitator. Facilitator contact was via emails, calls, face-to-face contact or messages through an online system. This contact ranged from ad-hoc contact, receiving feedback after homework, 10–20-minute calls and weekly emails. Facilitator credentials differed significantly between studies. Most studies used mental health professionals (k = 15*)* (e.g., Clinical Psychologists, Psychiatrists, CBT Therapists, Psychiatric Nurses). Six studies used psychology students and two studies did not require qualifications or used peer mentors. One study did not report therapist credentials.

The two most common statistical method used to examine potential predictors was multiple regression (k = 4) and logistic regression (k = 4), followed by correlations (k =

3). These methods all assess linear relationships between two variables and can assess whether a potential predictor has a relationship with outcome. Other studies used a variety of statistical methods such as chi-square, linear mixed-effect models, pairwise comparisons, multilevel modelling, linear and quadratic effects, piecewise growth model, repeated measures ANOVA, spearman's rank correlation coefficient and a T-test.

Table 2

Study Characteristics

Study	Design	Diagnostic Criteria	Diagnosis	Comparison Condition(s)	N Analysed	Mean Age	Treatment Duration (sessions/modules)	Anxiety Outcome Measure	N Predictors Reported
Baigent et al. (2023)	Cohort	GAD-7	Anxiety	-	2732	43.4	6	GAD-7	4
Nordgren et al. (2013)	RCT	DSM-IV	Anxiety	-	27	39.3	10	CORE-OM	3
Lawn et al. (2019)	Cohort	ICD-10	Anxiety	-	427	-	6 - 8	GAD-7	3
Dryman et al. (2017)	Cohort	-	SAD	-	3384	29.82	12	SPIN	7
González-Robles et al. (2021)	RCT	DSM-IV	Emotional disorder	TAU	63	-	12	BAI	6
Berger et al. (2013)	RCT	DSM-IV	SAD, PD, Agoraphobia, GAD	Waitlist Control	88	35/34.4	8	BAI	10
Chen et al. (2020)	Cohort	DSM-IV	SAD	-	107	24.78	8	SIAS & SPS	18
Nordmo et al. (2015)	RCT	DSM-IV	SAD	ICBT with face to face	23	25.6	9	SIAS & SPS	2
Hedman et al. (2013)	Cohort	DSM-IV	PD	-	451	37.3	10	PDSS-SR	1
Schønning & Nordgreen (2021)	Cohort	DSM-IV	SAD or PD	-	575	31.8	9	BSQ or SPS	2
Newman et al. (2021a)	RCT	DSM-V	GAD	Waitlist Control	117	20.1	40 (days)	DASS & PSWQ	1
El Alaoui et al. (2015a)	Cohort	DSM-IV	SAD	-	764	32.51	-	LSAS-SR	5
Haug et al. (2015)	RCT	DSM-IV	PD or SAD	Manualised CBT	85	32.4	1 - 10	CSR & SRC	11

Table 2

Study Characteristics

Study	Design	Diagnostic Criteria	Diagnosis	Comparison Condition(s)	N Analysed	Mean Age	Treatment Duration (sessions/modules)	Anxiety Outcome Measure	N Predictors Reported
El Alaoui et al. (2015)	Cohort	DSM-V	SAD	-	547	32.6	11	LSAS-SR	2
Andersson et al. (2008)	RCT	DSM-IV	PD	Manualised CBT	25	34.2	10	ACQ & BSQ	8
Newman et al. (2021)	RCT	DSM-V	GAD	Waitlist Control	50	21.62	40 (days)	DASS; PSWQ; STAI-IT	5
Seeley et al. (2017)	RCT	GAD-7	Anxiety	Waitlist Control	55	74.2	10	GAD-7	6
Lindegaard et al. (2020)	RCT	DSM-IV	SAD	Psychodynamic psychotherapy	13	41.4	10	LSAS-SR & GAD-7	2
Hadjistavropoulos et al. (2016)	Cohort	DSM-IV	GAD	-	58	40.22	12	GAD-7	1
Shalom et al. (2020)	RCT	DSM-V DSM-V	SAD	Waitlist Control	101	31.2	11	LSAS-SR & SPIN	1
Stolz et al. (2018)	RCT		SAD	Waitlist Control	120	34.6/34.7	8	SPS; SIAS; LSAS-SR	1
Schulz et al. (2016)	RCT	DSM-IV	SAD	Waitlist Control	149	35.38	8	SPS; SIAS	8
Gellatly et al. (2018)	Cohort	GAD-7	Anxiety	-	724	-	6 - 12	GAD-7	1
Mathiasen et al. (2018)	Cohort	ICD-10	SAD, PD, Agoraphobia, GAD	-	143	36.8	9	GAD-7	8

Note. RCT = Randomised Controlled; GAD-7 = Generalised Anxiety Disorder Assessment; DSM-IV = Diagnostic and Statistical Manual of Mental Disorders; ICD-10 = International Classification of Diseases 10th Revision; SAD = Social Anxiety Disorder; GAD = Generalised Anxiety Disorder; PD = Panic Disorder; TAU = treatment as usual; CORE-OM = Clinical Outcomes in Routine Evaluation – Outcome Measure; SPIN = Social Phobia Inventory; BAI = Becks Anxiety Inventory; SIAS = Social Interaction Anxiety

Scale; SPS = Social Phobia Scale; PDSS-SR = Panic Disorder Severity Scale – Self Report; BSQ = Body Sensations Questionnaire; DASS = Depression Anxiety Stress Scales; PSWQ = Penn State Worry Questionnaire; LSAS-SR = Liebowitz Social Anxiety Scale – Self Report; ACQ = Anxiety Control Questionnaire; STAI-IT = State-Trait Anxiety Inventory

Risk of Bias Assessment

Of the RCTs, five studies were considered to have low to medium risk of bias, six medium and two medium-to-high. A consistent methodological feature creating risk of bias was the lack of allocation concealment and blinding due to the nature of the research. There were no power analyses completed specifically for the predictor analyses. Another common reason for potential risk of bias included missing statistical information such as p-values and effect sizes. Other reasons consist of samples with poor generalisability (e.g., university students), small sample sizes and missing information regarding the methods and interventions used.

For the longitudinal cohort studies, four were considered to have low risk of bias, five low to medium and two medium. All of these studies used validated and reliable outcome measures; however, some were self-report which raises questions of validity due to social desirability bias. The most common methodological problem was the lack of follow-up or lack of information about follow-up. A further issue was the precision of the results as many of the studies lacked information about confidence intervals and effect sizes. A summary of further details about all the risk of bias assessments can be found in Appendix E.

Predictors of Outcome

The studies investigated a total of 116 predictors of outcome. Some predictors were consistently assessed across studies such as age, gender and number of sessions attended. However, the results described in the final reports were insufficient for a large proportion of the variables to complete a robust meta-analysis. Clinical variables (e.g., comorbidity, baseline anxiety, social functioning) and treatment variables (e.g., time spent in programme, number of challenged negative thoughts and working

alliance) were investigated more than demographic variables (e.g., employment status, marital status etc.). All significant and non-significant predictors were grouped into ten categories: demographics (k = 6); baseline severity (k = 6); comorbidity (k = 4); within session engagement (k = 13); between session engagement (k = 3); engagement with practitioner (k = 5); other treatments (k = 1); experience (k = 2); process (k = 10) and verbal fluency (k = 1). Results were narratively synthesised across the 10 predictor categories. Some authors are repeated within these tables if their predictors were examined on more than one outcome measure, for example Chen et al. (2020) assessed predictors against the Social Phobia Scale (SPS) and the Social Interaction Anxiety Scale (SIAS) but these are the same participants.

Demographics

Six studies examined demographics at the beginning of treatment as a predictor of outcome. The p-values, Pearson's r and risk of bias is for each of these predictors are summarised in Table 3. Direction of effect is reported were possible, this was difficult to interpret where studies have not indicated how they transformed the categorical variables.

Table 3

Demographic Predictors of Outcome

Study	Predictor	P	r	Effect size	Direction of effect (better outcomes)	Risk of bias
Baigent et al. (2023)	Gender	0.24	0.03	Small	-	Low-Medium
Chen et al. (2020) (SIAS)	Gender	0.02*	0.28	Small	Females	Medium
Chen et al. (2020) (SPS)	Gender	0.26	0.16	Small	Females	Medium

Study	Predictor	P	r	Effect size	Direction of effect (better outcomes)	Risk of bias
Mathiason et al. (2018)	Gender	0.84	0.24	Small	-	Low-Medium
Dryman et al. (2017)	Gender	0.74	0.01	Small	-	Low-Medium
Berger et al. (2013)	Gender	Non-sig	-	-	-	Medium
Baigent et al. (2023)	Age	0.15	0.003	Small	-	Low-Medium
Chen et al. (2020) (SIAS)	Age	0.12	0.2	Small	Younger	Medium
Chen et al. (2020) (SPS)	Age	0.05*	0.24	Small	Younger	Medium
El Alaoui et al. (2015a)	Age	Non-sig	-	-	-	Low-Medium
Mathiason et al. (2018)	Age	0.99	0.21	Small	-	Low-Medium
Berger et al. (2013)	Age	Non-sig	-	-	-	Medium
Baigent et al. (2023)	Employment Status	0.08	0.06	Small	Employed	Low-Medium
Chen et al. (2020) (SIAS)	Student Status	0.19	0.18	Small	Student	Medium
Chen et al. (2020) (SPS)	Student Status	0.1	0.21	Small	Student	Medium
Chen et al. (2020) (SIAS)	Employed	0.47	0.12	Small	Unemployed	Medium
Chen et al. (2020) (SPS)	Employed	0.31	0.15	Small	Unemployed	Medium
El Alaoui et al. (2015a)	Employment Status	<0.05*	-	-	-	Low-Medium
Mathiason et al. (2018)	Employment Status	<0.001*	-	-	Unemployed	Low-Medium
Baigent et al. (2023)	Relationship Status	0.002*	0.08	Small	Married/de facto	Low-Medium
Berger et al. (2013)	Relationship Status	Non-sig	-	-	-	Medium
Mathiason et al. (2018)	Relationship Status	<0.001*	-	-	Married/de facto	Low-Medium
Chen et al. (2020) (SIAS)	Education Level	0.58	0.11	Small	More educated	Medium
Chen et al. (2020) (SPS)	Education Level	0.32	0.16	Small	More educated	Medium

Study	Predictor	P	r	Effect size	Direction of effect (better outcomes)	Risk of bias
Berger et al. (2013)	Education Level	Non-sig	-	-	-	Medium
Mathiason et al. (2018)	Education Level	0.05*	-	-	More educated	Low-Medium
Chen et al. (2020) (SIAS)	Monthly Income	0.56	0.11	Small	More income	Medium
Chen et al. (2020) (SPS)	Monthly Income	0.71	0.09	Small	Less income	Medium

**Significance p<0.05*

Gender was examined as a predictor in five studies; however, Chen et al. (2020) investigated it twice for two different outcome measures (Social Interaction Anxiety Scale & Social Phobia Scale). Gender was a non-significant predictor in five studies (Baigent et al., 2023; Berger et al., 2013; Chen et al., 2020; Dryman et al., 2017; Mathiason et al., 2018). However, in Chen et al. (2020), gender was significant when outcome was assessed using the Social Interaction Anxiety Scale (SIAS; Mattick & Clarke, 1998), suggesting that females had better outcomes than males. Most studies didn't report the direction of effect and so it cannot be concluded that there are any patterns in the relationship between gender and treatment outcome.

A similar trend was found for age, it was a non-significant predictor in five studies (Baigent et al., 2023; Berger et al., 2013; Chen et al., 2020; El Alaoui et al., 2015a; Mathiason et al., 2018). However, Chen et al. (2020) found that younger participants had better outcomes when the outcome was assessed using the Social Phobia Scale (SPS; Mattick & Clarke, 1998). Most studies didn't report the direction of effect and so it cannot be concluded that there are any patterns in the effect of age on outcome.

Four studies investigated employment status as a predictor of outcome and results were mixed. Two studies found it be a non-significant (Baigent et al., 2023; Chen et

al., 2020) and two studies found it to be significant (El Alaoui et al., 2015a; Mathiason et al., 2018). Although, these were contradictory in their direction as two studies found unemployed participants had better outcomes, whilst Baigent et al. (2023) found that employed participants had better outcomes.

Relationship status was examined by three studies with two finding significance (Baigent et al., 2023; Mathiason et al., 2018) and were consistent in their direction i.e., Baigent et al. (2023) found that being divorced, separated or single was associated with a decreased chance of reliable improvement and Mathiason et al. (2018) found that married participants or those living with a partner had better outcomes. Berger et al. (2013) did not find relationship status to be a significant predictor of outcome. This highlights a pattern in the direction of effect, that individuals that are married/living with a partner are more likely to have better outcomes. However, the effect size is small and one study did find this relationship to be non-significant.

Education level was mostly non-significant (Berger et al., 2013; Chen et al., 2020). However, Mathiason et al. (2018) and Chen et al. (2020) found that participants with the highest education level were more likely to have better outcomes after CBT-GSH, albeit with a small effect size (or effect size not reported).

Chen et al. (2020) examined monthly income as a predictor but on both the SIAS and the SPS, this was non-significant and the direction of effect was mixed.

Baseline Severity

Seven studies examined baseline severity at the beginning of treatment as a predictor of outcome. The p-values, Pearson's r and risk of bias for each of these predictors are summarised in Table 4. Direction of effect is reported were possible.

Table 4

Baseline Severity Predictors of Outcome

Study	Predictor	P	r	Effect size	Direction of effect (better outcomes)	Risk of bias
Lawn et al. (2019)	Baseline Anxiety	0.019*	0.14	Small	Mild-Moderate	Low
Mathiason et al. (2018)	Baseline Anxiety	<0.001*	0.41	Medium	Higher	Low-Medium
González-Robles et al. (2021)	Baseline Anxiety	<0.001*	0.48	Medium	Higher	Low-Medium
Mathiason et al. (2018)	Baseline Depression	0.01*	0.23	Small	Higher	Low-Medium
Lawn et al. (2019)	Baseline Depression	0.008*	0.15	Small	Moderate	Low
González-Robles et al. (2021)	Baseline Depression	Non-sig	0.05	Small	Higher	Low-Medium
Dryman et al. (2017)	Baseline Social Phobia	0.03*	0.11	Small	Lower	Low-Medium
Haug et al. (2015) (CR)	Baseline Social Functioning	<0.01*	0.31	Medium	Higher	Medium-High
Haug et al. (2015) (SR)	Baseline Social Functioning	<0.01*	0.36	Medium	Higher	Medium-High
Haug et al. (2015) (CR)	Baseline Level of Impairment	<0.05*	0.28	Small	Lower	Medium-High
Haug et al. (2015) (SR)	Baseline Level of Impairment	Non-sig	0.17	Small	Lower	Medium-High
Andersson et al. (2008) (ACQ)	Baseline Agoraphobic Avoidance	Non-sig	0.05	Small	Lower	Medium
Andersson et al. (2008) (BSQ)	Baseline Agoraphobic Avoidance	Non-sig	0.04	Small	Lower	Medium
Haug et al. (2015) (CR)	Baseline Severity	<0.001*	0.51	Large	Higher	Medium-High
Haug et al. (2015) (SR)	Baseline Severity	<0.001*	0.59	Large	Higher	Medium-High
El Alaoui et al. (2015a)	Baseline Functioning	<0.001*	-	-	-	Low-Medium

**Significance p<0.05*

Within this category there were multiple significant predictors found, however there are discrepancies in whether low baseline severity or high baseline severity predicted better outcomes.

Lawn et al. (2019) found that mild to moderate levels of anxiety and depression predicted higher rates of recovery. Similarly, Dryman et al. (2017) found that participants with lower scores on the Social Phobia Inventory (SPIN; Connor et al., 2000) at baseline, were more likely to achieve reliable and clinically significant change. Haug et al. (2015) found that participants with lower levels of general impairment and higher levels of social functioning at baseline were more likely to have better outcomes. And within the same study, a baseline severity assessment found that higher severity was associated with worse outcomes, both clinician rated and self-report.

However, Mathiason et al. (2018) found that higher baseline levels of anxiety and depression indicated larger improvements. González-Robles et al. (2021) also found that participants with higher baseline anxiety were more likely to improve after treatment.

Comorbidity

Four studies examined comorbidity at the beginning of treatment as a predictor of outcome. The p-values, Pearson's r and risk of bias is for each of these predictors are summarised in Table 5. Direction of effect of comorbid diagnoses is reported were possible.

Table 5

Comorbidity Predictors of Outcome

Study	Predictor	P	r	Effect size	Direction of effect	Risk of bias
Chen et al. (2020) (SIAS)	Comorbid Depression	0.28	0.16	Small	Better outcome	Medium
Chen et al. (2020) (SPS)	Comorbid Depression	0.82	0.06	Small	Better outcome	Medium
Haug et al. (2015) (SR)	Comorbid Depression	<0.05*	0.23	Small	Better outcome	Medium-High
Chen et al. (2020) (SIAS)	Comorbid Anxiety	0.07	0.23	Small	Worse outcome	Medium
Chen et al. (2020) (SPS)	Comorbid Anxiety	0.06	0.24	Small	Worse outcome	Medium
Haug et al. (2015) (CR)	Comorbid Anxiety	<0.05*	0.29	Small	Better outcome	Medium-High
Haug et al. (2015) (CR)	Comorbid Personality Disorder	<0.01*	0.29	Small	Better outcome	Medium-High
Andersson et al. (2008) (ASQ)	Comorbid Personality Disorder	<0.05*	0.48	Medium	Worse outcome	Medium
Andersson et al. (2008)	Comorbid Personality Disorder	Non-sig	0.39	Medium	Worse outcome	Medium
Haug et al. (2015) (CR)	Comorbid PD or SAD	Non-sig	-	-	Worse outcome	Medium-High
Haug et al. (2015) (SR)	Comorbid PD or SAD	<0.05*	0.21	Small	Worse outcome	Medium-High
Berger et al. (2013)	Comorbidity Presence	Non-sig	-	-	-	Medium

**Significance p<0.05; PD = Panic Disorder; SAD = Social Anxiety Disorder*

Having a comorbid diagnosis of depression or anxiety (if the study examined an alternative anxiety disorder e.g., phobia) were found to be significant predictors of a worse outcome (Haug et al., 2015) but also were non-significant predictors of a better outcome (Chen et al., 2020). Personality disorder was a consistently significant predictor across two studies (Andersson et al., 2008, Haug et al., 2015), however in different directions. Andersson et al. (2008) found that a diagnosis was related to worse outcomes, whereas in Haug et al. (2015), results suggest that a personality

disorder diagnosis was a positive predictor of residual change scores. Haug et al. (2015) also found that having a diagnosis of social anxiety disorder or panic disorder significantly predicted worse outcomes but only when the outcome measure was self-reported. The relationship between comorbidity and outcome, mostly had a small effect size apart from Andersson et al. (2008) with a medium effect. However, the direction of effect is mixed across the studies and so regardless of significance, it cannot be concluded whether having a comorbid diagnosis predicts better or worse outcomes.

Within Session Engagement

Fourteen studies examined levels of engagement within treatment as a predictor of outcome. The p-values, Pearson's r and risk of bias is for each of these predictors are summarised in Table 6. Direction of effect is reported were possible.

Table 6

Within Session Engagement Predictors of Outcome

Study	Predictor	P	r	Effect size	Direction of effect (better outcome)	Risk of bias
Lawn et al. (2019)	No. Sessions Attended	<0.001*	0.31	Medium	More sessions	Low
Gellatly et al. (2018)	No. Sessions Attended	<0.01*	0.42	Medium	More sessions	Medium
Newman et al. (2021)	No. Sessions Completed	Non-sig	0.12	Small	More sessions	Low-Medium
Berger et al. (2013)	No. Sessions Completed	<0.01*	0.27	Small	More sessions	Medium
Chen et al. (2020) (SIAS)	No. Completed Modules	0.005*	0.32	Medium	More modules	Medium
Chen et al. (2020) (SPS)	No. Completed Modules	0.15	0.19	Small	More modules	Medium

Study	Predictor	P	r	Effect size	Direction of effect (better outcome)	Risk of bias
Newman et al. (2021a)	Time in Programme	0.004*	0.2	Small	More time	Medium-High
Nordmo et al. (2015) (SPS)	Time in Programme	<0.001*	0.43	Medium	More time	Medium
Nordmo et al. (2015) (SIAS)	Time in Programme	<0.001*	0.33	Medium	More time	Medium
Schulz et al. (2016) (SPS)	Time in Programme	0.011*	-	-	-	Low-Medium
Schulz et al. (2016) (SIAS)	Time in Programme	0.039*	-	-	-	Low-Medium
Mathiason et al. (2018)	Time in Programme	0.35	0.05	Small	-	Low-Medium
Dryman et al. (2017)	Days in Programme	<0.001*	0.26	Small	More days	Low-Medium
El Alaoui et al. (2015)	Treatment Adherence	0.001*	0.07	Small	Better adherence	Low
El Alaoui et al. (2015a)	Treatment Adherence	<0.001*	-	-	-	Low-Medium
Dryman et al. (2017)	Graduated Programme	<0.001*	0.41	Medium	Graduates	Low-Medium
Seeley et al. (2017)	No. Completed Workbooks	Non-sig	0.25	Small	More workbooks	Medium
González-Robles et al. (2021)	No. logins	Non-sig	0.15	Small	More logins	Low-Medium
Newman et al. (2021)	No. visits to platform	Non-sig	0.12	Small	More visits	Low-Medium
Newman et al. (2021)	Login time	Non-sig	0.2	Small	More logins	Low-Medium
Seeley et al. (2017)	Programme Satisfaction	Non-sig	0.26	Small	Higher satisfaction	Medium
Seeley et al. (2017)	Workbook Usability	Non-sig	0.17	Small	Higher usability	Medium

**Significance p<0.05*

Number of sessions attended/modules completed was a significant predictor of outcome in most of the studies that examined it as a predictor (Berger et al., 2013; Chen et al., 2020; Gellatly et al., 2018; Lawn et al., 2019), apart from Newman et al. (2021) and Chen et al. (2020) when the outcome measure was the SPS. Related to

this, whether the participant “graduated” from the programme was a significant predictor (Dryman et al., 2017).

Time spent in the programme and time engaging with the treatment were mostly all significant predictors of outcome (Dryman et al., 2017; Newman et al., 2021a; Nordmo et al., 2015; Schulz et al., 2016) apart from Mathiason et al. (2018). Similarly, treatment adherence was consistently a significant predictor of outcome (El Alaoui et al., 2015; El Alaoui et al., 2015a). All other variables of within session engagement were non-significant.

Regardless of significance, across all variables categorised as within session engagement, where reported, engaging with the treatment more was consistently related to better treatment outcome with small to medium effect sizes.

Between Session Engagement

Three studies examined levels of engagement between treatment sessions as a predictor of outcome. The p-values, Pearson’s r and risk of bias is for each of these predictors are summarised in Table 7. Direction of effect is reported were possible.

Table 7

Between Session Engagement Predictors of Outcome

Study	Predictor	P	r	Effect size	Direction of effect (better outcome)	Risk of bias
Dryman et al. (2017)	No. of Exposures	0.008*	0.15	Small	More exposures	Low-Medium
Stolz et al. (2018)	No. of Exposures	<0.01*	-	-	More exposures	Low-Medium
Schulz et al. (2016) (SPS)	No. of Exposures	0.004*	-	-	-	Low - Medium

Schulz et al. (2016) (SIAS)	No. of Exposures	0.026*	-	-	-	Low - Medium
Schulz et al. (2016) (SPS)	No. Diary Entries	0.026*	-	-	-	Low - Medium
Schulz et al. (2016) (SIAS)	No. Diary Entries	0.02*	-	-	-	Low - Medium
Schulz et al. (2016) (SPS)	No. Challenged Thoughts	0.012*	-	-	-	Low - Medium
Schulz et al. (2016) (SIAS)	No. Challenged Thoughts	0.033*	-	-	-	Low - Medium

**Significance p<0.05*

All potential predictors within this category were significantly associated with outcome. Dryman et al. (2017), Schulz et al. (2016) and Stolz et al. (2018) all found that participants that engaged better with the between session tasks such as number of entries into their anxiety diary and number of exposures, significantly predicted better outcomes.

Engagement With the Practitioner

Five studies examined levels of engagement with practitioner as a predictor of outcome. The p-values, Pearson's r and risk of bias is for each of these predictors are summarised in Table 8. Direction of effect is reported were possible.

Table 8

Engagement With Practitioner Predictors of Outcome

Study	**Predictor**	**P**	**r**	**Effect size**	**Direction of effect (better outcomes)**	**Risk of bias**
Seeley et al. (2017)	No. Peer Mentor Sessions	Non-sig	0.18	Small	More sessions	Medium
Dryman et al. (2017)	No. Calls with Coach	0.03*	-	-	More calls	Low-Medium

González-Robles et al. (2021)	No. Reviews	Non-sig	0.14	Small	More reviews	Low-Medium
González-Robles et al. (2021)	No. Calls	Non-sig	0.01	Small	Less calls	Low-Medium
González-Robles et al. (2021)	Duration of Calls	Non-sig	0.06	Small	Longer calls	Low-Medium
Newman et al. (2021)	No. Messages to Coach	Non-sig	0.09	Small	More messages	Low-Medium
Newman et al. (2021)	No. Messages from Coach	Non-sig	0.12	Small	More messages	Low-Medium
Berger et al. (2013)	No. Messages	Non-sig	-	-	-	Medium

**Significance p<0.05*

Dryman et al. (2017) found that the greater number of calls the participants had with the practitioner the better their outcome, but none of the other studies found any significant associations between contacts with their coach and outcome (Berger et al., 2013; González-Robles et al., 2021; Newman et al., 2021; Seeley et al., 2017). However, apart from González-Robles et al. (2021) who found that having less calls was related to better outcomes, the direction of effect in the rest of the studies suggested that more contact was related to better outcomes, albeit with small effect sizes. Due to this discrepancy and lack of significance, it cannot be concluded that contact with the practitioner was consistently associated with better treatment outcomes.

Process

Ten studies examined the process within CBT-GSH as a predictor of outcome. The p-values, Pearson's r and risk of bias is for each of these predictors are summarised in Table 9. Direction of effect is reported were possible.

Table 9

Process Predictors of Outcome

Study	Predictor	P	r	Effect size	Direction of effect (better outcome)	Risk of bias
Seeley et al. (2017)	Working Alliance	Non-sig	0.07	Small	Higher WA	Medium
Nordgren et al. (2013)	Working Alliance (pre)	0.43	0.17	Small	Higher WA	Low-Medium
Nordgren et al. (2013)	Working Alliance (mid)	0.019*	0.47	Medium	Higher WA	Low-Medium
Nordgren et al. (2013)	Working Alliance (post)	0.037*	0.42	Medium	Higher WA	Low-Medium
Berger et al. (2013)	Working Alliance	0.33	0.15	Small	Higher WA	Medium
Hadjistavropoulos et al. (2016)	Working Alliance (mid)	0.12	0.13	Small	Higher WA	Medium
Lindegaard et al. (2020)	Working Alliance (mid)	0.001*	0.56	Large	Higher WA	Low-Medium
Andersson et al. (2008) (ACQ)	Treatment Credibility	Non-sig	0.01	Small	Higher credibility	Medium
Andersson et al. (2008) (BSQ)	Treatment Credibility	Non-sig	0.1	Small	Higher credibility	Medium
El Alaoui et al. (2015a)	Treatment Credibility	<0.01*	-	-	Higher credibility	Low-Medium
Seeley et al. (2017)	Change in PHQ-9	Non-sig	0.32	Medium	Larger change	Medium
Dryman et al. (2017)	Social Phobia after Cognitive Restructuring	<0.001*	0.26	Small	Lower	Low-Medium
Shalom et al. (2020)	Sudden Gains	0.045*	0.2	Small	More sudden gains	Medium
Lindegaard et al. (2020)	Treatment Preference Strength	0.57	0.1	Small	-	Medium
Schønning & Nordgreen (2021)	Self-efficacy (PD)	<0.00*	-	-	-	Low
Schønning & Nordgreen (2021)	Self-efficacy (SAD)	<0.00*	-	-	-	Low

**Significance p<0.05; PD = Panic Disorder; SAD = Social Anxiety Disorder*

Higher levels of working alliance were found to be a significant predictor of better outcomes in two studies (Lindegaard et al., 2020; Nordgren et al., 2013) but non-significant in three studies (Berger et al., 2013; Hadjistavropoulos et al., 2016; Seeley et al., 2017). Perceived treatment credibility also had mixed results as a potential predictor as El Alaoui et al. (2015a) found it to be significant, whereas Andersson et al. (2008) found it to be non-significant. However, having a higher working alliance and higher perceived credibility was consistently associated with better outcomes with small to large effect sizes.

Having higher levels of self-efficacy was a significant predictor of outcome in participants with panic disorder and social anxiety disorder (Schønning & Nordgreen, 2021). The presence of sudden gains was also a significant predictor (Shalom et al., 2020) as was SPIN scores after engaging in cognitive restructuring (Dryman et al., 2017).

The strength of treatment preference and change in PHQ-9 scores during treatment were both non-significant (Lindegaard et al., 2020; Seeley et al., 2017).

Experience

Two studies examined whether the experience the practitioner had of delivering the treatment was associated with outcome (El Alaoui et al., 2015; Hedman et al., 2013). Both were significantly associated ($p<0.001$), suggesting those with more experience had better outcomes. Both studies have low risk of bias, however the effect size in El Alaoui et al. (2015) is small.

Other Treatment

One study examined previous treatment and medication status as a potential predictor of outcome (Berger et al., 2013). However, both were reported as non-significant and without effect sizes. The study is rated as having a medium risk of bias.

Verbal Fluency

One study examined verbal fluency as a potential predictor of outcome (Andersson et al., 2008). However, this was non-significant and the direction of effect indicated that worse verbal fluency had better outcomes. The study showed medium levels of risk of bias.

Discussion

Main Findings

This is the first systematic review of the literature on predictors of outcome after CBT-GSH for anxiety disorders. A total of 24 studies examined 116 variables across 10 domains (demographics, baseline severity, comorbidity, within session engagement, between session engagement, engagement with the practitioner, process, experience delivering the treatment, other treatment and verbal fluency). Just under half of the examined variables were significant predictors of outcome (48%), however, most of these were inconclusive across examined studies. There are also limited categories of predictors with consensus in the direction of the predictor (i.e., whether a high vs low score predicted better outcomes). However, risk of bias between studies may have impacted this. Two categories of examined treatment-related variables consistently significantly predicted outcome, between session engagement and experience in delivering the treatment. Although some results were non-significant, there was a consistent pattern in the direction of effect of within session engagement (more engaged participants were more likely to have better outcomes) and process variables

(participants with a higher working alliance and higher perceived treatment credibility were more likely to have better outcomes). Due to the non-significance of the result, but consistency in direction of effects these variables could be considered as partial predictors of outcome. No reliable baseline predictors of outcome were identified. Key findings are discussed alongside limitations and implications.

Inconsistencies and Risk of Bias

Numerous significant predictors were identified across the studies; however, there was a lack of consensus between studies. This was particularly the case for demographics and comorbidity. It is important to consider the risk of bias within the studies that have contradicting significance of predictors. Within demographic predictors, most of the significant variables are from a study with medium risk of bias (Chen et al., 2020). The authors do not account for confounding variables within this study and so, this could have impacted the results causing biased or confounded estimates of significance. The effect sizes of the significant variables within this category are also small and thus, we cannot conclude any demographic variable reliably predicts outcome after CBT-GSH for anxiety disorders. Previous research has shown that demographic variables predict treatment adherence to CBT-GSH (Batterham et al., 2008) rather than outcome.

Within comorbidity variables, all studies were rated as having a medium or medium to high risk of bias. Personality Disorder diagnosis was a significant predictor across two studies (Andersson et al., 2008; Haug et al., 2015) but in different directions (i.e., one stated it was advantageous to outcome and the other not). Thus, this review cannot conclude that having a comorbidity predicts better or worse outcome. Overall, this

tends to fit with existing literature, that comorbidity does not impede treatment response (Allen et al., 2010).

It is less clear why there is contradicting impact between process variables and within session engagement variables. However, both are related to the treatment and the practitioner. Clearly, there is a large variance in sample size, number of treatment sessions, amount of input from the practitioner and the practitioner's qualification as discussed earlier. This may explain some of the variance between the studies. Previous research has shown that process factors only have a moderate impact on CBT outcomes (Keijsers et al., 2000) which may also explain some of the variance. However, due to the consistency and the effect sizes in the associations between outcome and within session engagement and process variables, it is suggested that these are considered as partial predictors of outcome with the above reasons taken into account as to why levels of significance differed. It is suggested that these variables are investigated further in future research to determine their impact on outcomes.

Apart from one paper (Mathiason et al., 2018), every study that investigated time receiving CBT-GSH as a predictor of outcome had a significant result with small to medium effect sizes. This suggests that it is likely that the more time participants spent within CBT-GSH, the better the outcomes which fits with the general literature on psychotherapy outcomes (Cahill et al., 2003). It is therefore valuable to consider who is most likely to dropout and provide extra support to keep patients engaged in treatment. For example, Edmonds et al. (2018) found that younger participants were more likely to drop out of CBT-GSH and so this population may need more adaptations to keep them engaged with the treatment, improving outcomes.

Variables concerning engagement with the practitioner were consistently non-significant, apart from one study (Dryman et al., 2017) that had one of the highest risk of bias ratings within the category. It can be concluded that this variable is not likely to predict outcome. This highlights that the competency of the LI practitioner is potentially more important than the amount of contact in this brief psychological intervention. This may explain the relationship identified in previous research between practitioner involvement in self-help and better outcomes (Furmark et al., 2009; Pleva & Wade, 2007). Verbal fluency and other treatments were also not significant, however were only investigated in one study each.

Baseline severity was also a mixed category of significant predictors. Baseline anxiety was a significant predictor in all studies that investigated this variable, however, the direction varied in that some studies suggested higher baseline anxiety predicted better outcomes (González-Robles et al., 2021; Mathiason et al., 2018) and one study suggested mild to moderate anxiety predicted better outcomes (Lawn et al., 2019). The same pattern occurred for baseline depression. Thus, this review cannot conclude whether mild or severe baseline anxiety/depression predicts outcome. This may be mediated by the amount of contact and the competency of the practitioner as previous research suggests that mild/moderate symptoms predict better outcomes in unguided CBT and moderate/severe symptoms predict better outcomes in CBT-GSH (Karyotaki et al., 2021). To support this theory, both González-Robles et al. (2021) and Mathiason et al. (2018) provided weekly telephone calls and Mathiason et al. (2018) used a qualified psychologist as the practitioner (González-Robles et al. (2021) did not report the practitioner qualifications). Whereas Lawn et al. (2019) does not make it clear how much the practitioner was involved within the GSH, but only one was a mental health professional, the others being health and community workers. This suggests that the

mild/moderate predictor of outcome may have had more of an unguided treatment, considering the practitioner's qualifications. Whilst the moderate/severe predictors may have had more intense support from the qualified psychologists. Additionally, guidelines recommend CBT-GSH for patients with mild/moderate symptoms (Andersson et al., 2019) and so it is questioned whether participants with moderate/severe symptoms should have accessed CBT-GSH within these studies.

Significant Predictors of Outcome

Two categories represented replicated significant predictors of outcome – between session engagement ('homework') and experience delivering the GSH. Due to missing reported information, it was impossible to assess how large the effect was of these predictor variables and the risk of bias within these studies was low and medium. It is therefore strongly suggested that the more engaged participants are in completing between session tasks, the better their outcomes and this fits with existing literature (Glenn et al., 2013). Therefore, the design of homework tasks is important during GSH as poorly understood tasks or tasks that are too difficult will not be completed. Similarly, it is suggested that the more experience the practitioner has in delivering CBT-GSH, the better the participant's outcomes. There is little evidence in the existing literature that investigates a practitioner's experience and treatment outcome. Published research tends to contradict this finding in that practitioner's experience does not impact outcomes after LI CBT (Andersson et al., 2012; Branson et al., 2018; Norton et al., 2014). It could be suggested that the longer the practitioner has been delivering CBT-GSH, the more likely they are able to competently adapt the psychoeducation to meet the needs of the individual patient. This was evidenced in Haug et al. (2016), that higher therapist competence was associated with better

outcomes. Therefore, experienced practitioners may likely flex the psychoeducational materials to meet patient need and may have well-honed interpersonal skills.

Strengths and Limitations

The review had several strengths including pre-registration of study protocol; searches across multiple datasets and reference lists and a breadth of evidence collected. The initial search terms were kept broad, and the researcher remained inclusive throughout the screening process. This has resulted in a wide range of studies from different countries with many predictors assessed. A further strength is the use of the second rater through the screening process and the risk of bias assessment. This has provided the review with inter-rater reliability (McHugh, 2012) and increases the reliability of the results. This could have been improved by the second rater screening all the titles/abstracts and full texts, rather than a percentage.

Although a strength of the review is the broad range of studies, it is acknowledged that this may also have introduced an element of bias from the extent of variability. As shown in Table 1, within the 24 included studies, there are 15 different outcome measures used to assess anxiety disorders and treatment duration ranges from 1 session/module to 12 sessions/modules, while typical GSH lasts 6-8 sessions (Fairburn, 2013). Practitioner qualifications also varied as some included studies were before the advent of the LI practitioner i.e., PWP. Therefore, it is questionable whether the included studies can be directly compared when the treatments and outcomes differ so greatly. However, all the treatments were developed from CBT principles and had some similarities in the way they were delivered (i.e., 92% of the studies delivered CBT through online lessons and participants were then supported by a LI coach).

The review was restricted to peer reviewed publications; thus, no grey literature was included which means that studies not yet published may have been missed (Pappas & Williams, 2011). This opens the review up to possible publication bias as significant findings are more likely to be published than non-significant (Franco et al., 2014). Additionally, some full-text papers were unattainable and therefore, some relevant data may also have been missed.

A significant limitation to the review is the quantity of missing information from the studies included. This limited the review, as a meta-analysis was then not feasible. Variables that were examined across studies lacked relevant information for the meta-analysis to be conducted. A meta-analysis can help increase validity and confidence in findings (Valentine et al., 2010). The lack of information also made it difficult to compare effect sizes across studies and impacted individual studies' risk of bias ratings, with 42% of studies categorised as medium/medium-high risk of bias. Due to this, the results of the review need to be considered with caution.

To assess risk of bias, CASP checklists were chosen due to their usability and appropriateness for methods included in this review (RCT & cohort studies). However, not all questions within the CASP checklists were relevant to the studies included as the checklists assess evidence and they do not allow for in-depth assessment of approaches to predictor analyses (i.e., appropriateness of statistical methods and the impact of these on risk of bias). Alternatively, the Cochrane risk of bias tool for randomised trials (Higgins et al., 2022) could have been considered as a more detailed quality assessment.

Clinical and Research Implications

The main aim of the present review was to understand what variables predict outcome after CBT-GSH. Identification of variables able to be measured at baseline would enable an initial step toward applying precision medicine to GSH interventions. If it is known what pre-treatment variables predict outcome, then it is possible to match patients to the most effective treatment for them. Unfortunately, none of the patient pre-treatment variables were consistent predictors of outcome (demographics, comorbidity, baseline severity, verbal fluency, other treatments). Thus, from this review, we cannot determine the most suitable candidate for a CBT-GSH intervention. Further research is needed examining multivariate models of patient variables that predict outcome after CBT-GSH, to add to the evidence base to reach a more conclusive answer to this question to support the improvement of outcomes. If in the future, we learn more about who responds to CBT-GSH and who does not, increasing options for LI psychological treatments should be considered within the NHS.

A secondary aim of the review was to consider the implications of knowing in-treatment predictors and whether the treatment offered needs to be adapted. A key finding of this review is that between session engagement consistently predicts outcomes of CBT-GSH. This finding fits with previous research, that patient engagement with homework assignments is a significant predictor of symptom improvement (Conklin & Strunk, 2015). This is important to help understand how to improve outcomes after CBT-GSH. There is limited published research focussing on improving between session engagement, however Conklin et al. (2018) suggest that therapist behaviours predict homework engagement. They found that the practitioner's emphasis on key elements of homework enhances engagement. This may suggest a link between the two significant predictors as practitioners with more experience may have a better understanding of where to place the emphasis on between session tasks. Future

research should focus on methods of increasing between session engagement to ensure patients benefit as much as possible from CBT-GSH. The review highlights the importance for clinicians to encourage between session engagement and ongoing training to support them to do this.

Conclusion

This review highlights that the evidence base of predictors of outcome following CBT-GSH for anxiety disorders is currently underdeveloped and therefore the overall conclusion is unclear. Although 48% of examined variables were significant predictors of outcome, there is a lack of consensus as to those variables that consistently predicted outcome. Between session engagement and practitioner experience emerged as the only consistent predictors of outcome. These variables may be related with more experienced practitioners having more skills in engaging patients in between session tasks. This provides a focus for future research, to consider methods to improve between session engagement and understand the mechanisms behind experienced practitioners having better outcomes. It may also be interesting to understand which patients engage with in between sessions tasks to target extra support to those who are struggling to, consequently, improving outcomes.

Due to moderate levels of risk of bias within the included studies and a lack of reported data, caution should be taken when interpreting the results. It should not be ruled out that some of the categories of examined variables are strong predictors of outcome.

References

Allen, L. B., White, K. S., Barlow, D. H., Shear, M. K., Gorman, J. M., & Woods, S. W. (2010). Cognitive-Behavior Therapy (CBT) for Panic Disorder: Relationship of Anxiety and Depression Comorbidity with Treatment Outcome. *Journal of Psychopathology and Behavioral Assessment*, *32*(2), 185–192. https://doi.org/10.1007/s10862-009-9151-3

Amin, R., Iqbal, A., Naeem, F., & Irfan, M. (2020). Effectiveness of a culturally adapted cognitive behavioural therapy-based guided self-help (CACBT-GSH) intervention to reduce social anxiety and enhance self-esteem in adolescents: a randomized controlled trial from Pakistan. *Behavioural and Cognitive Psychotherapy*, *48*(5), 503–514. https://doi.org/10.1017/S1352465820000284

Andersson, G., Carlbring, P., & Furmark, T. (2012). Therapist experience and knowledge acquisition in internet-delivered CBT for social anxiety disorder: A randomized controlled trial. *PloS One*, *7*(5), e37411–e37411. https://doi.org/10.1371/journal.pone.0037411

Andersson, G., Carlbring, P., & Grimlund, A. (2008). Predicting treatment outcome in internet versus face to face treatment of panic disorder. *Computers in Human Behavior*, *24*(5), 1790–1801. https://doi.org/10.1016/j.chb.2008.02.003

Andersson, G., Paxling, B., Roch-Norlund, P., Östman, G., Norgren, A., Almlöv, J., Georén, L., Breitholtz, E., Dahlin, M., Cuijpers, P., Carlbring, P., & Silverberg, F. (2012). Internet-Based Psychodynamic versus Cognitive Behavioral Guided Self-Help for Generalized Anxiety Disorder. *Psychotherapy and Psychosomatics*, *81*(6), 344–355. https://doi.org/10.1159/000339371

Andersson, G., Titov, N., Dear, B. F., Rozental, A., & Carlbring, P. (2019). Internet-delivered psychological treatments: From innovation to implementation. *World Psychiatry, 18*(1), 20–28. https://doi.org/10.1002/wps.20610

Baigent, M., Smith, D., Battersby, M., Lawn, S., Redpath, P., & McCoy, A. (2023). The Australian version of IAPT: clinical outcomes of the multi-site cohort study of NewAccess. *Journal of Mental Health*, *32*(1), 341–350. https://doi.org/10.1080/09638237.2020.1760224

Batterham, P. J., Neil, A. L., Bennett, K., Griffiths, K. M., & Christensen, H. (2008). Predictors of adherence among community users of a cognitive behavior therapy website. *Patient Preference and Adherence*, *2*, 97–105.

Beard, J. I., & Delgadillo, 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

Berger, T., Boettcher, J., & Caspar, F. (2013). Internet-Based Guided Self-Help for Several Anxiety Disorders: A Randomized Controlled Trial Comparing a Tailored With a Standardized Disorder-Specific Approach. *Psychotherapy, 51*(2), 207–219. https://doi.org/10.1037/a0032527

Brandenburg, A. B. (2017). *Cognitive behavioral therapy for anxiety disorders: predictors of outcome* (Master's thesis, University of Twente).

Branson, A., Myles, P., Mahdi, M., & Shafran, R. (2018). The Relationship between Competence and Patient Outcome with Low-Intensity Cognitive Behavioural Interventions. *Behavioural and Cognitive Psychotherapy*, *46*(1), 101–114. https://doi.org/10.1017/S1352465817000522

Butler, R. M., O'Day, E. B., Swee, M. B., Horenstein, A., & Heimberg, R. G. (2021). Cognitive behavioral therapy for social anxiety disorder: Predictors of treatment outcome in a quasi-naturalistic setting. *Behavior therapy, 52*(2), 465-477. https://doi.org/10.1016/j.beth.2020.06.002

Cahill, J., Barkham, M., Hardy, G., Rees, A., Shapiro, D. A., Stiles, W. B., & Macaskill, N. (2003). Outcomes of patients completing and not completing cognitive therapy for depression. *The British journal of clinical psychology, 42*(Pt 2), 133–143. https://doi-org.sheffield.idm.oclc.org/10.1348/014466503321903553

Centre for Reviews and Dissemination (2009). *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare*. University of York

Chen, H., Rodriguez, M. A., Qian, M., Kishimoto, T., Lin, M., & Berger, T. (2020). Predictors of treatment outcomes and adherence in internet-based cognitive behavioral therapy for social anxiety in China. *Behavioural and Cognitive Psychotherapy*, *48*(3), 291–303. https://doi.org/10.1017/S1352465819000730

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

Conklin, L. R., & Strunk, D. R. (2015). A session-to-session examination of homework engagement in cognitive therapy for depression: Do patients experience immediate benefits? *Behaviour Research and Therapy*, *72*, 56–62. https://doi.org/10.1016/j.brat.2015.06.011

Conklin, L. R., Strunk, D. R., & Cooper, A. A. (2018). Therapist behaviors as predictors of immediate homework engagement in cognitive therapy for depression. *Cognitive Therapy and Research*, *42*(1), 16–23. https://doi.org/10.1007/s10608-017-9873-6

Connor, K. M., Davidson, J. R. T., Churchill, L. E., Sherwood, A., Weisler, R. H., & FOA, E. (2000). Psychometric properties of the Social Phobia Inventory (SPIN): New self-rating scale. *British Journal of Psychiatry*, *176*(4), 379–386. https://doi.org/10.1192/bjp.176.4.379

Coull, G., & Morris, P. G. (2011). The clinical effectiveness of CBT-based guided self-help interventions for anxiety and depressive disorders: a systematic review. *Psychological Medicine*, *41*(11), 2239–2252. https://doi.org/10.1017/S0033291711000900

Critical Appraisal Skills Programme. (CASP). (2022). *CASP checklist*. https://casp-uk.net/casp-tools-checklists/

Cuijpers, P., Donker, T., van Straten, A., Li, J., & Andersson, G. (2010). Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychological medicine,* *40*(12), 1943-1957. https://doi.org/10.1017/S0033291710000772

Delgadillo, J., McMillan, D., Lucock, M., Leach, C., Ali, S., & Gilbody, S. (2014). Early changes, attrition, and dose-response in low intensity psychological interventions. *British Journal of Clinical Psychology*, *53*(1), 114–130. https://doi.org/10.1111/bjc.12031

Delgadillo, J., Rhodes, L., Moreea, O., McMillan, D., Gilbody, S., Leach, C., Lucock, M., Lutz, W., & Ali, S. (2018). Relapse and Recurrence of Common Mental Health Problems after Low Intensity Cognitive Behavioural Therapy: The WYLOW Longitudinal Cohort Study. *Psychotherapy and Psychosomatics*. *87*(2), 116-117. https://doi.org/10.1159/000485386

Dryman, M. T., McTeague, L. M., Olino, T. M., & Heimberg, R. G. (2017). Evaluation of an Open-Access CBT-Based Internet Program for Social Anxiety: Patterns of Use, Retention, and Outcomes. *Journal of Consulting and Clinical Psychology*, *85*(10), 988–999. https://doi.org/10.1037/ccp0000232

Edmonds, M., Hadjistavropoulos, H. D., Schneider, L. H., Dear, B. F., & Titov, N. (2018). Who benefits most from therapist-assisted internet-delivered cognitive behaviour therapy in clinical practice? Predictors of symptom change and dropout. *Journal of Anxiety Disorders*, *54*, 24–32. https://doi.org/10.1016/j.janxdis.2018.01.003

El Alaoui, S., Hedman, E., Kaldo, V., Hesser, H., Kraepelien, M., Andersson, E., Rück, C., Andersson, G., Ljótsson, B., & Lindefors, N. (2015). Effectiveness of Internet-Based Cognitive-Behavior Therapy for Social Anxiety Disorder in Clinical Psychiatry. *Journal of Consulting and Clinical Psychology*, *83*(5), 902–914. https://doi.org/10.1037/a0039198

El Alaoui, S. E., Ljótsson, B., Hedman, E., Kaldo, V., Andersson, E., Rück, C., Andersson, G., & Lindefors, N. (2015a). Predictors of symptomatic change and adherence in internet-based cognitive behaviour therapy for social anxiety disorder in routine psychiatric care. *PloS One*, *10*(4), e0124258–e0124258. https://doi.org/10.1371/journal.pone.0124258

Eriksen, M. B., & Frandsen, T. F. (2018). The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: a systematic review. *Journal of the Medical Library Association, 106*(4), 420–431. https://doi.org/10.5195/jmla.2018.345

Fairburn, C. G. (2013). *Overcoming binge eating (2nd ed.)*. Guilford Press.

Fauskanger Bjaastad, J., Henningsen Wergeland, G. J., Mowatt Haugland, B. S., Gjestad, R., Havik, O. E., Heiervang, E. R., & Öst, L. G. (2018). Do clinical experience, formal cognitive behavioural therapy training, adherence, and competence predict outcome in cognitive behavioural therapy for anxiety disorders in youth?. *Clinical psychology & psychotherapy, 25*(6), 865-877. https://doi.org/10.1002/cpp.2321

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: unlocking the file drawer. *Science*, *345*(6203), 1502–1505. https://doi.org/10.1126/ science.1255484

Furmark, T., Carlbring, P., Hedman, E., Sonnenstein, A., Clevberger, P., Bohman, B., Eriksson, A., Hållén, A., Frykman, M., Holmström, A., Sparthan, E., Tillfors, M., Ihrfelt, E. N., Spak, M., Eriksson, A., Ekselius, L., & Andersson, G. (2009). Guided and unguided self-help for social anxiety disorder: randomised controlled trial. *British Journal of Psychiatry*, *195*(5), 440–447. https://doi.org/10.1192/bjp.bp.108.060996

Gaudiano, B. A. (2008). Cognitive-behavioural therapies: achievements and challenges. *BMJ Ment Health*, *11*(1), 5-7. https://doi.org/10.1136/ebmh.11.1.5

Gega, L., Marks, I., & Mataix-Cols, D. (2004). Computer-aided CBT self-help for anxiety and depressive disorders: Experience of a London clinic and future

directions. *Journal of clinical psychology*, *60*(2), 147-157. https://doi.org/10.1002/jclp.10241

Gellatly, J., Chisnall, L., Seccombe, N., Ragan, K., Lidbetter, N., & Cavanagh, K. (2018). @Home eTherapy Service for People with Common Mental Health Problems: an Evaluation. *Behavioural and Cognitive Psychotherapy*, *46*(1), 115–120. https://doi.org/10.1017/S1352465817000297

Glenn, D., Golinelli, D., Rose, R. D., Roy-Byrne, P., Stein, M. B., Sullivan, G., Bystritksy, A., Sherbourne, C., & Craske, M. G. (2013). Who Gets the Most Out of Cognitive Behavioral Therapy for Anxiety Disorders? The Role of Treatment Dose and Patient Engagement. *Journal of Consulting and Clinical Psychology*, *81*(4), 639–649. https://doi.org/10.1037/a0033403

González-Robles, A., Suso-Ribera, C., Díaz-García, A., García-Palacios, A., Castilla, D., & Botella, C. (2021). Predicting response to transdiagnostic iCBT for emotional disorders from patient and therapist involvement. *Internet Interventions*, *25*, 100420–100420. https://doi.org/10.1016/j.invent.2021.100420

Haby, M. M., Donnelly, M., Corry, J., & Vos, T. (2006). Cognitive behavioural therapy for depression, panic disorder and generalized anxiety disorder: a meta-regression of factors that may predict outcome. *Australian & New Zealand Journal of Psychiatry, 40*(1), 9-19. https://doi.org/10.1080/j.1440-1614.2006.01736

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency

and Open Synthesis. *Campbell Systematic Reviews*, *18*, e1230. https://doi.org/10.1002/cl2.1230

Hadjistavropoulos, H. D., Pugh, N. E., Hesser, H., & Andersson, G. (2016). Therapeutic Alliance in Internet-Delivered Cognitive Behaviour Therapy for Depression or Generalized Anxiety. *Clinical Psychology and Psychotherapy*, *24*(2), 451–461. https://doi.org/10.1002/cpp.2014

Haug, T., Nordgreen, T., Öst, L.-G., Kvale, G., Tangen, T., Andersson, G., Carlbring, P., Heiervang, E. R., & Havik, O. E. (2015). Stepped care versus face-to–face cognitive behavior therapy for panic disorder and social anxiety disorder: Predictors and moderators of outcome. *Behaviour Research and Therapy*, *71*, 76–89. https://doi.org/10.1016/j.brat.2015.06.002

Haug, T., Nordgreen, T., Öst, L.-G., Tangen, T., Kvale, G., Hovland, O. J., Heiervang, E. R., & Havik, O. E. (2016). Working alliance and competence as predictors of outcome in cognitive behavioral therapy for social anxiety and panic disorder in adults. *Behaviour Research and Therapy*, *77*, 40–51. https://doi.org/10.1016/j.brat.2015.12.004

Hedman, E., Andersson, E., Ljótsson, B., Andersson, G., Andersson, E., Schalling, M., Lidefors, C., & Rück, C. (2012). Clinical and genetic outcome determinants of internet-and group-based cognitive behavior therapy for social anxiety disorder. *Acta Psychiatrica Scandinavica, 126*(2), 126-136. https://doi.org/10.1111/j.1600-0447.2012.01834.x

Hedman, E., Ljótsson, B., Rück, C., Bergström, J., Andersson, G., Kaldo, V., Jansson, L., Andersson, E., Blom, K., El Alaoui, S., Falk, L., Ivarsson, J., Nasri, B., Rydh, S., & Lindefors, N. (2013). Effectiveness of Internet-based cognitive behaviour

therapy for panic disorder in routine psychiatric care. *Acta Psychiatrica Scandinavica*, *128*(6), 457–467. https://doi.org/10.1111/acps.12079

Higgins, J. P., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions (vol. 4*). John Wiley and Sons.

Higgins, J., Savović, J., Page, M., Elbers, R., Sterne, J. (2022). Assessing risk of bias in a randomized trial. In Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (Eds). *Cochrane handbook for systematic reviews of interventions (Version 6.3).* Cochrane. www.training.cochrane.org/handbook.

Kampman, M., Keijsers, G. P., Hoogduin, C. A., & Hendriks, G. J. (2008). Outcome prediction of cognitive behaviour therapy for panic disorder: Initial symptom severity is predictive for treatment outcome, comorbid anxiety or depressive disorder, cluster C personality disorders and initial motivation are not. *Behavioural and Cognitive Psychotherapy, 36*(1), 99-112. https://doi.org/10.1017/S1352465807004018

Karyotaki, E., Efthimiou, O., Miguel, C., Bermpohl, F. M. genannt, Furukawa, T. A., Cuijpers, P., Riper, H., Patel, V., Mira, A., Gemmil, A. W., Yeung, A. S., Lange, A., Williams, A. D., Mackinnon, A., Geraedts, A., van Straten, A., Meyer, B., Björkelund, C., Knaevelsrud, C., & Mohr, D. C. (2021). Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA Psychiatry, 78*(4), 361–371. https://doi.org/10.1001/jamapsychiatry.2020.4364

Keijsers, G. P. J., Schaap, C. P. D. R., & Hoogduin, C. A. L. (2000). the impact of interpersonal patient and therapist behavior on outcome in Cognitive-Behavior Therapy. *Behavior Modification*, *24*(2), 264–297. https://doi.org/10.1177/0145445500242006

König, I. R., Fuchs, O., Hansen, G., von Mutius, E., & Kopp, M. V. (2017). What is precision medicine? *European respiratory journal*, *50*(4). https://doi.org/10.1183/13993003.00391-2017

Kosorok, M. R., & Laber, E. B. (2019). Precision Medicine. *Annual Review of Statistics and Its Application*, *6*(1), 263–286. https://doi.org/10.1146/annurev-statistics-030718-105251

Kunas, S. L., Lautenbacher, L. M., Lueken, U., & Hilbert, K. (2021). Psychological predictors of cognitive-behavioral therapy outcomes for anxiety and depressive disorders in children and adolescents: a systematic review and meta-analysis. *Journal of Affective Disorders, 278*, 614-626. https://doi.org/10.1016/j.jad.2020.09.092

Lawn, S., Huang, N., Zabeen, S., Smith, D., Battersby, M., Redpath, P., Glover, F., Venning, A., Cameron, J., & Fairweather-Schmidt, K. (2019). Outcomes of telephone-delivered low-intensity cognitive behaviour therapy (LiCBT) to community dwelling Australians with a recent hospital admission due to depression or anxiety: MindStep. *BMC Psychiatry*, *19*(1), 2–2. https://doi.org/10.1186/s12888-018-1987-1

Lenhard, W. & Lenhard, A. (2016). Computation of effect sizes. Retrieved from: https://www.psychometrica.de/effect_size.html. *Psychometrica*. https://doi.org/10.13140/RG.2.2.17823.92329

Lindegaard, T., Hesslow, T., Nilsson, M., Johansson, R., Carlbring, P., Lilliengren, P., & Andersson, G. (2020). Internet-based psychodynamic therapy vs cognitive behavioural therapy for social anxiety disorder: A preference study. *Internet Interventions: The Application of Information Technology in Mental and*

Behavioural Health, *20*, 100316–100316. https://doi.org/10.1016/j.invent.2020.100316

Mathiasen, K., Riper, H., Andersen, T. E., & Roessler, K. K. (2018). Guided internet-based cognitive behavioral therapy for adult depression and anxiety in routine secondary care: Observational study. *Journal of Medical Internet Research*, *20*(11), e10927–e10927. https://doi.org/10.2196/10927

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour research and therapy*, *36*(4), 455-470. https://doi.org/10.1016/S0005-7967(97)10031-6

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282. https://doi.org/10.11613/bm.2012.031

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7). https://doi.org/10.1371/journal.pmed.1000097

Mululo, S. C. C., Menezes, G. B. D., Vigne, P., & Fontenelle, L. F. (2012). A review on predictors of treatment outcome in social anxiety disorder. Brazilian *Journal of Psychiatry, 34*, 92-100. https://doi.org/10.1590/S1516-44462012000100016

National Institute for Health and Clinical Excellence (NICE). (2011, May 25). *Common mental health disorders identification and pathways to care.* https://www.nice.org.uk/guidance/cg123

Nordgren, L. B., Carlbring, P., Linna, E., & Andersson, G. (2013). Role of the working alliance on treatment outcome in tailored internet-based cognitive behavioural therapy for anxiety disorders: Randomized controlled pilot trial. *JMIR Research Protocols*, *2*(1), e4–e4. https://doi.org/10.2196/resprot.2292

Nordmo, M., Sinding, A. I., Carlbring, P., Andersson, G., Havik, O. E., & Nordgreen, T. (2015). Internet-delivered cognitive behavioural therapy with and without an initial face-to-face psychoeducation session for social anxiety disorder: A pilot randomized controlled trial. *Internet Interventions: the Application of Information Technology in Mental and Behavioural Health*, *2*(4), 429–436. https://doi.org/10.1016/j.invent.2015.10.003

Norton, P. J., Little, T. E., & Wetterneck, C. T. (2014). Does experience matter? Trainee experience and outcomes during transdiagnostic cognitive-behavioral group therapy for anxiety. *Cognitive Behaviour Therapy*, *43*(3), 230–238. https://doi.org/10.1080/16506073.2014.919014

Newman, M. G., Jacobson, N. C., Rackoff, G. N., Bell, M. J., & Taylor, C. B. (2021). A randomized controlled trial of a smartphone-based application for the treatment of anxiety. *Psychotherapy Research*, *31*(4), 443–454. https://doi.org/10.1080/10503307.2020.1790688

Newman, M. G., Kanuri, N., Rackoff, G. N., Jacobson, N. C., Bell, M. J., & Taylor, C. B. (2021a). A randomized controlled feasibility trial of internet-delivered guided self-help for generalized anxiety disorder (GAD) among university students in India. *Psychotherapy*, *58*(4), 591–601. https://doi.org/10.1037/pst0000383

Nicholls, K. (2023, August 2). *Key statistics about mental health treatment and services.* Counselling Directory. https://www.counselling-directory.org.uk/talk-therapy-stats.html#accessingmentalhealthservices

ONS. (2024, May 9). *Quarterly personal well-being estimates – seasonally adjusted*. https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/datasets/quarterlypersonalwellbeingestimatesseasonallyadjusted

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & McGuinness, L. A. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *PLOS Medicine*, *18*(3), e1003583. https://doi.org/10.1371/journal.pmed.1003583

Pappas, C., & Williams, I. (2011). Grey literature: its emerging importance. *Journal of Hospital Librarianship*, *11*(3), 228–234. https://doi.org/10.1080/15323269.2011.587100

Pleva, J., & Wade, T. D. (2007). Guided self-help versus pure self-help for perfectionism: A randomised controlled trial. *Behaviour Research and Therapy*, *45*(5), 849–861. https://doi.org/10.1016/j.brat.2006.08.009

Powell, C. L. Y. M., Chiu, C. Y., Sun, X., & So, S. H. (2024). A meta-analysis on the efficacy of low-intensity cognitive behavioural therapy for generalised anxiety disorder. *BMC Psychiatry*, *24*(1), 10–10. https://doi.org/10.1186/s12888-023-05306-6

Rea, L. M., & Parker, R. A. (1992). *Designing and conducting survey research: a comprehensive guide*. Jossey-Bass Publishers.

Schat, A., Van Noorden, M. S., Noom, M. J., Giltay, E. J., Van der Wee, N. J. A., Vermeiren, R. R. J. M., & Zitman, F. G. (2013). Predictors of outcome in outpatients with anxiety disorders: the Leiden routine outcome monitoring study. *Journal of psychiatric research, 47*(12), 1876-1885. https://doi.org/10.1016/j.jpsychires.2013.09.005

Schønning, A., & Nordgreen, T. (2021). Predicting treatment outcomes in guided internet-delivered therapy for anxiety disorders—the role of treatment self-efficacy. *Frontiers in Psychology*, *12*, 712421–712421. https://doi.org/10.3389/fpsyg.2021.712421

Schulz, A., Stolz, T., Vincent, A., Krieger, T., Andersson, G., & Berger, T. (2016). A sorrow shared is a sorrow halved? A three-arm randomized controlled trial comparing internet-based clinician-guided individual versus group treatment for social anxiety disorder. *Behaviour Research and Therapy*, *84*, 14–26. https://doi.org/10.1016/j.brat.2016.07.001

Seeley, J. R., Manitsas, T., & Gau, J. M. (2017). Feasibility study of a peer-facilitated low intensity cognitive-behavioral intervention for mild to moderate depression and anxiety in older adults. *Aging & Mental Health*, *21*(9), 968–974. https://doi.org/10.1080/13607863.2016.1186152

Shafran, R., Myles-Hooton, P., Bennett, S., & Öst, L. G. (2021). The concept and definition of low intensity cognitive behaviour therapy. *Behaviour Research and Therapy*, *138*, 103803. https://doi.org/10.1016/j.brat.2021.103803

Shalom, J. G., Strauss, A. Y., Huppert, J. D., Andersson, G., Agassi, O. D., & Aderka, I. M. (2020). Predicting Sudden Gains Before Treatment Begins: An Examination of Pretreatment Intraindividual Variability in Symptoms. *Journal of Consulting and Clinical Psychology, 88*(9), 809–817. https://doi.org/10.1037/ccp0000587

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159

Stolz, T., Schulz, A., Krieger, T., Vincent, A., Urech, A., Moser, C., Westermann, S., & Berger, T. (2018). A mobile app for social anxiety disorder: a three-arm randomized controlled trial comparing mobile and pc-based guided self-help interventions. *Journal of Consulting and Clinical Psychology*, *86*(6), 493–504. https://doi.org/10.1037/ccp0000301

Turpin, G. (2010). IAPT good practice guide to using self-help materials. *NMHDU/IAPT*, 1-40.

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? *Journal of Educational and Behavioral Statistics*, *35*(2), 215–247. https://doi.org/10.3102/ 1076998609346961

Wergeland, G. J. H., Fjermestad, K. W., Marin, C. E., Bjelland, I., Haugland, B. S. M., Silverman, W. K., Ost, L-G., Fauskanger Bjaastad, J., Oeding, K., Havik, O. E., & Heiervang, E. R. (2016). Predictors of treatment outcome in an effectiveness trial of cognitive behavioral therapy for children with anxiety disorders. *Behaviour research and therapy, 76*, 1-12. https://doi.org/10.1016/j.brat.2015.11.001

Appendices

Appendix	Title	Page
A	PRISMA 2020 Checklist	57
B	Full Search Strategy	59
C	CASP Checklist: Cohort Studies	60
D	CASP Checklist: RCTs	66
E	Summary of Risk of Bias Assessment	69

Appendix A. PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Page 3
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Page 3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 7
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 9
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Page 10-11
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 11
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Page 59
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Page 12
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 13
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Page 13-14
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Page 13-14
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Page 12-13
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Page 13-14
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Page 13-14
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Page 13-14
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Page 13-14
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Page 13-14
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	N/A
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Page 12-13
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Page 12-13

Section and Topic	Item #	Checklist item	Location where item is reported
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Page 15
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	No citations included
Study characteristics	17	Cite each included study and present its characteristics.	Page 16-19
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Page 20-21; 69-73
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Page 21-32
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Page 21-32
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	N/A
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	N/A
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Page 21-36; 69-73
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Page 21-36; 69-73
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Page 32-33; 36-37
	23b	Discuss any limitations of the evidence included in the review.	Page 37-38
	23c	Discuss any limitations of the review processes used.	Page 37-38
	23d	Discuss implications of the results for practice, policy, and future research.	Page 39-40
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Page 9
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Page 9
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Page ii
Competing interests	26	Declare any competing interests of review authors.	Page ii
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Page ii

Appendix B. Search Strategy

PUBMED

(anxiety[MeSH Terms]) AND ("CBT" OR "cognitive behav* therapy") AND ("self help" OR "psychoeducation" OR "low intensity" OR "self care" OR "self management" OR "guided self help" OR "minimal intervention" OR "minimal contact")

PSYCNET

Did not use MeSH terms because: PsycInfo records include MeSH terms if the document has also been indexed by PubMed (this accounts for approximately 35% of PsycInfo records), but MeSH terms are not mapped to APA Thesaurus terms or vice versa. A MeSH term in a PsycInfo record is hyperlinked to PubMed, therefore selecting it triggers a PubMed search.

Any Field: "CBT" *OR* **Any Field**: "cognitive behav* therapy" *AND* **Any Field**: "self help" *OR* **Any Field**: "psychoeducation" *OR* **Any Field**: "low intensity" *OR* **Any Field**: "self management" *OR* **Any Field**: "guided self help" *OR* **Any Field**: "minimal intervention" *OR* **Any Field**: "minimal contact" *AND* **Any Field**: "anxiety" *OR* **Any Field**: "phobia" *OR* **Any Field**: "panic" *NOT* **Any Field**: "depression" *NOT* **Any Field**: "animal" *NOT* **Any Field**: "major depression"

SCOPUS

"CBT" *OR* "cognitive behav* therapy" *AND* "self help" *OR* "psychoeducation" *OR* "low intensity" *OR* "self management" *OR* "guided self help" *OR* "minimal intervention" *OR* "minimal contact" *AND* "anxiety" *OR* "phobia" *OR* "panic" *NOT* "depression" *NOT* "animal" *NOT* "major depression"

Appendix C. CASP Checklist for Cohort Study

Paper for appraisal and reference:.....

Section A: Are the results of the study valid?

1. Did the study address a clearly focused issue?

Yes ☐
Can't Tell ☐
No ☐

HINT: A question can be 'focused' in terms of
- the population studied
- the risk factors studied
- is it clear whether the study tried to detect a beneficial or harmful effect
- the outcomes considered

Comments:

2. Was the cohort recruited in an acceptable way?

Yes ☐
Can't Tell ☐
No ☐

HINT: Look for selection bias which might compromise the generalisability of the findings:
- was the cohort representative of a defined population
- was there something special about the cohort
- was everybody included who should have been

Comments:

Is it worth continuing?

CASP
Critical Appraisal Skills Programme

3. Was the exposure accurately measured to minimise bias?

Yes ☐
Can't Tell ☐
No ☐

HINT: Look for measurement or classification bias:
- did they use subjective or objective measurements
- do the measurements truly reflect what you want them to (have they been validated)
- were all the subjects classified into exposure groups using the same procedure

Comments:

4. Was the outcome accurately measured to minimise bias?

Yes ☐
Can't Tell ☐
No ☐

HINT: Look for measurement or classification bias:
- did they use subjective or objective measurements
- do the measurements truly reflect what you want them to (have they been validated)
- has a reliable system been established for detecting all the cases (for measuring disease occurrence)
- were the measurement methods similar in the different groups
- were the subjects and/or the outcome assessor blinded to exposure (does this matter)

Comments:

CASP
Critical Appraisal Skills Programme

5. (a) Have the authors identified all important confounding factors?

Yes ☐
Can't Tell ☐
No ☐

HINT:
- list the ones you think might be important, and ones the author missed

Comments:

5. (b) Have they taken account of the confounding factors in the design and/or analysis?

Yes ☐
Can't Tell ☐
No ☐

HINT:
- look for restriction in design, and techniques e.g. modelling, stratified-, regression-, or sensitivity analysis to correct, control or adjust for confounding factors

Comments:

6. (a) Was the follow up of subjects complete enough?

Yes ☐
Can't Tell ☐
No ☐

HINT: Consider
- the good or bad effects should have had long enough to reveal themselves
- the persons that are lost to follow-up may have different outcomes than those available for assessment
- in an open or dynamic cohort, was there anything special about the outcome of the people leaving, or the exposure of the people entering the cohort

6. (b) Was the follow up of subjects long enough?

Yes ☐
Can't Tell ☐
No ☐

CASP
Critical Appraisal Skills Programme

Comments:	

Section B: What are the results?

7. What are the results of this study?

HINT: Consider
- what are the bottom line results
- have they reported the rate or the proportion between the exposed/unexposed, the ratio/rate difference
- how strong is the association between exposure and outcome (RR)
- what is the absolute risk reduction (ARR)

Comments:	

8. How precise are the results?

HINT:
- look for the range of the confidence intervals, if given

Comments:	

CASP
Critical Appraisal Skills Programme

9. Do you believe the results?

Yes ☐
Can't Tell ☐
No ☐

HINT: Consider
- big effect is hard to ignore
- can it be due to bias, chance or confounding
- are the design and methods of this study sufficiently flawed to make the results unreliable
- Bradford Hills criteria (e.g. time sequence, dose-response gradient, biological plausibility, consistency)

Comments:

Section C: Will the results help locally?

10. Can the results be applied to the local population?

Yes ☐
Can't Tell ☐
No ☐

HINT: Consider whether
- a cohort study was the appropriate method to answer this question
- the subjects covered in this study could be sufficiently different from your population to cause concern
- your local setting is likely to differ much from that of the study
- you can quantify the local benefits and harms

Comments:

11. Do the results of this study fit with other available evidence?

Yes ☐
Can't Tell ☐
No ☐

Comments:

CASP
Critical Appraisal Skills Programme

12. What are the implications of this study for practice?

Yes ☐
Can't Tell ☐
No ☐

HINT: Consider
- one observational study rarely provides sufficiently robust evidence to recommend changes to clinical practice or within health policy decision making
- for certain questions, observational studies provide the only evidence
- recommendations from observational studies are always stronger when supported by other evidence

Comments:	

Appendix D. CASP Checklist for Randomised Controlled Trials

CASP
Critical Appraisal Skills Programme

Study and citation: .....

Section A: Is the basic study design valid for a randomised controlled trial?			
1. Did the study address a clearly focused research question? *CONSIDER:* *Was the study designed to assess the outcomes of an intervention?* *Is the research question 'focused' in terms of:* • *Population studied* • *Intervention given* • *Comparator chosen* • *Outcomes measured?*	Yes ☐	No ☐	Can't tell ☐
2. Was the assignment of participants to interventions randomised? *CONSIDER:* • *How was randomisation carried out? Was the method appropriate?* • *Was randomisation sufficient to eliminate systematic bias?* • *Was the allocation sequence concealed from investigators and participants?*	Yes ☐	No ☐	Can't tell ☐
3. Were all participants who entered the study accounted for at its conclusion? *CONSIDER:* • *Were losses to follow-up and exclusions after randomisation accounted for?* • *Were participants analysed in the study groups to which they were randomised (intention-to-treat analysis)?* • *Was the study stopped early? If so, what was the reason?*	Yes ☐	No ☐	Can't tell ☐

Section B: Was the study methodologically sound?			
4. • **Were the participants 'blind' to intervention they were given?** • **Were the investigators 'blind' to the intervention they were giving to participants?** • **Were the people assessing/analysing outcome/s 'blinded'?**	Yes ☐ ☐ ☐	No ☐ ☐ ☐	Can't tell ☐ ☐ ☐
5. Were the study groups similar at the start of the randomised controlled trial? *CONSIDER:* • *Were the baseline characteristics of each study group (e.g. age, sex, socio-economic group) clearly set out?* • *Were there any differences between the study groups that could affect the outcome/s?*	Yes ☐	No ☐	Can't tell ☐

CASP
Critical Appraisal Skills Programme

6. Apart from the experimental intervention, did each study group receive the same level of care (that is, were they treated equally)? *CONSIDER:* • *Was there a clearly defined study protocol?* • *If any additional interventions were given (e.g. tests or treatments), were they similar between the study groups?* • *Were the follow-up intervals the same for each study group?*	Yes ☐	No ☐	Can't tell ☐

Section C: What are the results?

7. Were the effects of intervention reported comprehensively? *CONSIDER:* • *Was a power calculation undertaken?* • *What outcomes were measured, and were they clearly specified?* • *How were the results expressed? For binary outcomes, were relative and absolute effects reported?* • *Were the results reported for each outcome in each study group at each follow-up interval?* • *Was there any missing or incomplete data?* • *Was there differential drop-out between the study groups that could affect the results?* • *Were potential sources of bias identified?* • *Which statistical tests were used?* • *Were p values reported?*	Yes ☐	No ☐	Can't tell ☐
8. Was the precision of the estimate of the intervention or treatment effect reported? *CONSIDER:* • *Were confidence intervals (CIs) reported?*	Yes ☐	No ☐	Can't tell ☐
9. Do the benefits of the experimental intervention outweigh the harms and costs? *CONSIDER:* • *What was the size of the intervention or treatment effect?* • *Were harms or unintended effects reported for each study group?* • *Was a cost-effectiveness analysis undertaken? (Cost-effectiveness analysis allows a comparison to be made between different interventions used in the care of the same condition or problem.)*	Yes ☐	No ☐	Can't tell ☐

CASP
Critical Appraisal Skills Programme

Section D: Will the results help locally?			
10. **Can the results be applied to your local population/in your context?** *CONSIDER:* • *Are the study participants similar to the people in your care?* • *Would any differences between your population and the study participants alter the outcomes reported in the study?* • *Are the outcomes important to your population?* • *Are there any outcomes you would have wanted information on that have not been studied or reported?* • *Are there any limitations of the study that would affect your decision?*	Yes ☐	No ☐	Can't tell ☐
11. **Would the experimental intervention provide greater value to the people in your care than any of the existing interventions?** *CONSIDER:* • *What resources are needed to introduce this intervention taking into account time, finances, and skills development or training needs?* • *Are you able to disinvest resources in one or more existing interventions in order to be able to re-invest in the new intervention?*	Yes ☐	No ☐	Can't tell ☐

APPRAISAL SUMMARY: *Record key points from your critical appraisal in this box. What is your conclusion about the paper? Would you use it to change your practice or to recommend changes to care/interventions used by your organisation? Could you judiciously implement this intervention without delay?*

Appendix E. Risk of Bias Tables

Y = Yes

N = No

CT = Can’t Tell

Study	*Focussed Research Question?*	*Appropriate Randomisation?*	*Are all participants accounted for?*	*Participants and Investigators blinded?*	*Are study groups similar?*	*Similar care across groups (except for experimental intervention)?*	*Effects of the intervention reported comprehensively?*	*Precision of treatment effect reported?*	*Benefits greater than cost/harm?*	*Results applicable to local population/your context?*	*Experimental intervention greater value than existing interventions?*	**Risk of bias**
Andersson et al. (2008)	Y	Y	Y	N N CT	Y	Y	N	N	CT	Y	Y	Medium
Berger et al. (2013)	Y	Y	Y	CT N N	Y	Y	Y	N	CT	Y	CT	Medium
Gonzalez-Robles et al. (2021)	Y	Y	Y	N N N	Y	Y	Y	Y	Y	Y	Y	Low - Medium
Haug et al. (2015)	Y	Y	Y	N N N	CT	Y	N	N	CT	Y	Y	Medium – High
Lindegaard et al. (2020)	Y	N	Y	N N N	Y	Y	Y	Y	CT	Y	Y	Medium
Newman et al. (2021)	Y	Y	Y	N N Y	Y	Y	Y	N	Y	Y	Y	Low – Medium
Newman et al. (2021a)	Y	Y	N	N N N	Y	Y	Y	N	CT	N	CT	Medium - High
Nordgren et al. (2013)	Y	Y	Y	N N Y	Y	Y	Y	N	CT	Y	Y	Low - Medium
Nordmo et al. (2015)	Y	Y	Y	N N N	Y	Y	Y	Y	CT	N	CT	Medium

Schulza et al. (2016)	Y	Y	Y	N N Y	Y	Y	Y	Y	Y	Y	Y	Low - Medium
Seeley et al. (2016)	Y	Y	Y	N N N	CT	Y	Y	Y	Y	Y	N	Medium
Shalom et al. (2020)	Y	CT	Y	N N CT	Y	Y	Y	N	CT	Y	Y	Medium
Stolz et al. (2018)	Y	Y	Y	N Y Y	Y	Y	Y	Y	CT	Y	Y	Low - Medium

Study	*Clearly Focussed Research Question?*	*Cases recruited in an acceptable way?*	*Was the exposure accurately measured to minimise bias?*	*Was the outcome accurately measured to minimise bias?*	*Were confounding factors identified?*	*Were confounding factors taken into account?*	*Was the follow-up of subjects complete enough?*	*Was the follow-up long enough?*	*Precision estimate of treatment effect reported?*	*Do you believe the results?*	*Results applicable to local population/our context?*	*Do results fit with the available evidence?*	**Risk of bias**
Baigent et al. (2023)	Y	Y	Y	Y	N	N	N	N	Y	Y	Y	Y	Low - Medium
Chen et al. (2020)	Y	Y	Y	Y	Y	N	N	N	N	CT	Y	Y	Medium
Dryman et al. (2017)	Y	Y	Y	Y	Y	CT	N	N	N	Y	Y	Y	Low - Medium
El Alaoui et al. (2015)	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Low
El Alaoui et al. (2015a)	Y	Y	Y	Y	Y	N	N	N	Y	Y	Y	CT	Low - Medium
Gellately et al. (2018)	Y	Y	Y	Y	N	N	N	N	N	CT	Y	Y	Medium
Hadjistavropoulus et al. (2016)	Y	Y	Y	Y	N	Y	N	N	Y	Y	Y	Y	Low - Medium
Hedman et al. (2013)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Low
Lawn et al. (2019)	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Low

Mathiason et al. (2018)	Y	Y	Y	Y	Y	Y	Y	CT	N	Y	Y	Y	Low - Medium
Schønning & Nordgreen (2021)	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Low

Part Two: Empirical Study

Predicting Optimal Treatment Allocation Using the Personalised Advantage Index for Patients Receiving Cognitive Analytic Guided Self-Help (CAT-GSH) and Cognitive Behavioural Guided Self-Help (CBT-GSH) for Anxiety

Abstract

Background: Cognitive behavioural therapy is delivered as treatment as usual in a guided self-help format (i.e., CBT-GSH) in primary care psychological services. Meadows & Kellett (2017) developed cognitive analytic therapy (CAT-GSH) for anxiety to increase patient choice and this has recently been trialled against CBT-GSH in a National Health Service (NHS) Talking Therapies service. Results suggested comparable effectiveness but showed evidence of outcome variability at an individual level, with approximately half of patients meeting recovery status. Understanding who is likely to respond best to which GSH intervention could support treatment matching to improve patient outcomes.

Methods: The present study used data from the patient preference trial comparing CAT-GSH and CBT-GSH for a subset of N = 209 patients receiving treatment for an anxiety disorder (Kellett et al., 2023). This study aimed to identify prognostic and prescriptive predictors of post-treatment anxiety scores. Separate predictive models were developed for CAT-GSH and CBT-GSH using 18 baseline sociodemographic, clinical and treatment preference variables. Two machine learning variable selection approaches (Elastic Net regularisation and Boruta random forest) were applied and evaluated, with the best fitting model selected for the final model development. Regression models were used to calculate a patient advantage index (PAI) based on predicted outcome for each intervention indicating which GSH each patient was predicted to respond best to. Finally, treatment outcomes were compared for patients who received their PAI-indicated optimal GSH with those who received their PAI-indicated non-optimal GSH.

Results: Variable selection using Elastic Net produced the best fitting model for both interventions, retaining 8 and 5 variables for the CAT-GSH and CBT-GSH models respectively. Two prognostic variables predicted improved outcomes for both interventions (baseline PHQ-9 and BAI scores), and 6 and 3 prescriptive variables predicted CAT-GSH outcome (baseline WSAS score; long-term condition; perinatal status; sexual orientation; previous CAT; previous CBT) and CBT-GSH outcome (indices of multiple deprivation; ethnicity; employment status) respectively. Patient preference status was not associated with outcome during either intervention.

Applying the PAI to the full sample indicated 63% received their model indicated optimal GSH. Patients who received their PAI-indicated optimal GSH experienced significantly higher rates of reliable and clinically significant change (RCSI). At post-treatment, 35.9% of patients who had their optimal GSH met RCSI and 16.6% of patients who did not have their optimal GSH met RCSI and this difference was significant (X^2 (1, N = 209) = 8.82, p = .003). At 24-week follow-up, 36.6% of patients who had their optimal GSH met RCSI and 19.2% of patients who did not have their optimal GSH met RCSI and this was again significant (X^2 (1, N = 209) = 7.04, p = .008).

Within the subgroup of the sample who had the largest indicated PAI value (larger than 1 standard deviation of the sample mean), only 17.7% of the sample had an optimal GSH identified (N = 37), and of these, 70.3% had received their optimal GSH. No patients met RCSI at post-treatment or follow-up if they did not have their optimal GSH. Thirty-one percent of patients that had their optimal GSH met RCSI and this difference was significant (X^2 (1, N = 37) = 4.32, p = .038).

Conclusion: This study indicates that treatment matching algorithms have the potential to improve outcomes and support treatment decision making for low intensity interventions for anxiety. Lack of an external test sample, unbalanced intervention samples and skewed data need to be held in mind when interpreting the results. Future research should continue to investigate the use of the PAI for GSH treatment matching but with larger and more balanced samples.

Keywords: *Anxiety, guided self-help, cognitive analytic therapy, cognitive behavioural therapy, machine learning, patient advantage index, patient preference*

Practitioner Points

- Matching patients to their optimal GSH results in better outcomes when the PAI-predicted benefit is large, therefore, implementing machine learning algorithms when allocating patients to GSH could improve outcomes for subsample of anxious patients accessing NHS Talking Therapies.
- Patient preference did not significantly predict outcome; however, it could be incorporated into clinical practice when an optimal treatment is not identified for any individual.
- Future research is needed with larger, more balanced samples, to develop an algorithm to predict optimal treatments within NHS Talking Therapies.

Introduction

Guided Self-Help for Anxiety

Anxiety is a primary driver of disability worldwide (Lozano et al., 2012). The National Institute of Clinical Excellence (NICE) recommends cognitive behavioural therapy (CBT) for the treatment of anxiety (NICE, 2011). To meet increased demand, CBT has been adapted to be delivered in a low intensity (LI), psychoeducational and self-help format, in which the treatment contract and the sessions themselves are brief (Bennett-Levy et al., 2010). The Improving Access to Psychological Therapies (IAPT – now National Health Service [NHS] Talking Therapies) uses a stepped-care approach so that a guided self-help version of CBT (CBT-GSH) is provided for individuals experiencing mild-moderate anxiety at step 2. CBT-GSH is provided by psychological wellbeing practitioners (PWPs). The PWP is likened to the role of a 'coach' rather than a traditional 'therapist' (Turpin, 2010).

There are limited recent reviews of the efficacy of CBT-GSH for anxiety. Cuijpers et al's (2010) meta-analysis suggested that CBT delivered as GSH has been shown to have positive effects on symptoms of anxiety, comparable with traditional CBT. Similarly, Priemer & Talbot (2013) found no significant differences in outcomes between studies of CBT-GSH and CBT for anxiety. However, a meta-analysis completed by Coull & Morris (2011) found that CBT-GSH was effective at post-treatment for anxiety, but that this was limited at follow-up and with clinical populations. High dropout rates for CBT-GSH (Chan & Adams, 2014) suggest low treatment acceptability (Milosevic et al., 2015) and there is evidence of high relapse rates following CBT-GSH intervention (Delgadillo et al, 2018). This evidence highlighted the

need for patients to be offered a wider choice of GSH to ensure improvements in acceptability, effectiveness and durability at step 2 within NHS Talking Therapies.

To address these issues, Meadows and Kellett (2017) developed a manualised version of cognitive analytic therapy guided self-help (CAT-GSH) for delivery across the range of anxiety disorders. CAT is a relational model which assumes that early life experiences influence relationships with others and the self (Ryle & Kerr, 2002) and CAT collaboratively identifies and changes these unhelpful roles and patterns. The evidence base for CAT for anxiety is supported by clinical trials (Boogar, Rezaei & Yosefi, 2013) and cohort studies (Tzouramanis et al., 2010). CAT-GSH has been shown to have high adherence to GSH principles, generated low dropout rates, was easy to deliver and was clinically effective with a durable short-term effect (Meadows & Kellett, 2017; Wray et al., 2022).

A randomised patient preference trial comparing CAT-GSH and CBT-GSH within an NHS Talking Therapies service indicated equivalent efficacy (Kellett et al., 2023). But there was variability in recovery rates, with only 42.4% of patients meeting reliable recovery in the CAT-GSH group and 50.9% of patients in the CBT-GSH group. This suggests treatment response heterogeneity, some patients responded better to the treatments than others. This may be explained by individual differences in patient characteristics (e.g., previous treatment, symptom severity, comorbidity), which could be used to match patients to the best treatment for them. To do this, there needs to be treatment choice which until now, has not been available in NHS Talking Therapies step 2. This brings a further dilemma in how to allocate treatments; the referenced trial offered patient preference or randomisation. Most patients favoured choosing their treatment over randomisation; however, this did not impact clinical outcomes. This

presents a debate in the value of patient preference and whether evidence-based treatment allocation is more effective.

Predicting Treatment Outcome

Precision mental health care employs data-driven methods to monitor patients' treatment response, model prognosis, and personalise the treatment (Delgadillo & Lutz, 2020). An array of prediction models have been successfully integrated into routine practice across medicine, and regularly support national clinical guidelines around treatment allocation (Damen et al., 2016; NICE, 2014). There are examples of prediction algorithms that have been developed to identify who would have the best treatment response to CBT and psychodynamic therapy (Schwartz et al., 2021), CBT and eye movement desensitization and reprocessing (Deisenhofer et al., 2018), CBT and interpersonal psychotherapy (Huibers et al., 2015) and CBT and antidepressant medication (DeRubeis et al., 2014). Machine learning methods are becoming increasingly popular to enhance variable selection within prediction models, and so increasing generalisability to new samples (Delgadillo et al., 2017). The introduction of greater treatment heterogeneity at step 2 increases the need to identify which patients respond better to differing treatments.

Prediction models enable prognostic (associated with overall outcome) and prescriptive (associated with differential outcomes between interventions) treatment outcome variables to be identified (Cohen & DeRubeis, 2018). DeRubeis et al. (2014) developed an approach, the *personalised advantage index* (PAI), that integrates multiple identified outcome predictors from different treatments into one statistical model. The PAI identifies in 2 or more comparable effective treatments, one intervention that is more effective for an individual by producing counterfactual

outcome predictions. The intervention that an individual is predicted to respond better to is considered their optimal treatment and enables outcomes from optimal and non-optimal received treatments to be compared.

The PAI in DeRubeis et al. (2014) predicted a clinically meaningful advantage for 60% of patients assigned to their predicted optimal treatment for depression. Huibers et al. (2015) also used the PAI and predicted a clinically meaningful advantage for 63% of patients. Headley et al., (2024) have recently called for studies using methods such as the PAI to improve the allocation of patients in routine practice to differing versions of efficacious GSH. PAI represents a promising empirical approach in guiding services regarding efficient treatment allocation. Targeted treatment has the potential to make best use of currently available evidence-based treatments, improving outcomes for patients at no additional cost (Delgadillo & Gonzalez Salas Duhne, 2020). In the effort to better match patients to CAT-GSH and CBT-GSH, it may be possible to then improve outcomes at step 2.

Clinical Trial

The patient preference trial of CBT-GSH vs CAT-GSH for anxiety delivered at step-2 NHS Talking Therapies gave patients the option of randomisation or to choose which treatment they preferred and routine NHS Talking Therapies outcome data was collected alongside the primary outcomes of the trial (Kellett et al., 2023). The data from this clinical trial was used in the current study to develop predictive models of treatment outcome and use the PAI to determine which GSH treatment would work best for whom. To summarise, Kellett et al. (2023) found that that were no significant differences in outcome at post-treatment and 24-week follow-ups on Beck's Anxiety Inventory (BAI; Beck & Steer, 1993) between CAT-GSH and CBT-GSH. This suggests

that CAT-GSH is comparable to CBT-GSH as the 'treatment as usual' at step 2. As the trial was patient preference, it is interesting to understand whether certain patients had better outcomes if they were prescribed a treatment based on a predictive model, rather than through personal choice or randomisation.

Aims

The main objective of this study was to develop and test a personalised treatment selection method to match patients to an optimal GSH treatment for anxiety.

The study had 3 aims:

1) To use a variable selection procedure to identify baseline characteristics of patients with anxiety that significantly predict treatment outcome for CBT-GSH & CAT-GSH.
2) To use identified predictors to develop separate predictive models for each intervention and calculate the PAI to indicate the optimal GSH for each patient.
3) To assess the usefulness of the models by comparing treatment response for patients who received their optimal versus non-optimal GSH.

Due to adoption of a data-driven approach to identify predictive variables, there were no specific hypotheses regarding which predictors would hold predictive value and therefore be retained in the final PAI models. There is limited existing research to provide suggestions of prognostic and prescriptive predictors, particularly as CAT-GSH is a new treatment. It was hypothesised that patients who received their optimal treatment as indicated by the PAI would have better outcomes than patients who received their non-optimal treatment.

Ethical Approval

The current study received ethical approval from the University of Sheffield Ethics Committee (see Appendix A) and did not require NHS ethical approval due to the use of secondary data.

Method

The study is reported according to the Transparent Reporting of Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD + AI) reporting guidelines for machine learning informed prediction model studies (Collins et al., 2015). The full checklist can be found in Appendix B, anything that has not been included in the current study is discussed within limitations.

Patient and Public Involvement

It was not possible to involve a patient and public involvement (PPI) group in the current study due to the secondary nature of the research. However, a PPI group were consulted at the beginning of the clinical trial to support the development of the trial materials, particularly the language used in the presentation of CBT-GSH and CAT-GSH.

Setting

The clinical trial designed by Kellett et al., had a partially randomised patient preference design (Torgerson & Sibbald, 1998). It was a single site study at the Oldham NHS Talking Therapies service hosted by the Pennine Care NHS Foundation Trust. If deemed suitable for a step 2 intervention within the NHS Talking Therapies service during triage, patients were given the option to participate in the trial. If the

patient accepted, they were offered a trial eligibility interview. Written informed consent was obtained from all patients. Data collection began on the 29th of January 2019.

Inclusion and Exclusion Criteria

Patients were included if they met the following criteria outlined in Table 1. If patients met eligibility criteria, they were offered randomisation (block randomisation by a third party), or they could choose CBT-GSH or CAT-GSH following psychoeducation about each treatment.

Table 1

Inclusion and Exclusion Criteria

Inclusion	Exclusion
Self-referred or been referred by their General Practitioner/other health or social care professional for the treatment of a common mental health problem.	Currently taking part in another NHS Talking Therapies step 2 intervention.
Met criteria for an anxiety disorder based on the Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998).	Didn't meet the criteria for an anxiety disorder or "caseness" on the BAI.
Scored above the cut-off for clinically significant symptoms on Beck's Anxiety Inventory (BAI; Beck & Steer, 1993).	Met criteria for comorbid depression and anxiety disorder, where the depression is the main concern.
Want to engage in GSH for anxiety.	Had a severe/chronic mental health difficulty - engaged with secondary mental health care services or Diagnosis of social phobia/post-traumatic stress disorder (treated at step 3).
Motivated to engage in treatment and able to attend six face-to-face GSH sessions.	Unable to read and write or need an interpreter.

Data Collection

The main trial sample size calculation indicated that 134 patients were required to detect a small-moderate effect size in treatment differences with 80% power and $p<.05$. In total, 271 patients were eligible for inclusion and were allocated to either CBT-GSH or CAT-GSH. Patients completed trial measures at baseline (week 0), post-treatment (week 8) and follow-up (week 24).

Interventions

GSH was delivered by qualified PWPs over the telephone due to the COVID-19 pandemic. All PWPs had passed an NHS Talking Therapies 1-year post-graduate certificate in CBT-GSH following a national curriculum (UCL, 2014) and attended a 2-day CAT-GSH training session. PWPs had 1-hour per week of individual case management supervision and were enrolled in group supervision monthly for 2-hours. Both interventions had a contract of 6-8 sessions which were 35-minutes long.

CBT-GSH is a low-intensity, structured psychological intervention based on the principles of CBT. Treatment followed the NHS Talking Therapies treatment protocol (Richards & Whyte, 2011). This is treatment as usual within step 2 of NHS Talking Therapies. The patient works through a standardized treatment manual (NICE, 2009; 2011) with regular support from a PWP focussing on changing their thoughts, feelings and behaviours. The aim of treatment is to support patients in learning techniques to help manage their symptoms.

CAT-GSH is a low-intensity, structured psychological intervention based on CAT principles (Meadows & Kellett, 2017) and follows the reformulation, recognition and revision structured approach. The content of the sessions is as follows: 1) identify snags, traps & dilemmas; use self-monitoring homework; 2) develop reciprocal roles

from early experiences and associated homework; 3) write a problem statement linking past to the present; 4) create diagrammatic reformulation; 5) identify exits associated with homework; 6) work on endings and relapse prevention.

Both rely on a psychoeducational workbook to deliver the content. They differ however in terms of content, style and focus. Firstly, CBT-GSH is based in the here and now and CAT-GSH has a past-present focus. Secondly, CBT-GSH requires an effective therapeutic relationship but does not use transference/countertransference, whereas CAT-GSH works with the therapeutic relationship and makes use of transference/countertransference. Thirdly, CAT-GSH is based on a dialogical and relational theoretical model and CBT-GSH is based on a cognitive behavioural model (Meadows & Kellett, 2017).

Measures

The following data was collected in the trial and were considered as candidate baseline predictors for the analysis – age; ethnicity; employment status; sexual orientation; identification of long-term condition; veteran status; perinatal status; psychotropic medication; indices of multiple deprivation; GAD-7 baseline severity; PHQ-9 baseline severity; WSAS baseline severity; BAI baseline severity; previous treatment; allocation choice. Only variables with sufficient data to enable reliable missing data imputation were used as candidate predictors and only variables empirically selected during variable selection were included in the final PAI models (see Data Analysis section).

Four clinical outcome measures were completed at baseline, post-intervention and follow-up (see Appendix C for information gathered):

Beck Anxiety Inventory (BAI) is a tool to assess symptoms of anxiety. Items were specifically selected to separate anxiety from depression. It has 21 items with a suggested cut-off for clinically significant anxiety as 16 (Beck & Steer, 1993). The BAI has high internal consistency and has good reliability (Fydrich, Dowdall, & Chambless, 1992).

Generalised Anxiety Disorder 7 (GAD-7) is an assessment tool to support the diagnosis of generalised anxiety disorder. It has 7 items with a maximum score of 21. Scores of 5, 10, and 15 are taken as the cut-off points for mild, moderate and severe anxiety, respectively. Using the threshold score of 10, the GAD-7 has a sensitivity of 89% and a specificity of 82% for GAD (Kroenke et al., 2007). Spitzer et al. (2006), found the questionnaire has good validity and reliability. It is routinely used within NHS Talking Therapies.

Patient Health Questionnaire 9 (PHQ- 9) is an assessment tool to support the recognition of depression in patients. It has 9 items with a maximum score of 27. Total scores of 5, 10, 15, and 20 represent cut-offs for mild, moderate, moderately severe and severe depression, respectively. It can be repeated over time to monitor changes and has been assessed as having excellent test-retest reliability (Kroenke et al., 2001).

Work and Social Adjustment Scale (WSAS) is a 5-item self-report measure that provides the impact of a disorder on daily life, it has a maximum score of 40. Scores between 10 and 20 are associated with significant functional impairment but less severe clinical symptomatology, over 20 suggests moderately severe psychopathology. Mundt et al. (2002) reported that the WSAS is reliable, valid and sensitive to change.

The main trial used the BAI as the primary outcome measure; however, the current study used the GAD-7 as the primary outcome measure as it is already used in NHS Talking Therapies as a routine outcome measure for anxiety which maximises clinical utility.

Sample Characteristics

The current study sample (N = 209) included patients who accessed ≥ 2 sessions of CBT-GSH or CAT-GSH to align with Talking Therapies definition of accessing a course of treatment (The National Collaborating Centre for Mental Health, 2018). This excluded 62 cases from the original clinical trial. The mean number of attended sessions was 5.7 (SD = 2.1). More patients accessed CAT-GSH (N = 154) than CBT-GSH (N = 55) and mostly this was through choice as 93.8% of patients were allocated to treatment from their preference, rather than randomisation.

Figure 1 shows a STROBE diagram (von Elm et al., 2017), demonstrating the flow of sample selection.

Figure 1

STROBE Diagram.

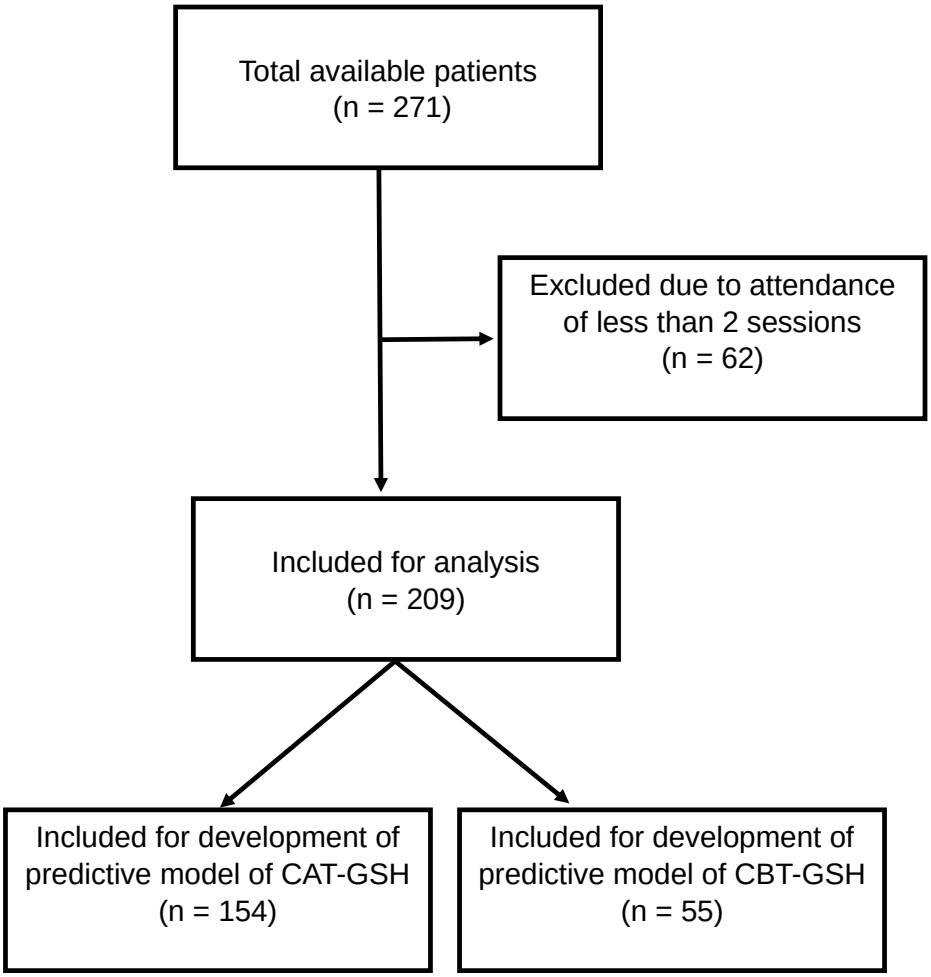


Table 2 presents a summary of sample characteristics including statistical comparisons between patients that had CAT-GSH or CBT-GSH. The samples were mostly matched, apart from significantly more females accessed CAT-GSH than CBT-GSH. There is also a significant difference in previous treatment, suggesting that more patients preferring CAT-GSH had received a previous treatment. Within the CAT-GSH sample, 94.8% had chosen that treatment and within the CBT-GSH sample, 90.9% had chosen that treatment.

Table 2

Summary of Sample Characteristics

	Full Sample (N=209)	CBT-GSH (N=55)	CAT-GSH (N=154)	Test statistic	P
Demographics					
Females[2]	75.6%	65.5%	79.2%	$X^2(1) = 4.16$	.041
Age[1]	36.49 (13.81)	36.18 (13.97)	36.60 (13.80)	$t\ (207) = 0.19$	.846
Ethnicity[2]					
White British	90.4%	94.5%	89%	$X^2(1) = 1.46$	.227
Other	9.6%	5.5%	11%		
IMD decile[1]				$t\ (207) = 0.26$	.793
1 = Poorest 10 = Affluent	4.14 (2.77)	4.05 (2.82)	4.17 (2.76)		
Unemployed[2]	12.9%	12.7%	13%	$X^2(1) = 0.002$	.961
Perinatal	6.2%	5.8%	7.3%	$t\ (207) = 0.38$	.354
Heterosexual	90%	89.6%	90.9%	$t\ (207) = 0.27$	.392
Previous CAT	1.4%	0	1.9%	$t\ (207) = 1.04$	.150
Allocation Choice[2]					
Preference	93.8%	90.9%	94.8%	$X^2(1) = 1.05$	.304
Randomised	6.2%	9.1%	5.2%		
Baseline Severity Measures					
GAD-7[1]	13.62 (4.81)	14.24 (4.67)	13.40 (4.85)	$t\ (207) = -1.11$	.267
PHQ-9[1]	13.65 (5.59)	14.09 (4.68)	13.49 (5.89)	$t\ (207) = -0.68$	.498
WSAS[1]	18.69 (8.57)	17.89 (9.23)	18.97 (8.33)	$t\ (207) = 0.80$	.422
BAI[1]	25.50 (9.82)	25.69 (9.57)	25.43 (9.94)	$t\ (207) = -0.17$	.865
LTC[2]	31.6%	34.5%	30.5%	$X^2(1) = 0.30$	.581
Previous Treatment[2]	45.9%	25.5%	53.2%	$X^2(1) = 12.61$	<.001
Medication[2]	56.5%	61.8%	54.5%	$X^2(1) = 0.87$	.350

Note. CBT-GSH = guided self-help cognitive-behavioural therapy; CAT-GSH = guided self-help cognitive-analytic therapy; IMD Decile = Index of multiple deprivation in deciles; PHQ-9 = Patient Health Questionnaire; GAD-7 = Generalized

Anxiety Disorder Questionnaire; WSAS = Work and Social Adjustment Scale; BAI = Beck's Anxiety Inventory; LTC = Long-term condition; 1 = Mean and Standard Deviation; 2 = Percentages

Sample Size Calculation

The sample size calculation proposed by Riley et al. (2019) provides an equation for multi-variate prediction models of continuous outcomes based on the required number of predictors, expected explained variance (R^2) and allowing for up to 10% out-of-sample prediction shrinkage. Although sufficient data on 18 baseline predictors was available from the trial, it was not expected that all variables would have predictive value. A variable selection procedure was applied prior to developing the models and it was anticipated that approximately seven of the predictors would be retained. Literature on previous psychological treatment prediction models for continuous outcomes reported R^2 values in the range of .35 to .45, indicating that baseline characteristics tend to explain on average 40% of variance in post-treatment scores (Friedl et al., 2020; Salomonsson et al., 2020; Schwartz et al., 2021; Senger et al., 2022). On this basis, a sample size calculation for a predictive model using an estimate value of R^2 =0.40, including 7 predictors and allowing for up to 10% shrinkage would require 79 patients in each treatment group.

The sample used in this study included 209 patients, however, this was not equal across both treatment groups as 55 patients received CBT-GSH and 154 patients received CAT-GSH. The number of patients in the CBT-GSH group falls short of the sample size calculation for suitable power. Due to this, keeping a subset of the data for an external cross-validation was not feasible. Using the full dataset maximised power for development of the models, as shown in DeRubeis et al. (2014).

Data Preparation

Missing data was previously imputed as part of the main trial analysis, however some of the candidate predictors for this study were not used for the primary main trial analysis and were therefore imputed for the purpose of this study. The R package "missForest" (Stekhoven & Bühlmann, 2012) on R (R Core Team, 2021) and Rstudio (Rstudio Team, 2021) was used to input missing data using a random forest approach, separately for the two interventions. This parametric method handles outliers well, is robust and reliable. Marital status was removed as a candidate predictor as it had too much missing data to be reliably imputed (86%).

If not already, categorical variables were collapsed into binary variables for the purpose of data analysis i.e., employment status became unemployed or employed/other; sexual orientation became heterosexual or not heterosexual; ethnicity became white British or minoritised background. Continuous variables were standardized into Z-scores and binary variables were also dummy coded as -0.5 and 0.5. This put the variables into common scales so they could be compared.

Variable Selection

Data was separated into CAT-GSH and CBT-GSH subsets for the purpose of variable selection and building predictive models for each treatment. Two machine learning approaches (a decision tree method and a penalized regression method) were used to conduct variable selection in each intervention dataset and the regression predictions were compared using evaluation metrics to identify the best fitting model. This meant that a total of four variable selection models were produced: separate decision-tree variable selection models for CAT-GSH and CBT-GSH and separate penalised regression variable selection models for CAT-GSH and CBT-GSH. A total of

18 variables were included as predictors in in each model, with post-treatment GAD-7 scores as the dependent variable. Baseline GAD-7 scores were not included in this analysis as it was later included in the regression models (by forced entry) to control for anxiety levels at screening (an approach adopted in other PAI studies; Moggia et al. 2023).

The first variable selection method was the Boruta approach, selected as it can handle multivariate interactions and was conducted in R using the “Boruta” package (Kursa & Rudnicki, 2010). Boruta is a form of random forest, which is a supervised machine learning algorithm that builds multiple decision trees using a bagging method (a combination of variable and bootstrapping samples; Breiman, 2001).The Boruta extension includes shadow variables (one continuous, one categorical) based on the distributions of other variables in the dataset and included in the model as a ‘noise’ variable (i.e., have no actual predictive power). Only predictor variables which are ranked higher than one (tentative inclusion) or both (confirmed inclusion) shadow variables are deemed to have reliable predictive power over and above noise and are retained (see Appendix D for supplementary data analysis information).

There are known potential issues of biased importance values and overfitting when using random forest approaches for variable selection, particularly when there are a smaller number of candidate variables (Tang et al., 2018). Therefore, a second variable selection method was also tested employing elastic net regularisation variable selection (Zou & Hastie, 2005). Elastic net is a linear regression technique that uses a penalty term to shrink coefficients of predictors that are unimportant. It identifies variables that are reliably associated with an outcome but also adds more or less “weight” to variables with stronger or weaker predictive value (see Appendix D).

To decide which predictive algorithm to use, variables identified by each model were entered into separate linear regressions to produce predicted outcomes. Leave-one-out cross-validation (LOOCV) was used to prevent over-fitting (Efron, 1982). LOOCV estimates each model without information about the participant whose score is being predicted therefore uses a sample of (*n*-1), with *n* being the sample size. This aims to reduce bias in predicted values. Analysis was performed using the "caret" package (Kuhn, 2008) within R. Predicted outcomes from each regression model were compared with the observed outcomes and evaluation metrics were compared. These included the Root Mean Squared Error (RMSE) which measured the average difference between predicted and actual values; R squared which represented the variance within the outcome measure explained by the regression model; Mean Absolute Error (MAE) which measured the errors between predicted and actual scores and Pearson's Correlation Coefficient which measured the strength of the relationship between predicted and observed scores. The model with the lowest error and highest correlation between actual and predicted scores was chosen as the preferred model.

PAI Estimation

The preferred regression model for each GSH intervention was used to predict post-treatment GAD-7 scores for both CAT-GSH and CBT-GSH in the full sample (N=209). This produced a predicted score for both treatment modalities for each patient. The PAI was estimated for every individual patient by calculating the difference between their predicted post-treatment GAD-7 score for each treatment (positive PAI value indicated greater benefit from CAT-GSH; a negative PAI indicated greater benefit from CBT-GSH).

Assessment of the PAI

Patients who had received the GSH intervention recommended by the PAI (based on a positive or negative value) were classified into the optimal treatment group, whereas those who did not were classified into the non-optimal group following the approach used by Moggia et al. (2023). Observed post-treatment and follow-up GAD-7 scores and reliable and clinically significant improvement (RCSI) rates were compared between the optimal and non-optimal groups using mixed analysis of variance (ANOVA; for baseline, post-treatment and 24-week follow-up timepoints) and Chi-squared tests (post-treatment and 24-week follow-up) respectively. RCSI was defined as GAD-7 change score > = 4 and post-treatment/follow-up scores to have moved below the clinical cut-off of 8 (Jacobson et al., 1999; The National Collaborating Centre for Mental Health, 2018). These analyses were conducted within IBM SPSS Statistics (Version 27; IBM Corp, 2020). An a priori power analysis conducted using G*Power version 3.1.9.7 (Faul et al., 2007) indicated the required sample size to achieve 80% power ($\alpha = .05$) for detecting a medium between-group effect in outcomes was N = 86 for a mixed ANOVA. Data was analysed to assess assumptions required for an ANOVA using the Shapiro-Wilks and Levene's Tests (see Appendix E for full results). Assumption testing suggested that homogeneity of variance was met, however, some of the data was not normally distributed. Due to the robustness of the ANOVA, the analysis was still completed with evidence suggesting that this was appropriate (Blanca et al., 2017).

To explore the impact on outcomes in those who had the biggest indicated PAI benefit, a secondary analysis subgroup was identified including optimal and non-optimal groups based on patients with a PAI value one standard deviation larger than the sample mean following the approach used by Delgadillo & Gonzalez Salas Duhne

(2020). If the PAI was smaller than this sum, no optimal treatment was indicated. Between-group comparison of outcomes analyses was repeated in this subsample.

A brief comparison of patient preference and optimal treatment allocation was also conducted by observing percentages to further investigate the impact of patient preference.

Results

Variable Selection

Results from the Boruta variable selection model for CAT-GSH and CBT-GSH are shown in Figure 2 and 3, respectively. Each model's variables are ranked from the highest to the lowest importance (according to the mean decrease in accuracy). For CAT-GSH, 5 variables (27.78%) were selected as potential predictors (i.e., previous CBT, sexual orientation, baseline BAI score, baseline PHQ-9 and baseline WSAS). For CBT-GSH, 4 variables (22.22%) were selected as potential predictors (i.e., employment status, baseline BAI, indices of multiple deprivation and baseline PHQ-9).

Figure 2

Boruta Model Variable Importance for CAT-GSH (green denotes retained variables ranked higher than all shadow variables denoted by blue; red denotes discarded variables)

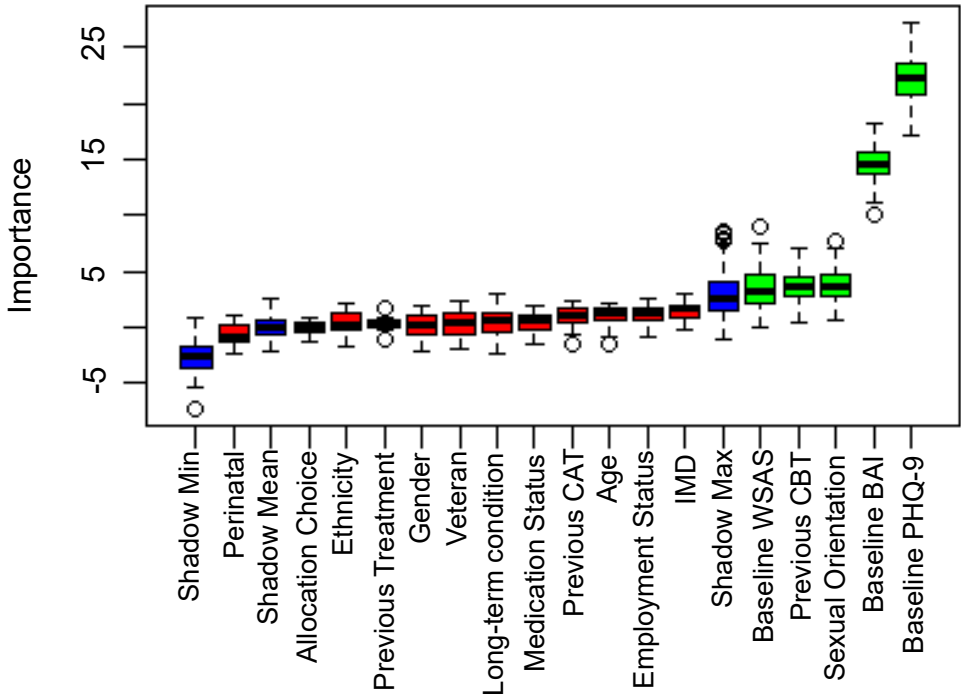
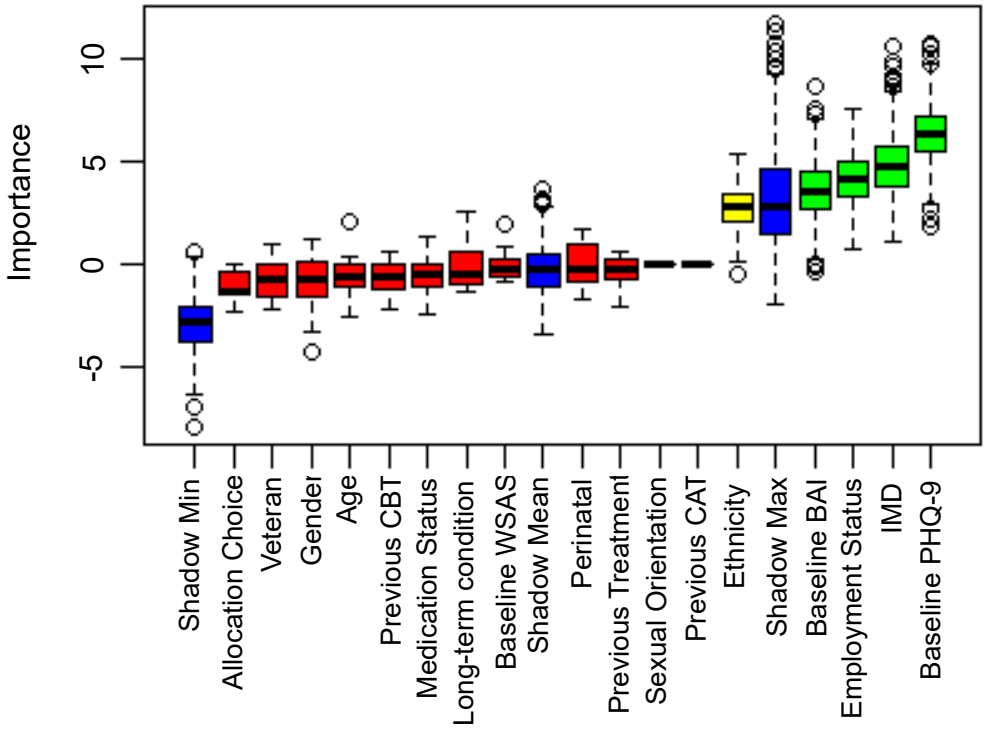


Figure 3

Boruta Variable Selection Model for CBT-GSH (green denotes retained variables ranked higher than all shadow variables denoted by blue; yellow denotes tentative included variables ranked higher than some shadow variables; red denotes discarded variables)



Results from the elastic net variable selection model for CAT-GSH and CBT-GSH are shown in Table 3. They present each model's variables 18 coefficients, shrinking unimportant variables to 0. For CAT-GSH, 8 variables (44.44%) were selected as potential predictors (i.e., baseline PHQ-9, baseline WSAS, baseline BAI, long-term condition, perinatal status, sexual orientation, previous CAT and previous CBT). For CBT-GSH, 5 variables (27.78%) were selected as potential predictors (i.e., indices of multiple deprivation, baseline PHQ-9, baseline BAI, ethnicity, and employment status).

Table 3

Elastic Net Variable Selection Coefficients

Variable	CAT-GSH Coefficient	CBT-GSH Coefficient
Age	0	0
IMD	0	-0.344432620
Baseline PHQ-9	2.11187076	0.602385780
Baseline WSAS	0.04951514	0
Baseline BAI	0.55903206	0.189781429
Allocation Choice	0	0
Gender	0	0
Ethnicity	0	-0.001057776
Employment Status	0	1.085421116
Long Term Condition	0.67118767	0
Medication Status	0	0
Veteran Status	0	0
Perinatal Status	1.07732136	0
Sexual Orientation	0.13674917	0
Previous Treatment	0	0
Previous CAT	0.51312583	0
Previous CBT	-1.43364222	0

Estimation of LOOCV Regressions

The evaluation metrics for the LOOCV regressions including the variables identified by each variable selection model are shown in Table 4. The elastic net models outperformed Boruta on all metrics for CAT-GSH. Due to both elastic net and Boruta selecting the same variables for CBT-GSH, all evaluation metrics were identical. Therefore, the elastic net variable selection was selected as the preferred model for both interventions to ensure congruence.

Table 4

LOOCV Regression Results

	CAT-GSH			
	RMSE	**R^2**	**MAE**	***r***
Boruta	4.09	0.32	3.21	0.62
Elastic Net	4.03	0.33	3.12	0.64
	CBT-GSH			
	RMSE	**R^2**	**MAE**	***r***
Boruta	4.50	0.10	3.62	0.54
Elastic Net	4.50	0.10	3.62	0.54

Final Selected Models

Prognostic variables across the treatments were baseline PHQ-9 scores and BAI scores with lower scores suggesting better outcomes. Prescriptive variables were also identified and within CBT-GSH, patients from a higher socioeconomic status were more likely to have better outcomes, along with White British patients and unemployed patients. Within CAT-GSH, better treatment outcomes were associated with lower baseline WSAS scores, a self-reported long-term condition, the perinatal period and identifying as not heterosexual. Previously engaging in CBT was associated with poorer outcomes, whilst previously engaging in CAT was associated with better outcomes.

Optimal Guided Self-Help

Based on +/- PAI values in the full sample (N = 209), CBT-GSH was indicated as the optimal treatment for 34.9% of the sample and CAT-GSH was indicated as the optimal

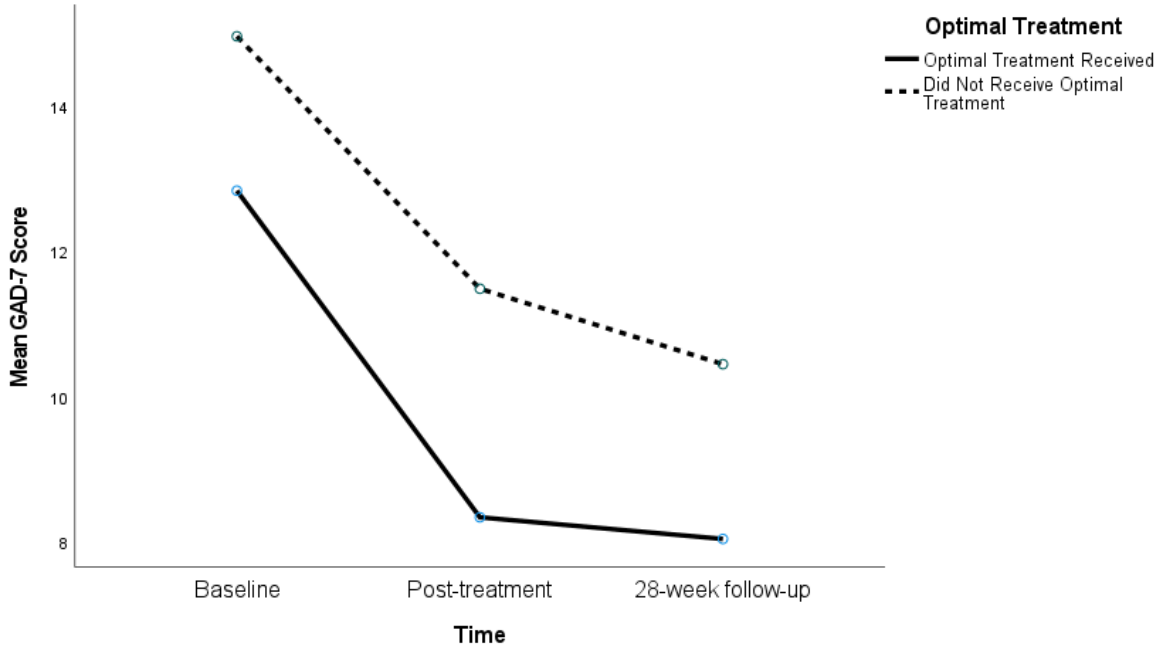
treatment for 65.1%. Classifications indicated 62.7% received their optimal GSH and 37.3% did not received their optimal GSH.

There was a significant between subjects' effect of receiving optimal GSH on GAD-7 outcomes (F (1,207) = 21.675, p < .001) and the partial Eta squared (η^2 = .10) suggested a medium effect size (Miles and Shevlin, 2001). Mauchly's test indicated that the assumption of sphericity had been violated within the repeated measures analysis, $\chi 2$ (2) = 39.62, p = <.001. Due to this, the Greenhouse-Geisser corrected results were reported. There was a significant within subjects' effect of time on GAD-7 outcomes (Greenhouse–Geisser's ε = .85, F (1.7, 352.35) = 117.54, p <. 001, η^2 = .36) suggesting a large effect size. However, the interaction effect between GAD-7 scores over time and receiving optimal GSH, was non-significant (Greenhouse–Geisser's ε = .85, F (1.7, 352.35) = 1.29, p = .273).

Figure 4 displays the mean GAD-7 scores at baseline, post-treatment and follow-up for patients who received optimal versus non-optimal GSH.

Figure 4

Mean GAD-7 scores across timepoints for the optimal versus non-optimal treatment groups in the full sample



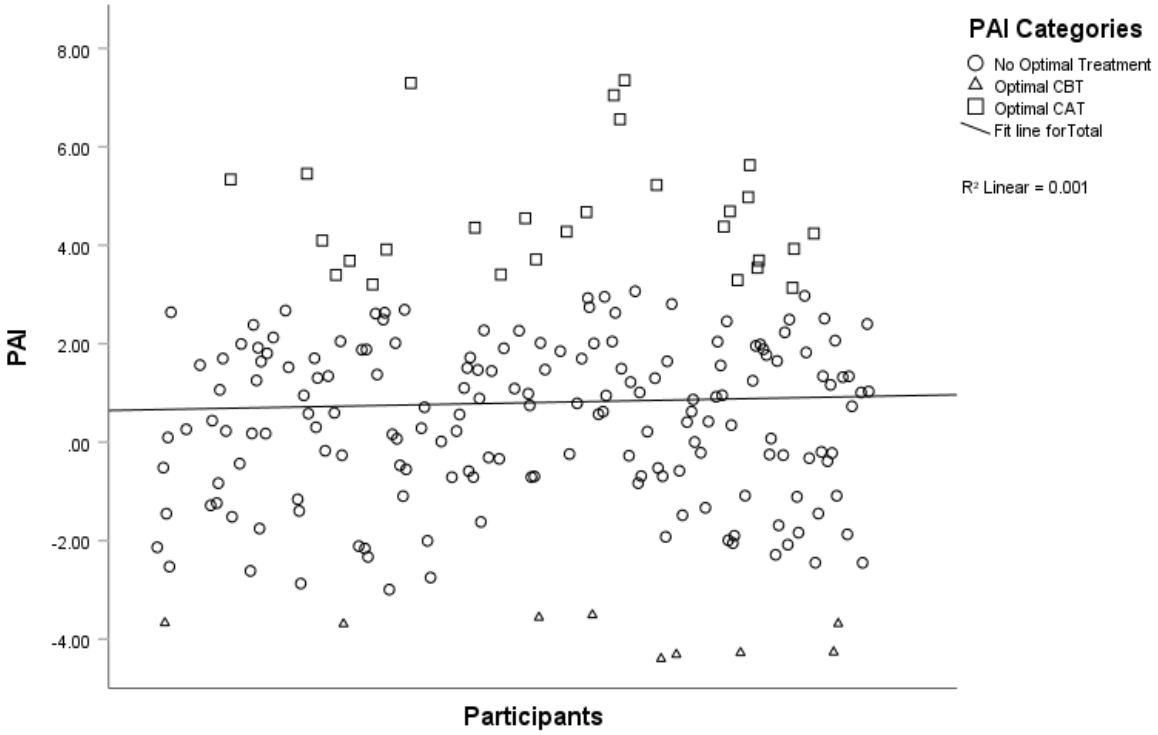
At post-treatment, significantly more patients who received their optimal GSH met RCSI (35.9%) compared to patients who did not (16.6%); X^2 (1, N = 209) = 8.82, p = .003). The odds ratio indicated that patients who received their PAI recommended GSH were more than twice as likely to recover (OR = 0.36). The same pattern was observed at 24-week follow-up, with significantly more patients who had their optimal treatment experiencing RCSI (36.6%) compared to those who did not (19.2%; X^2 (1, N = 209) = 7.04, p=.008) indicating that receiving the PAI recommended GSH more than doubled the chance of longer-term recovery (OR = 0.41).

Subgroup Analysis

A subgroup of N = 37 who had the largest indicated PAI benefit (PAI >= 3.08 [mean +/- 1 SD]) was identified. Based on +/- PAI values in the subgroup, CBT-GSH was indicated as the optimal treatment for 4.3% of the sample and CAT-GSH was indicated as the optimal treatment for 13.4%. No optimal treatment was identified for 82.3% of the subgroup (PAI < 3.08). Figure 5 represents the distribution of the PAI.

Figure 5

Distribution of Optimal Treatments



Optimal Guided Self-Help – Subgroup

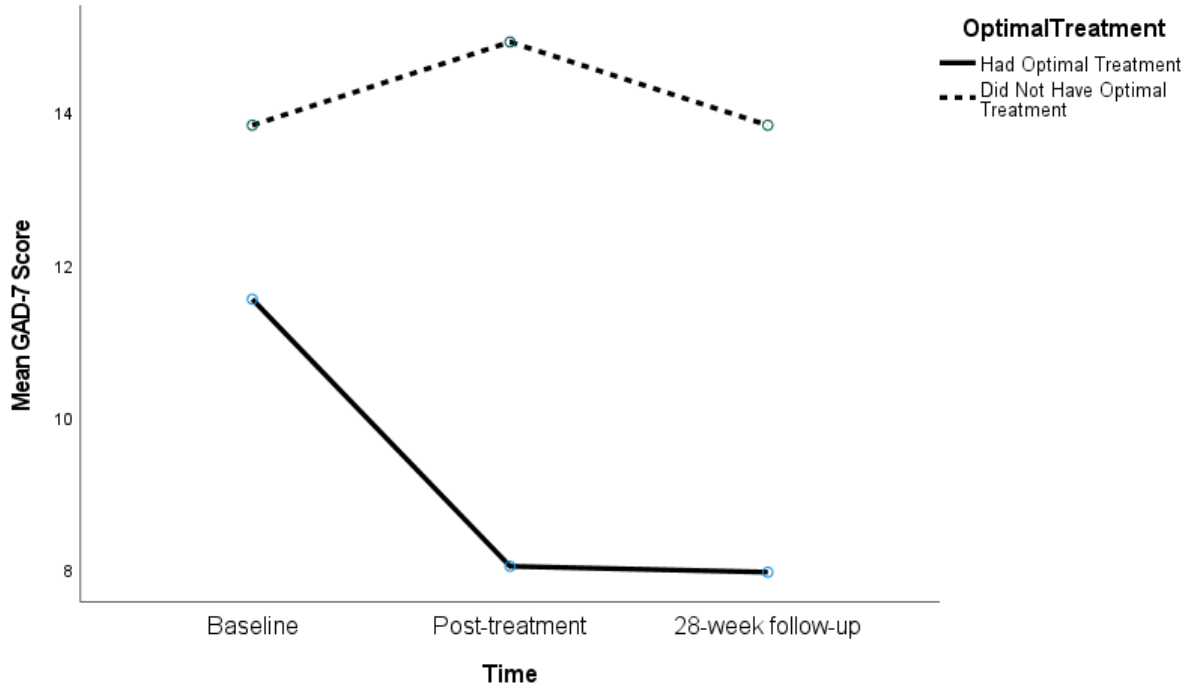
Classifications indicated 70.3% received their optimal GSH and 29.7%% did not received their optimal GSH.

There was a significant between subjects' effect of receiving optimal GSH on GAD-7 outcomes (F (1, 35) = 12.296, p = .001) and the partial Eta squared ($\eta 2$ = .26) suggested a large effect size (Miles and Shevlin, 2001). Mauchly's test indicated that the assumption of sphericity had been violated ($\chi 2$ (2) = 14.94, p = <.001). Due to this, the Greenhouse-Geisser corrected results were reported. There was a non-significant within subjects' effect of time on GAD-7 outcomes (Greenhouse–Geisser's ε = .78, F (1.48, 51.64) = 2.19, p = .119). The interaction effect between GAD-7 scores over time and receiving optimal GSH, was significant (Greenhouse–Geisser's ε = .78, F (1.48, 51.64) = 3.83, p = .040, $\eta 2$ = .01), suggesting a medium effect size.

Figure 6 displays the mean GAD-7 scores at baseline, post-treatment and follow-up for patients who received optimal versus non-optimal GSH.

Figure 6

Mean GAD-7 scores across timepoints for the optimal versus non-optimal treatment groups in the subgroup



Of patients that did not have their optimal GSH, none met RCSI at post-treatment or 24-week follow-up. Those patients that had their optimal GSH then 30.8% met RCSI and this difference was significant (X^2 (1, N = 37) = 4.32, p = .038). The odds ratio also suggests that patients who had their optimal GSH had a better outcome (OR = 0.69).

At post-treatment, significantly more patients who received their optimal GSH met RCSI (30.8%) compared to patients who did not (0%); X^2(1, N = 37) = 4.32, p = .038). The odds ratio indicated that patients who received their PAI recommended GSH were more likely to recover (OR = 0.69). The number of patients meeting RCSI criteria at 24-week follow-up did not change from post-treatment.

Patient Preference

Table 6 reports the optimal GSH recommendations in the context of the patient's trial treatment preference. Of patients whom the PAI recommended CAT-GSH, 74.26% stated a preference for it. Of those recommended CBT-GSH, 61.64% preferred CAT-GSH. We can assume that patient preference has not impacted the PAI as although more than half of patients showed a preference for CAT, they would have had better outcomes from CBT-GSH according to this model.

Table 6

Optimal Treatment Conditions and Patient Preference

	Preference for CAT (%)	**Preference for CBT (%)**	**Randomised (%)**
Optimal CAT-GSH	74.26	19.85	5.89
Optimal CBT-GSH	61.64	31.51	6.85

Discussion

This paper is the first to test whether better matching of patients to differing GSH treatments could improve the effectiveness of these brief LI interventions. This has not been possible in the past due to limited treatment options delivered at step 2, but this is starting to change. Kellett et al. (2023) for example were the first to introduce an analytically informed one-to-one GSH into an NHS Talking Therapies services, and Lemma and Fonagy (2013) have developed and evaluated group psychoeducational psychodynamic therapy. The results of this study have shown the importance of the plurality of treatment options at step 2 and provide preliminary evidence that it may be

possible to match patients to different GSH treatments as a method of improving outcomes. The present study also produces an argument for evidence-based treatment allocation championing patient preference.

Main Findings

Results suggest that patients who received their optimal GSH were significantly more likely to have lower anxiety scores and meet criteria for RCSI for anxiety outcomes (GAD-7), compared to patients who received their non-optimal GSH within the subgroup, who the PAI predicted would have the most differential treatment benefit. Within the full sample, patients who received their optimal GSH were significantly more likely to meet criteria for a RCSI for anxiety outcomes (GAD-7), compared to patients who received their non-optimal GSH. However, having an optimal treatment did not significantly affect the trajectory of anxiety outcome scores. This difference in results between the full sample and subgroup may be due to the stronger optimal treatment recommendations indicated by the largest predicted benefit (PAI >= mean+1SD). These results suggest there is a subgroup of patients (18%) who have a differential response to GSH and could benefit from treatment matching.

This study suggests that matching patients to their optimal GSH could support better treatment outcomes from a brief LI intervention for subgroup of patients. This fits with existing evidence of the use of the PAI in estimating optimal treatments that have been conducted with traditional psychotherapies. DeRubeis et al. (2014), used a similar method and found that patients matched to their optimal treatment for depression (medication versus CBT) had superior outcomes. Similarly, Huibers et al. (2015), found that patients fared better having their predicted optimal treatment than non-optimal (cognitive therapy or interpersonal psychotherapy) for depression. It is

recommended that the PAI should continue to be researched within clinical populations to develop robust algorithms for matching patients to their optimal treatments.

Patients who received their indicated non-optimal treatment had higher average baseline GAD-7 scores. It is unclear why this is the case; it may be due to the decision to force entry baseline GAD-7 scores into the PAI model rather than include them during the variable selection process of the prediction models. Although baseline anxiety severity was captured by the BAI during variable selection, consideration of how GAD-7 scores may interact with other variables during variable selection should be considered within future research. The high rate of randomisation refusal in the patient preference trial meant there was a lack of randomisation in the sample and these differences may also reflect selection biases in the data that was not controlled for.

Prediction Model

Only two variables were prognostic and were both related to baseline clinical severity (baseline PHQ-9 score, baseline BAI score). Similar findings have been seen for low intensity CBT treatments for anxiety with baseline depression and anxiety identified as significant predictors of outcome (González-Robles et al., 2021; Lawn et al., 2019; Mathiason et al., 2018). The remaining variables were prescriptive. Patients who were from a minoritized background, were employed and had a lower socioeconomic status were predicted to have poorer outcomes after CBT-GSH. Patients who had higher baseline WSAS scores, no self-reported long-term condition, were not in the perinatal period, identified as heterosexual and had previously engaged in CBT were predicted to have poorer outcomes after CAT-GSH.

Some of the prescriptive variables fit with existing evidence as El Alaoui et al., 2015; Delgadillo et al., 2016 and Mathiason et al., 2018 suggest that employment status significantly predicts outcomes. Monthly income has previously been associated with outcomes but not significantly (Chen et al., 2020). Delgadillo et al. (2017) reported that an accumulation of disadvantages, such as minority ethnic status, may impact psychological improvement. Furthermore, Delgadillo & Gonzalez Salas Duhne (2020) also found that minoritised ethnic groups had poorer outcomes after a CBT intervention. Overall, the variables selected in the model predicting outcomes after CBT-GSH fit with the wider literature.

Due to the novelty of a LI CAT treatment, there was no existing evidence on predictors of outcome. There is also a limited evidence base investigating predictors of outcome after CAT. Ryle and Golynkina (2000) suggested that employment status and previous self-harm were significant predictors of outcome for borderline personality disorder (BPD) following CAT. This study did not find employment status to be a predictor of outcome, however the two papers investigated different outcomes (i.e., anxiety versus BPD symptoms). It cannot be concluded whether the variables included in the predictive models for CAT-GSH fit with the current evidence base.

It is unclear why the variables that were chosen for the CAT-GSH prediction model were deemed important and this is partly due to a lack of previous evidence. On closer review, perinatal status, sexual orientation and previous CAT were significantly skewed. Only 6.2% of the sample were within the perinatal period and 10% identified as not heterosexual. There were also only 3 patients (1.4%) that had engaged in traditional CAT previously, which were all in the CAT-GSH group. This may provide some explanation as to why these variables were included within the prediction model as the data was skewed. These may be adding noise rather than useful predictive

power and highlight the need to externally validate these models in a new sample to see if the findings replicate before firm conclusions can be drawn or applied in clinical practice.

The elastic net algorithm was selected as the best performing model over the Boruta approach. This aligns with previous research that suggests decision-tree approaches can be less accurate when the number of predictors is small (Tang et al., 2018). However, the evaluation metrics of the selected CBT-GSH model were very poor and suggest clinical utility of the current model may be limited. Retraining in a larger sample and externally validating would test whether the poor fit was due to the limited sample size.

Patient Preference

Patient preference was not identified as an indicator of outcome. However, patient preference samples were skewed with only 6.2% of the sample randomised. Significantly more patients chose CAT-GSH than CBT-GSH. Biases behind patient preference should be considered such as previous treatment and interests in certain therapy models which in turn, might influence treatment choice and outcomes (Kawathekar, 2023).

This highlights further questions about the usefulness of treatment allocation via patient preference versus treatment optimisation using artificial intelligence (AI). Previous studies suggest that using machine learning to match patients to treatment has the potential to improve outcomes for patients (Delgadillo & Gonzalez Salas Duhne, 2020). Whilst evidence for the usefulness of patient preference is mixed with some studies finding no significant effect of patient preference on outcomes (Dunlop et al., 2017). Yet, a meta-analysis investigating the impact of patient preference found

a small significant effect and suggested that patient preference was underestimated (Swift & Callahan, 2009). There are mechanisms underlying why patient preference is beneficial to pursue such as motivation to engage and therapeutic alliance (Gelhorn et al., 2011). However, optimisation through AI could be viewed as currently, the most objective method of matching patients to treatment.

Machine learning is becoming more widely used as a method of treatment matching and some studies have suggested that this should be combined with patient preference to ensure a shared decision-making process (Hamilton et al., 2024). It may be that this is the most cautious and ethical way forward to embed machine learning into mental health treatment. This is considering public views of machine learning, such as feelings that humans are too unique to be understood by machines (Ipsos, 2017).

Within the subgroup analysis, there was a large proportion of the sample identified that did not have a strong optimal treatment identified (82%), which suggests that between CAT-GSH and CBT-GSH, most of the sample would have had similar outcomes. This is important when considering clinical implications of this study and whether for this subsample of patients, patient preference can be incorporated into treatment choice if evidence suggests that there is a small significant effect on outcome (Swift & Callahan, 2009). This raises the possibility of also conducting patient preference trials where patients are allocated to a treatment using AI but are then allowed an ‘opt-out’ based on their treatment preferences, once they better understand what the different treatments entail.

Limitations

Due to the infancy of the evidence base for CAT-GSH, there were some limitations in the sample. In the Kellett et al. (2023) patient preference trial significantly more patients received CAT-GSH (N = 154) than CBT-GSH (N = 55). This resulted in external cross validation of the prediction model being unfeasible. It was therefore inappropriate to separate the samples into training and test samples. For example, the prediction model for CBT-GSH would then have been trained on around 27 cases, negatively impacting on the model's validity and accuracy. While the decision was made to maximise the sample size for training the models, this algorithm should be viewed as a proof of concept for treatment matching with GSH interventions, an approach advocated by Fransén et al. (2022) & Lutz et al. (2018) when sample sizes are suboptimal and underpowered. The use of internal cross validation within the building of the predictive model increased the reliability, but external cross validation of these models in an independent sample is required to evaluate whether this treatment matching algorithm generalises to new data before its clinical utility can be determined. Use of routine outcome data from NHS Talking Therapies services implementing the CAT-GSH intervention alongside standard CBT-GSH could be used to validate these models in a larger sample using established statistical matching methods (e.g., propensity score matching) to better control for baseline differences.

It is important to consider the difference in sample sizes in relation to results as these will have had a negative impact on model development. Predictive models could have been skewed as significantly more women had CAT-GSH than CBT-GSH as did patients who had had a previous treatment.

Model evaluation metrics for the Boruta method and Elastic Net regularisation for CBT-GSH were exactly the same and explained less variance than the CAT-GSH predictive model. This is because both models used almost the same variables (Elastic Net added ethnicity) which may be a consequence of a small sample. A larger sample size might have yielded a predictive model with greater accuracy. It could be that less patients were identified as having CBT-GSH as their optimal treatment due to a weaker prediction model.

Furthermore, categorising optimal treatments by having a PAI of either above or below 0 (as Moggia et al., 2023), leaves it unclear whether some patients had an optimal treatment that was clinically meaningful which was reflected in the non-significant results. However, this analysis yielded more power than the subgroup analysis, which did not meet the sample size requirements for suitable power (N = 37).

The impact of health inequalities between the different sociodemographic groups may not have been accurately presented within the prediction models, due to their transformations into binary variables. For example, the spectrum of different ethnicities within the study was not considered (white British or minoritised background), similarly with sexual orientation (heterosexual or not heterosexual). Future research should consider including more variables for the prediction model to capture the range of patient's sociodemographic backgrounds.

Treatment Options and Future Research

These findings suggest some patients within this sample could have had better outcomes if they had engaged in a different version of GSH. This is an important finding as currently in NHS services, CBT-GSH is the only treatment widely available at step 2. It is suggested that services should continue to adopt other therapeutic

models into GSH to offer to patients' wider choice of interventions. Choice needs to be backed up with clear psychoeducational materials so that patient preferences are informed by the clearest descriptions and the best evidence. The algorithm developed within this study for matching patients to each GSH should be evaluated further within a larger sample of patients. It is hoped that with more robust testing of the algorithm, it could be incorporated into NHS services to match patients to treatment, to ensure the best possible outcomes.

Conclusion

Results from this study have suggested that matching patients to treatments can improve outcomes for brief GSH interventions for a subsample of patients. This study supports the use of machine learning and prediction models to enable treatment matching, suggesting that using trained models holds the potential of improving outcomes. However, where there is no strong recommended treatment, patient preference could guide treatment allocation. Future research should continue to build on this evidence base, indexing improvements to patient outcomes to increase public confidence in the use of AI. The study again underlines the need for treatment plurality at Step 2 of the NHS Talking Therapies. It seems particularly important to match patients to treatments at step 2 due to their brevity (6-8 sessions), as there is less time and opportunity to adapt the GSH when they are not effective for the individual. The results need to be considered with caution due to the unbalanced samples and lack of external validation to assess generalisability. There are no firm conclusions from the individual predictors of outcome, particularly for CAT-GSH. Further research work is clearly indicated with larger, more balanced samples to better understand who CAT-GSH may work best for.

References

Beck, A. T., & Steer, R. A. (1993). *Beck Anxiety Inventory manual*. Psychological Corporation.

Bennett-Levy, J., Richards, D. A., & Farrand, P. (2010). Low intensity CBT interventions: A revolution in mental health care. *In J. Bennett-Levy & P. Farrand (Eds.), Oxford guide to low intensity CBT interventions* (Vol. 1). https://doi.org/10.1093/med:psych/9780199590117.003.0001

Blanca, M. J., Alarcón Postigo, R., Arnau Gras, J., Bono Cabré, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema, 29(4),* 552-557. https://doi.org/10.7334/psicothema2016.383

Boogar, I.R., Rezaei, A.M., & Yosefi, A. (2013). The effectiveness of cognitive analytic therapy on the severity of symptoms in patients with obsessive-compulsive disorder. *Practice in Clinical Psychology, 1*, 197-204.

Breckler, S. J. (1990). Applications of covariance structure modelling in psychology: Cause for concern?. *Psychological bulletin, 107*(2), 260. https://doi.org/10.1037/0033-2909.107.2.260

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chan, S.W.Y. & Adams, M. (2014). Service use, dropout rate and clinical outcomes: A comparison between high and low intensity treatments in an IAPT service. *Cognitive and Behavioural Psychotherapy, 42 (6)*, 747-759. https://doi.org/10.1017/S1352465813000544

Chen, H., Rodriguez, M. A., Qian, M., Kishimoto, T., Lin, M., & Berger, T. (2020). Predictors of treatment outcomes and adherence in internet-based cognitive behavioral therapy for social anxiety in China. *Behavioural and Cognitive Psychotherapy*, *48*(3), 291–303. https://doi.org/10.1017/S1352465819000730

Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment selection in depression. *Annual Review of Clinical Psychology*, *14*, 209-236. https://doi.org/10.1146/annurev-clinpsy-050817-084746

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Journal of British Surgery, 102*(3), 148-158. https://doi.org/10.1111/1471-0528.13244

Coull, G., & Morris, P. G. (2011). The clinical effectiveness of CBT-based guided self-help interventions for anxiety and depressive disorders: a systematic review. *Psychological Medicine*, *41*(11), 2239–2252. https://doi.org/10.1017/S0033291711000900

Cuijpers, P., Donker, T., Van Straten, A., Li, J., & Andersson, G. (2010). Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychological Medicine, 40*, 1943. https://doi.org/10.1017/S0033291710000772

Damen, J. A., Hooft, L., Schuit, E., Debray, T. P., Collins, G. S., Tzoulaki, I., & Moons, K. G. (2016). Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ, 353*. https://doi.org/10.1136/bmj.i2416

Deisenhofer, A.K., Delgadillo, J., Rubel, J. A., Böhnke, J. R., Zimmermann, D., Schwartz, B., & Lutz, W. (2018). Individual treatment selection for patients with posttraumatic stress disorder. *Depression and Anxiety, 35*(6), 541–550. https://doi.org/10.1002/da.22755

Delgadillo, J., & Gonzalez Salas Duhne, P. (2020). Targeted prescription of cognitive–behavioral therapy versus person-centred counselling for depression using a machine learning approach. *Journal of Consulting and Clinical Psychology, 88*(1), 14. https://doi.org/10.1037/ccp0000476

Delgadillo, J., Huey, D., Bennett, H., & McMillan, D. (2017). 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., & Lutz, W. (2020). A development pathway towards precision mental health care. *JAMA psychiatry, 77*(9), 889-890. https://doi.org/10.1001/jamapsychiatry.2020.1048

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

Delgadillo, J., Rhodes, L., Moreea, O., McMillan, D., Gilbody, S., Leach, C., Lucock, M., Lutz, W., Ali, S. (2018). Relapse and Recurrence of Common Mental Health Problems after Low Intensity Cognitive Behavioural Therapy: The WYLOW Longitudinal Cohort Study. *Psychotherapy and Psychosomatics, 87*(2),116-117. https://doi.org/10.1159/000485386

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. *PloS one, 9*(1), e83875. https://doi.org/10.1371/journal.pone.0083875

Dunlop, B. W., Kelley, M. E., Aponte-Rivera, V., Mletzko-Crowe, T., Kinkead, B., Ritchie, J. C., Nemeroff, C. B., Craighead, W. E., & Mayberg, H. S. (2017). Effects of patient preferences on outcomes in the Predictors of Remission in Depression to Individual and Combined Treatments (PReDICT) study. *American Journal of Psychiatry, 174*(6), 546-556. https://doi.org/10.1176/appi.ajp.2016.16050517

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Society for Industrial and Applied Mathematics (SIAM).

El Alaoui, S. E., Ljótsson, B., Hedman, E., Kaldo, V., Andersson, E., Rück, C., Andersson, G., & Lindefors, N. (2015). Predictors of symptomatic change and adherence in internet-based cognitive behaviour therapy for social anxiety disorder in routine psychiatric care. *PloS One*, *10*(4), e0124258–e0124258. https://doi.org/10.1371/journal.pone.0124258

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191. https://doi.org/10.3758/BF03193146

Fransén, J., Lundin, J., Fredén, F., & Huss, F. (2022). A proof-of-concept study on mortality prediction with machine learning algorithms using burn intensive care

data. *Scars, burns & healing, 8*, 20595131211066585. https://doi.org/10.1177/20595131211066585

Friedl, N., Krieger, T., Chevreul, K., Hazo, J. B., Holtzmann, J., Hoogendoorn, M., Kleiboer, A., Mathiasen, K., Urech, A., Riper, H., & Berger, T. (2020). Using the personalized advantage index for individual treatment allocation to blended treatment or treatment as usual for depression in secondary care. *Journal of Clinical Medicine*, *9*(2), 490. https://doi.org/10.3390/jcm9020490

Fydrich, T., Dowdall, D., & Chambless, D. L. (1992). Reliability and validity of the beck anxiety inventory. *Journal of anxiety disorders*, *6*(1), 55-61. https://doi.org/10.1016/0887-6185(92)90026-4

Gelhorn, H. L., Sexton, C. C., & Classi, P. M. (2011). Patient preferences for treatment of major depressive disorder and the impact on health outcomes: a systematic review. Primary Care Companion for CNS Disorders, 13(5). https://doi.org/10.4088/PCC.11r01161

González-Robles, A., Suso-Ribera, C., Díaz-García, A., García-Palacios, A., Castilla, D., & Botella, C. (2021). Predicting response to transdiagnostic iCBT for emotional disorders from patient and therapist involvement. *Internet Interventions*, *25*, 100420–100420. https://doi.org/10.1016/j.invent.2021.100420

Hamilton, D. E., Albright, J., Seth, M., Painter, I., Maynard, C., Hira, R. S., Sukul, D., & Gurm, H. S. (2024). Merging machine learning and patient preference: a novel tool for risk prediction of percutaneous coronary interventions. *European Heart Journal*. https://doi.org/10.1093/eurheartj/ehad836

Headley, E., Kellett, S., Bee, C., Lancashire, J., Aadahl, V., Bone, C., & Power, N. (2024). Types and mechanisms of idiographic change during guided self-help for anxiety. *Psychology and Psychotherapy, Theory, Research and Practice* (submitted).

Huibers, M. J., Cohen, Z. D., Lemmens, L. H., Arntz, A., Peeters, F. P., Cuijpers, P., & DeRubeis, R. J. (2015). Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the personalized advantage index approach. *PloS one, 10*(11), e0140771. https://doi.org/10.1371/journal.pone.0140771

IBM Corp. (2020). *IBM SPSS Statistics for Windows (Version 27.0)* [Computer software]. IBM Corp.

Ipsos, M. O. R. I. (2017). *Public views of machine learning. Findings From Public Research Engagement Conducted on Behalf of the Royal Society*. London: Ipsos MORI, The Royal Society.

Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *Journal of consulting and clinical psychology, 67*(3), 300. https://doi.org/10.1037/0022-006X.67.3.300

Kawathekar, U., S. (2023, July 26). *From Expertise to Collaboration: Understanding the Role of Patient Preferences in Clinical Decision Making*. JOSPT. https://www.jospt.org/do/10.2519/jospt.blog.20230726/full/#:~:text=Role%20Preferences%20in%20Clinical%20Decision,relationship%20with%20the%20healthcare%20provider.

Kellett, S., Bee, C., Smithies, J., Aadahl, V., Simmonds-Buckley, M., Power, N., Duggan-Williams, C., Fallon, N., & Delgadillo, J. (2023). Cognitive–behavioural versus cognitive–analytic guided self-help for mild-to-moderate anxiety: a pragmatic, randomised patient preference trial. *The British Journal of Psychiatry, 223*(3), 438-445. https://doi.org/10.1192/bjp.2023.78

Kroenke K., Spitzer R. L. & Williams J. B. W. (2001). The PHQ-9: Validity of a brief depression measure. *Journal of General Internal Medicine, 16*, 606–613.

Kroenke, K., Spitzer, R. L., Williams, J. B., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Annals of internal medicine, 146*(5), 317-325.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, *36*(11), 1–13. https://doi.org/10.18637/jss.v036.i11.

Lawn, S., Huang, N., Zabeen, S., Smith, D., Battersby, M., Redpath, P., Glover, F., Venning, A., Cameron, J., & Fairweather-Schmidt, K. (2019). Outcomes of telephone-delivered low-intensity cognitive behaviour therapy (LiCBT) to community dwelling Australians with a recent hospital admission due to depression or anxiety: MindStep. *BMC Psychiatry*, *19*(1), 2–2. https://doi.org/10.1186/s12888-018-1987-1

Lemma, A., & Fonagy, P. (2013). Feasibility study of a psychodynamic online group intervention for depression. *Psychoanalytic Psychology*, *30*(3), 367–380. https://doi.org/10.1037/a0033239

Lozano, R., Naghavi, M., Lim, S, Aboyans, V., Abraham, J., Adair, T., Ahn, S. Y., Al Mazroa, M. A., Anderson, R. H., Anderson, L. M., Andrews, K. G., Baddour, L. M., Barker-Collo, S., Bartels, D. H., Bell, M. L., Bennett, D., Bikbov, B., Boufoud, S., Burch, M., Colquhoun, S. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet (British Edition), 380*(9859), 2095–2128. https://doi.org/10.1016/S0140-6736(12)61728-0

Lutz, W., Schwartz, B., Hofmann, S. G., Fisher, A. J., Husen, K., & Rubel, J. A. (2018). Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: A methodological proof-of-concept study. *Scientific reports*, *8*(1), 7819. https://doi.org/10.1038/s41598-018-25953-0

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological bulletin, 111*(3), 490. https://doi.org/10.1037/0033-2909.111.3.490

Mathiasen, K., Riper, H., Andersen, T. E., & Roessler, K. K. (2018). Guided internet-based cognitive behavioral therapy for adult depression and anxiety in routine secondary care: Observational study. *Journal of Medical Internet Research*, *20*(11), e10927–e10927. https://doi.org/10.2196/10927

Meadows, J., & Kellett, S. (2017). Development and evaluation of Cognitive Analytic Guided Self-Help (CAT-SH) for use in IAPT services. *Behavioural and cognitive psychotherapy, 45*(3), 266-284. https://doi.org/10.1017/S1352465816000485

Miles, J., & Shelvin, M. (2001). *Applying Regression.* Sage Publications.

Milosevic, I., Levy, H.C., Alcolado, G.M. & Radomsky, A.S. (2015). The treatment acceptability/adherence scale: moving beyond the assessment of treatment effectiveness. *Cognitive Behaviour Therapist, 44*, 456-469. https://doi.org/10.1080/16506073.2015.1053407

Moggia, D., Saxon, D., Lutz, W., Hardy, G. E., & Barkham, M. (2023). Applying precision methods to treatment selection for moderate/severe depression in person-centered experiential therapy or cognitive behavioral therapy. *Psychotherapy Research*, 1-16. https://doi.org/10.1080/10503307.2023.2269297

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

National Collaborating Centre for Mental Health. (2018, June). *The Improving Access to Psychological Therapies Manual.* https://www.rcpsych.ac.uk/docs/default-source/improving-care/nccmh/iapt/the-iapt-manual-v6.pdf?sfvrsn=79ece7e5_2

National Institute of Health and Clinical Excellence (NICE). (2009, October 28). *Clinical guideline 90. depression: the treatment and management of depression in adults*. http://www.nice.org.uk/nicemedia/pdf/CG90NICEguideline.pdf

National Institute of Health and Clinical Excellence (NICE). (2011). *Common mental health disorders: identification and pathways to care.* Clinical Guideline 123. London, UK: National Institute for Health and Clinical Excellence.

National Institute for Health and Clinical Excellence. (NICE). (2014) *Cardiovascular disease: risk assessment and reduction, including lipid modification*. London, UK: National Institute for Health and Care Excellence.

Priemer, M., & Talbot, F. (2013). CBT guided self-help compares favourably to gold standard therapist-administered CBT and shows unique benefits over traditional treatment. *Behaviour Change*, *30*(4), 227–240. https://doi.org/10.1017/bec.2013.22

R Core Team. (2021, n.d.). *R: A language and environment for statistical computing* [computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

RStudio Team. (2021, n.d.). *Rstudio: Integrated development for R* [computer software]. RStudio, Inc. http://www.rstudio.com/

Richards, D., & Whyte, M. (2011). *Reachout. national programme student materials to support the delivery of training for psychological wellbeing practitioners delivering low intensity interventions*. Rethink Publishing

Riley, R. D., Snell, K. I., Ensor, J., Burke, D. L., Harrell Jr, F. E., Moons, K. G., & Collins, G. S. (2019). Minimum sample size for developing a multivariable prediction model: part I–continuous outcomes. *Statistics in medicine, 38*(7), 1262-1275.

Rodriguez, J. D., Perez, A., & Lozano, J. A. (2009). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE transactions on pattern analysis and machine intelligence, 32*(3), 569-575. https://doi.org/10.1109/TPAMI.2009.187

Ryle, A., & Golynkina, K. (2000). Effectiveness of time-limited cognitive analytic therapy of borderline personality disorder: Factors associated with outcome. *British Journal of Medical Psychology*, *73*(2), 197-210. https://doi.org/10.1348/000711200160426

Ryle, A., & Kerr, I. (2002). *Introducing cognitive analytic therapy: Principles and practice*. Wiley.

Salomonsson, S., Santoft, F., Lindsäter, E., Ejeby, K., Ingvar, M., Öst, L. G., Lekander, M., Ljótsson, B. & Hedman-Lagerlöf, E. (2020). Predictors of outcome in guided self-help cognitive behavioural therapy for common mental disorders in primary care. *Cognitive Behaviour Therapy, 49*(6), 455-474. https://doi.org/10.1080/16506073.2019.1669701

Schwartz, B., Cohen, Z. D., Rubel, J. A., Zimmermann, D., Wittmann, W. W., & Lutz, W. (2021). Personalized treatment selection in routine care: Integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. *Psychotherapy Research*, *31*(1), 33-51. https://doi.org/10.1080/10503307.2020.1769219

Sheehan, D.V., Lecrubier, Y., Harnett-Sheehan, K., Amorimm P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar G. (1998). The mini-international neuropsychiatric interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview. *Journal of Clinical Psychiatry, 59*, 22-33.

Spitzer, R.L., Kroenke, K., Williams, J.B., Löwe, B. (2006) A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine, 166*, 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics, 28(*1), 112–118. https://doi.org/10.1093/bioinformatics/btr597

Swift, J. K., & Callahan, J. L. (2009). The impact of client treatment preferences on outcome: A meta-analysis. *Journal of clinical psychology, 65*(4), 368-381. https://doi.org/10.1002/jclp.20553

Tang, C., Garreau, D., & von Luxburg, U. (2018). When do random forests fail? *Advances in neural information processing systems*, *31*.

Torgerson, D., & Sibbald B. (1998). Understanding controlled trials: what is a patient preference trial? *British Medical Journal, 316*, 360-365. https://doi.org/10.1136/bmj.316.7128.360

Turpin, G. (2010). *Good practice guidelines on the use of self-help materials within IAPT services*. IAPT document.

Tzouramanis, P., Adamopoulou, A., Bozikas, V., Voikli, M., Zagora, C., Lombtzianidou, M., Mamouzelos, E., & Garyfallos, G. (2010). Evaluation of cognitive-analytic therapy (CAT) outcome in patients with panic disorder. *Psychiatrike, 21*, 287–293.

University College London (UCL) (2014, n.d.). PWP best practice guide. https://www.ucl.ac.uk/pals/sites/pals/files/pwp_training_review_appendix_8_-_pwp_best_practice_guide.pdf

Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet*, *370*(9596), 1453-1457. https://doi.org/10.1016/S0140-6736(07)61602-X

Wray, A., Kellett, S., Bee, C., Smithies, J., Aadahl, V., Simmonds-Buckley, M., & McElhatton, C. (2022). The acceptability of cognitive analytic guided self-help in an Improving Access to Psychological Therapies service. *Behavioural and Cognitive Psychotherapy*, *50*(5), 493–507. https://doi.org/10.1017/S1352465822000194

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Appendices

Appendix	Title	Page
A	Ethical Approval Letter	132
B	TRIPOD Checklist	133
C	Information gathered from Kellett et al. (2023)	135
D	Supplementary Data Analysis Information	136
E	ANOVA Assumption Tests Results	138

Appendix A. Ethical Approval Letter

Downloaded: 14/05/2024
Approved: 07/12/2022

Caroline Wojnarowski
Registration number: 210154825
Psychology
Programme: Doctorate in Clinical Psychology

Dear Caroline

PROJECT TITLE: Predicting Optimal Treatment Outcomes Using the Personalised Advantage Index for Patients Receiving Cognitive Analytic Guided Self-Help (CAT-GSH) and Cognitive Behavioural Guided Self-Help (CBT-GSH) for Anxiety
APPLICATION: Reference Number 050718

This letter confirms that you have signed a University Research Ethics Committee-approved self-declaration to confirm that your research will involve only existing research, clinical or other data that has been robustly anonymised. You have judged it to be unlikely that this project would cause offence to those who originally provided the data, should they become aware of it.

As such, on behalf of the University Research Ethics Committee, I can confirm that your project can go ahead on the basis of this self-declaration.

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since full ethical review may be required.

Yours sincerely

Department Of Psychology Research Ethics Committee
Departmental Ethics Administrator

Appendix B. TRIPOD Checklist

TRIPOD+AI

Version: 11-January-2024

Section/Topic	Item	Development / evaluation[1]	Checklist item	Reported on page
TITLE				
Title	1	D;E	Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted	Page 76
ABSTRACT				
Abstract	2	D;E	See TRIPOD+AI for Abstracts checklist	Page 77-79
INTRODUCTION				
Background	3a	D;E	Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, including references to existing models	Page 80-81
	3b	D;E	Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users (e.g., healthcare professionals, patients, public)	Page 80-83
	3c	D;E	Describe any known health inequalities between sociodemographic groups	N/A
Objectives	4	D;E	Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	Page 84
METHODS				
Data	5a	D;E	Describe the sources of data separately for the development and evaluation datasets (e.g., randomised trial, cohort, routine care or registry data), the rationale for using these data, and representativeness of the data	Page 85-88
	5b	D;E	Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up	Page 86
Participants	6a	D;E	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including the number and location of centres	Page 85-86
	6b	D;E	Describe the eligibility criteria for study participants	Page 86-87
	6c	D;E	Give details of any treatments received, and how they were handled during model development or evaluation, if relevant	Page 88-89
Data preparation	7	D;E	Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups	Page 95
Outcome	8a	D;E	Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale for choosing this outcome, and whether the method of outcome assessment is consistent across sociodemographic groups	Page 89-91
	8b	D;E	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	N/A
	8c	D;E	Report any actions to blind assessment of the outcome to be predicted	N/A
Predictors	9a	D	Describe the choice of initial predictors (e.g., literature, previous models, all available predictors) and any pre-selection of predictors before model building	Page 95-97
	9b	D;E	Clearly define all predictors, including how and when they were measured (and any actions to blind assessment of predictors for the outcome and other predictors)	Page 95-97 Appendix C
	9c	D;E	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	N/A
Sample size	10	D;E	Explain how the study size was arrived at (separately for development and evaluation), and justify that the study size was sufficient to answer the research question. Include details of any sample size calculation	Page 94
Missing data	11	D;E	Describe how missing data were handled. Provide reasons for omitting any data	Page 95
Analytical methods	12a	D	Describe how the data were used (e.g., for development and evaluation of model performance) in the analysis, including whether the data were partitioned, considering any sample size requirements	Page 95-97
	12b	D	Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardisation).	Page 95-97
	12c	D	Specify the type of model, rationale[2], all model-building steps, including any hyperparameter tuning, and method for internal validation	Page 95-97
	12d	D;E	Describe if and how any heterogeneity in estimates of model parameter values and model performance was handled and quantified across clusters (e.g., hospitals, countries). See TRIPOD-Cluster for additional considerations[3]	Page 95-97
	12e	D;E	Specify all measures and plots used (and their rationale) to evaluate model performance (e.g., discrimination, calibration, clinical utility) and, if relevant, to compare multiple models	Page 95-97
	12f	E	Describe any model updating (e.g., recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings	N/A
	12g	E	For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface)	Page 95-97
Class imbalance	13	D;E	If class imbalance methods were used, state why and how this was done, and any subsequent methods to recalibrate the model or the model predictions	N/A
Fairness	14	D;E	Describe any approaches that were used to address model fairness and their rationale	N/A
Model output	15	D	Specify the output of the prediction model (e.g., probabilities, classification). Provide details and rationale for any classification and how the thresholds were identified	Page 95-97

Training versus evaluation	16	D;E	Identify any differences between the development and evaluation data in healthcare setting, eligibility criteria, outcome, and predictors	N/A
Ethical approval	17	D;E	Name the institutional research board or ethics committee that approved the study and describe the participant-informed consent or the ethics committee waiver of informed consent	Page 85
OPEN SCIENCE				
Funding	18a	D;E	Give the source of funding and the role of the funders for the present study	Page ii
Conflicts of interest	18b	D;E	Declare any conflicts of interest and financial disclosures for all authors	Page ii
Protocol	18c	D;E	Indicate where the study protocol can be accessed or state that a protocol was not prepared	N/A
Registration	18d	D;E	Provide registration information for the study, including register name and registration number, or state that the study was not registered	N/A
Data sharing	18e	D;E	Provide details of the availability of the study data	Page ii
Code sharing	18f	D;E	Provide details of the availability of the analytical code[4]	Page ii
PATIENT & PUBLIC INVOLVEMENT				
Patient & Public Involvement	19	D;E	Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement.	Page 85
RESULTS				
Participants	20a	D;E	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Page 91-92
	20b	D;E	Report the characteristics overall and, where applicable, for each data source or setting, including the key dates, key predictors (including demographics), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data. A table may be helpful. Report any differences across key demographic groups.	Page 92-93
	20c	E	For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome).	N/A
Model development	21	D;E	Specify the number of participants and outcome events in each analysis (e.g., for model development, hyperparameter tuning, model evaluation)	Page 91-92
Model specification	22	D	Provide details of the full prediction model (e.g., formula, code, object, application programming interface) to allow predictions in new individuals and to enable third-party evaluation and implementation, including any restrictions to access or re-use (e.g., freely available, proprietary)[5]	Page 99-103
Model performance	23a	D;E	Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid presentation.	Page 102-103
	23b	D;E	If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD Cluster for additional details[3].	N/A
Model updating	24	E	Report the results from any model updating, including the updated model and subsequent performance	N/A
DISCUSSION				
Interpretation	25	D;E	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies	Page 109-113
Limitations	26	D;E	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalizability	Page 115-116
Usability of the model in the context of current care	27a	D	Describe how poor quality or unavailable input data (e.g., predictor values) should be assessed and handled when implementing the prediction model	N/A
	27b	D	Specify whether users will be required to interact in the handling of the input data or use of the model, and what level of expertise is required of users	N/A
	27c	D;E	Discuss any next steps for future research, with a specific view to applicability and generalizability of the model	Page 116-117

Appendix C. Demographic & treatment history data collected within initial trial and suitable for analysis (Kellett et al., 2023)

Variable	**Method of Measurement (Transformed)**
Age	Continuous
Ethnicity	White British or Minoritised Ethnicity
Employment Status	In Employment or Not Employed
Sexual Orientation	Heterosexual or Not Heterosexual
Gender	Male or Female
Identification of Long-Term Condition (LTC)	Self-reported LTC or no LTC
Veteran Status	Veteran or Not a Veteran
Perinatal Status	Pregnant or Not Pregnant
Psychotropic Medication	Not Taking Medication or Taking Medication
Indices of Multiple Deprivation (IMD)	Continuous
GAD-7 baseline severity	Continuous
PHQ-9 baseline severity	Continuous
WSAS baseline severity	Continuous
BAI baseline severity	Continuous
Previous Treatment	Previous Treatment Disclosed or No Previous Treatment
Previous CAT	Had CAT or Not Had CAT
Previous CBT	Had CBT or Not Had CBT
Allocation Choice	Treatment Preference or Randomisation

Appendix D. Supplementary Data Analysis Information

Random Forest (Boruta method)

Random forest builds multiple decision trees using a bagging method (a combination variable and bootstrapping samples; Breiman, 2001). A variable is selected to create a node, this is then analysed to establish threshold values to split observations (a tree branch). A new variable is then selected for each branch with remaining observations split through recursive partitioning. Various decision trees are build based on bootstrapped datasets and a random selection of variables. Data not included in the bootstrapped datasets are the out-of-the-bag (OOB) cases. This prevents overfitting of the model to the current dataset. The use of random selection reduces the overriding effects of stronger predictive variables as it uses information from weaker predictors. The importance of each predictor is expressed as a “mean decrease in accuracy” which presents how much a variable increases or decreases the models accuracy. Such approaches handle multicollinearity well and can be well suited to multivariable prediction models.

One of the limitations of standard random forest variable selection is that it provides a rank-ordered list of variable importance but does not remove any variables. Researchers are required to set criteria for what threshold will be used to retain predictors in the model and this has been applied in different ways, including retaining only predictors reaching 90% importance in the model (Schwartz et al. 2021) or positive scores of mean decrease in prediction accuracy (Moggia et al. 2023). The Boruta method aims to address this issue by extending the random forest approach to include an embedded variable selection criterion. Shadow variables (one continuous and one categorical) are created based on the distributions of other variables in the

dataset and included in the model as a ‘noise’ variable (i.e., have no actual predictive power). Only variables which are ranked higher than one (tentative inclusion) or both (confirmed inclusion) shadow variables are deemed to have reliable predictive power over and above noise and are retained.

Elastic Net Regularisation

This method combines both the Least Absolute Shrinkage and Selection Operator (LASSO) and Ridge penalties. LASSO penalisation shrinks coefficients with no predictive value to zero and Ridge penalisation shrinks coefficients with less predictive value towards zero, therefore not excluding any predictors. Elastic Net reduces overfitting of the model, increases prediction accuracy and model generalisability (Zou & Hastie, 2005). Multiple potential models were tested using 10-fold cross-validation (Rodriguez et al., 2009), this enables selection of predictors to be less influenced by extreme outliers (Breckler, 1990; MacCallum et al., 1992). The Elastic net model which had the best accuracy and lowest prediction error (i.e., best combination of LASSO and Ridge penalties was chosen). The best fitting model was based on the one standard error rule, the most parsimonious model whose error was no more than one standard error above the error of the best model was chosen. As a result, this chosen model was more likely to generalise to other data and not be too specific (overfitted) to the current data, which might have been the case if the “best model” was chosen.

Appendix E. ANOVA Assumption Testing Results

Test of Normality Full Sample (*significance at p<0.05)

Time	Group	Shapiro-Wilk Statistic	df	Significance
Baseline	*Optimal*	.970	131	.006*
	Non-optimal	.941	78	.001*
Post-treatment	*Optimal*	.955	131	<.001*
	Non-optimal	.957	78	.010*
Follow-up	*Optimal*	.949	131	<.001*
	Non-optimal	.967	78	.041*

Test of Normality Subgroup (*significance at p<0.05)

Time	Group	Shapiro-Wilk Statistic	df	Significance
Baseline	*Optimal*	.968	26	.575
	Non-optimal	.829	11	.022*
Post-treatment	*Optimal*	.877	26	.005*
	Non-optimal	.943	11	.554
Follow-up	*Optimal*	.933	26	.092
	Non-optimal	.909	11	.235

Test of Homogeneity of Variance Full Sample

Time	Levene Statistic	df	Significance
Baseline	.521	1, 207	.471
Post-treatment	.359	1, 207	.550
Follow-up	.869	1, 207	.352

Test of Homogeneity of Variance Subgroup

Time	Levene Statistic	df	Significance
Baseline	1.676	1, 35	.204
Post-treatment	.055	1, 35	.816
Follow-up	.254	1, 35	.617