Exploring the Prevalence and Predictors of Deterioration After Low Intensity Cognitive Behavioural Therapy

Itzia Perez Morales

A thesis submitted in partial fulfilment of the requirements for the award of Doctor of Clinical Psychology

Clinical Psychology Unit, Department of Psychology
The University of Sheffield

November 2024

Declaration

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctorate in Clinical Psychology in the University of Sheffield. This thesis is my own work and it has not been used to obtain another degree or qualification or in any other institution.

Structure and Word Count

Literature Review

Excluding references and tables: 6,251

Including references and tables: 12,543

Empirical Study

Excluding references and tables: 6,938

Including references and tables: 10,755

Total Word Count

Excluding references and tables: 13,189

Including references and tables: 23,298

Lay Summary

This research investigates clinical deterioration in adults following low-intensity cognitive behavioural therapy (LICBT), a brief psychological intervention for mild to moderate common mental health problems. LICBT is widely implemented in primary care mental health services in England. While its effectiveness is well-documented, its potential for negative outcomes (e.g., a worsening of mental health symptoms) remains underexplored.

The first chapter of this thesis presents a systematic review examining deterioration rates and risk factors for deterioration after LICBT. Forty-four studies were initially included, with only seven reporting deterioration rates (15.9%). The vast majority of studies were internet-based interventions targeting depression and anxiety in European community settings. Deterioration rates ranged from 0% to 6% using a wide variety of clinical outcome measures and statistical methods to calculate deterioration rates. No studies conducted subgroup analyses to identify potential risk factors for deterioration.

The second chapter details an empirical study using data from patients who accessed LICBT in NHS Talking Therapies Services in England. Statistical methods were used to identify two types of deterioration: those likely to be caused by the treatment itself (iatrogenic harm) and those related to natural illness progression. The estimated deterioration rates were around 16% of patients who accessed LICBT (11% iatrogenic and 5% chronic). Common characteristics identified in participants experiencing deterioration include having more than one co-occurring mental health problem, severe impairment to daily activities, unemployment, and living in socioeconomically deprived neighbourhoods. Although further research is needed, these findings suggests that deterioration in LICBT may be more common than previously thought.

Acknowledgements

This thesis would not have been completed without the support from my research supervisor Dr Jaime Delgadillo. He was instrumental in developing my research skills, especially in regard to machine learning and statistics. Thank you for lending me your expertise these past three years, thank you for your faith, your patience, and your guidance.

I would also like to thank my family as I wouldn't be here without them. To my mum, who has been an incredible source of support and inspiration in the most challenging and rewarding moments. To my grandma Bella, who will always be by my side reminding me that there is always something worth fighting for, parties to attend, people to see, and joy to be had. To my aunt Monica, who showed me the wonders of studying abroad and shared my passion for books, travel, science, and wonder. To my little brother Daniel, someone I love dearly and wish to motivate as much as he inspires me. I also have to include my partner James, my only family in this continent, you have no idea how much you have improved my life, you have helped me accomplish so many ambitions and keep encouraging to achieve even more and bigger things, your support helped me finish this thesis and this degree. Shout out to my best friends in Mexico, Dana and Mau, and those in the UK, Nailah, Emma, Alex, and EA, you kept me sane, gave me strength, and cheered me on throughout this journey. Thank you all.

Finally, I want to show my appreciation and gratitude to my clinical tutor, Dr Kate Rayner. Kate, you never failed to give me advice, support, and kindness when I needed it the most.

Table of Contents

Chapter One: Literature Review

Deterioration after Low-Intensity CBT:

A Systematic Review

Abstract

Background: Low-intensity cognitive behavioural therapy (LICBT) is a widely used psychological intervention for mild to moderate common mental health conditions. While its effectiveness is well-documented, deterioration rates have not been systematically explored.

Methods: A systematic search investigating deterioration in adults following LICBT for common mental health problems was conducted across four electronic databases (OVID, SCOPUS, EBSCO, and Web of Science). All eligible studies were included in a narrative synthesis with a focus on those reporting deterioration rates.

Results: Of the 44 eligible studies, most of them were randomised controlled trials (90.9%) conducted in Europe (75%). Most studies were internet-based interventions (70.5%), targeted depression (79.5%), and delivered in community settings (63.6%). Only 7 studies (15.9%) reported deterioration rates using ITT analyses and RCI criteria. Deterioration rates ranged from 0% to 5.5% post-treatment and between 3% and 6% at follow-up. Internet-based interventions reported rates between 0% and 5.5% post-treatment and between 5.5% and 6% at follow-up; group interventions 3% post-treatment and follow-up; and GSH 3% post-treatment and 6% at follow-up. No sociodemographic or clinical characteristics were associated with deterioration due to a lack of subgroup analyses in all studies.

Conclusions: This review highlights a significant gap in reporting deterioration rates following LICBT. While those found were low, the absence of standardised data collection, reporting frameworks, and subgroup analyses limits the precision, reliability and generalizability of the findings. Future research should calculate and report deterioration prevalence rates. Subgroup analyses to identify deterioration predictors should also be considered when possible.

Keywords: Low intensity cognitive behavioural therapy, deterioration, systematic review

Practitioner Points:

- Mental health professionals should be aware of about possible adverse effects in general psychotherapeutic work and particular to the psychological model or intervention used. This includes deterioration rates and associated predictors and moderators.
- Information regarding possible adverse outcomes, such as non-response and clinical deterioration, should be discussed with patients to improve their understanding on the risks and benefits associated. This can include a rationale for the routine use of psychometric outcome measures and enable risk-benefit informed consent.
- Clinical practitioners should routinely use validated psychometric measures at different timepoints, at least before, during and after the intervention, to promptly identify treatment non-response or deterioration.
- Clinicians should remain attentive to commonly recognised risk factors associated with clinical deterioration in psychotherapy, such as high baseline symptom severity, psychiatric comorbidity, and socio-economic challenges.
- Clinical supervisors and service managers should also be aware of possible adverse outcomes in psychotherapy and ensure that practitioners collect routine outcome data to identify reliable change, improvement, and deterioration.

Introduction

Psychotherapy is widely regarded as a safe and effective treatment for mental health conditions, yet there is a growing body of research that emphasises the importance of examining its potential for harm (Schermuly-Haupt, Linden, & Rish, 2018). While few specific psychological interventions have been found to be harmful (Klatte, 2023), adverse effects have been found across psychotherapeutic modalities in a minority of cases (Lambert, 2013). Yet, the risks associated with psychotherapy are often underreported and underexplored (Lilienfeld, 2007).

Low-intensity cognitive behavioural therapy (LICBT) has emerged as a key intervention for common mental health disorders in primary care settings. In the United Kingdom, the National Health Service (NHS) provides LICBT through NHS Talking Therapies, previously known as Improving Access to Psychological Therapies (IAPT) services. These services are commissioned to treat mild to moderate common mental health conditions following a stepped care model at step 2 and step 3, where step 2 includes LICBT in the form of psychoeducational groups, pure or guided self-help (GSH), and internet-based CBT (Clark, 2011).

LICBT are brief CBT-based psychological interventions aimed towards mild to moderate depression, generalised anxiety disorder (GAD), panic disorder, specific phobias, and obsessive-compulsive disorder (OCD) following National Institute for Health and Care Excellence (NICE) guidelines (National Collaborating Centre for Mental Health, 2018). LICBT interventions are highly structured based on protocols and CBT-based psychoeducational workbooks lasting ≤ 8 sessions with support from a psychological wellbeing practitioner (PWP) trained in LICBT (Bennett-Levy et al., 2010).

While the efficacy of LICBT in improving symptoms of depression and anxiety is well-documented (e.g. Andrews et al. 2010; Bower et al., 2013; Cuijpers et al., 2010; Powell et al., 2024), prevalence rates, predictors, and moderators of symptomatic deterioration in LICBT have not been thoroughly investigated. Some studies have found limited post-treatment effectiveness with high relapse rates at 6, 12 and 18 months follow-up in up to 59% of cases (Ali et al., 2017; Coull & Morris, 2011; Delgadillo et al., 2018) and published data from routine clinical practice in NHS Talking Therapies shows that approximately 6% of patients deteriorate after LICBT in primary care at a national level in the United Kingdom (National Health Service [NHS], 2023), yet to date there are no systematic reviews on the topic of LICBT deterioration.

The underreporting and lack of exploration of deterioration in the literature is not particular to LICBT, partly due to a lack of consensus regarding the definition, classification, assessment, and reporting guidelines of deterioration in clinical research (Rozental et al., 2018). The term deterioration can be defined under treatment failure, treatment non-response, and relapse (Lambert, 2013), and even its most common definition as increased symptom severity can vary according to the domain of interest and outcome assessment methodologies (Lazar, 2017).

Recognizing the potential harms and hidden costs of any treatment is crucial for patients, clinicians, managers, and policymakers to make informed decisions (Ernst, 2001). This is reflected in the Consolidated Standards for Reporting Trials (CONSORT) guidelines, which emphasize the importance of monitoring and reporting adverse effects in psychological research (Ioannidis et al., 2004). To enable this, attempts to define deterioration have been made, including the use of statistical methods to identify significant clinical deterioration as seen in Intention-To-Treat (ITT) and Reliable Change Index (RCI; Jacobson and Truax, 1991).

ITT analyses include data from all randomized participants and taking into account their assigned group in clinical trials to provide an unbiased estimate of effectiveness and deterioration rates (McCory, 2017). While the RCI uses psychometric data to determine reliable change, improvement, or deterioration using participant's standardized scores, and the measure's standard error (Jacobson and Truax, 1991). These analyses are used in the literature to identify clinical deterioration, defined as worsening symptom progression (Linden, 2012).

Another gap in the literature is the lack of consensus and limited high-quality research on treatment effect moderators and predictors of adverse effects and deterioration, including patient, therapist, and the intervention's characteristics. This not only makes challenging the identification of harmful psychotherapeutic interventions, but also the identification of which individuals could find these interventions detrimental (Teachman, White, & Lilienfeld, 2021).

Moderators and predictors are baseline variables that have been found to be associated with treatment outcomes through statistical analyses, where moderators provide information regarding the conditions under which interventions can be effective or detrimental, and predictors provide information on which variables have a treatment effect without interacting with the treatment itself (Sextl-Plotz et al., 2024). Moderators and predictors can include patient sociodemographic and clinical characteristics (e.g. age and diagnosis), or characteristics related to the intervention provided (e.g. delivery method).

Information on moderators and predictors can be investigated through sub-group analyses, although such research requires high-quality qualitative studies, systematic reviews, or meta-analyses with significant statistical power, including large sample sizes, diverse populations, appropriate outcome measures, and rigorous methodologies. While some psychotherapeutic outcome predictors and moderators have been identified in the literature, the

vast majority are related to positive clinical outcomes such as improvement and recovery, instead of adverse effects such as deterioration (Teachman, White, & Lilienfeld, 2021).

Aims

This systematic review addresses these critical gaps by searching and synthesizing evidence on clinical deterioration following LICBT interventions (psychoeducational groups, pure or guided self-help, and internet-based) for mild to moderate common mental health disorders (depression, GAD, panic disorder, specific phobias, and OCD) in adults (over 18 years-old). This includes studies reporting and investigating clinical deterioration, defined as worsening symptom progression using ITT or RCI to calculate prevalence rates and subgroup analyses that aim to identify deterioration predictors or moderators.

Research Questions

The research questions that guided this study are as follows:

1) What is the percentage of studies that report deterioration rates following LICBT interventions for common mental health disorders?

2) What are the prevalence rates found, and those that can be calculated with available data?

3) What are the sociodemographic and clinical characteristics of cases classed as having deteriorated?

Methods

Study Protocol Registration

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Page et al., 2021) guidelines and was pre-registered in the International Prospective Register of Systematic Review (PROSPERO) database on October 18th, 2024 (protocol identification reference: CRD42024601920).

The study protocol pre-registration had to be amended to include more details in regard to the search restrictions (publication dates and languages), and on the data synthesis strategy.

Search Strategy

Systematic searches were carried out in the following electronic databases: OVID (Medline and PsycINFO), SCOPUS, EBSCO (CINAHL), and Web of Science. The search was conducted on October 19th, 2024, following the pre-registered protocol's search strategy.

Search term variations on the topic of interest (deterioration), intervention (LICBT), condition of interest (diagnoses), population (adults), type of study (cohort and RCT), and clinical setting (primary care, outpatient, and community) were included, as seen in Table 1. The search terms were informed by previous systematic reviews on deterioration (Lazar, 2017; Thorpe, 2012), LICBT interventions (Cremers et al., 2019; Cuijpers, et al., 2010; Powell et al., 2024), and medical subject headings (MeSH) terminology by the National Institute of Health.

All retrieved records were screened for suitability using the protocol's inclusion and exclusion criteria, initially through titles, abstracts, and indexes, followed by full-text screening. Forward and backward citation searching of papers matching the eligibility criteria were conducted by hand to identify any further studies missing in the original search. Reasons for exclusion during the screening stage were recorded and tabulated.

Table 1.

Search Terminology

Term of Interest	Search Terms Variations
Clinical Outcome	(deteriorat*) OR (symptomatic deteriorat*) OR (clinical deteriorat*) OR (negative outcom*) OR (negative respon*) OR (harmful treatment) OR (iatrogenic treatment effect*) OR (treatment failure)

Intervention	(low intensity cognitive behavio* therap*) OR (low intensity CBT) OR (LI CBT) OR (LICBT) OR (step 2 cognitive behavio* therap*) OR (Step 2 CBT) OR (guided self-help) OR (GSH) OR (assisted self-help) OR (facilitated self-help) OR (supervised self-help) OR (supported self-help) OR (internet based cognitive behavio* therap*) OR (internet based CBT) OR (internet delivered cognitive behavio* therap*) OR (internet delivered CBT) OR (internet cognitive behavio* therap*) OR (internet CBT) OR (iCBT) OR (computeri* cognitive behavio* therap*) OR (computeri* CBT) OR (cCBT) OR (web-based cognitive behavio* therap*) OR (web-based CBT) OR (digitally enabled therap*) OR (group cognitive behavio* therap*) OR (group CBT) OR (cognitive behavio* therap* group) OR (CBT group) OR (cognitive behavio* therap* based group) OR (CBT based group) OR (cognitive behavio* therap* group workshop) OR (CBT group workshop) OR (cognitive behavio* therap* workshop) OR (CBT workshop) OR (cognitive behavio* therap* psycho educatio* group) OR (CBT psycho educatio* group) OR (CBT psycho educatio*) OR (psycho educatio*) OR (cognitive behavio* therap*) OR (CBT) OR (behavio* activation) OR (problem solving) OR (exposure and response prevention) OR (ERP)
Target Conditions	(affective disorde*) OR (mood disorde*) OR (anxiety disorde*) OR (depressi*) OR (generali* anxiety disorder) OR (GAD) OR (worry) OR (panic) OR (phobi*) OR (obsessive compulsive disorder) OR (OCD) OR (anxiety)
Population	(adults) OR (middle age) OR (elderly) OR (older adult)

Research Design	(cohort) OR (randomi* clinical trial) OR (randomi*) OR (clinical trial) OR (RCT)
Setting	(primary care) OR (outpatient) OR (community)

Inclusion and Exclusion Criteria for Eligibility

Studies were eligible for inclusion if they pertained to adults aged 18 years or older who had received LICBT for mild or moderate common mental health disorders (depression, GAD, phobias, panic disorder, and OCD), as defined by NICE guidelines (National Institute for Health and Care Excellence [NICE], 2022; NICE 2014).

Interventions included any form of LICBT, including pure and guided self-help, psychoeducational groups or workshops, problem-solving, behavioural activation, and exposure and response prevention (ERP), delivered through individual, group, or online formats with ≤ 9 sessions (as LICBT is normally 8 sessions, sometimes with an additional session as introduction or booster) based on CBT principles (Bennet-Levy et al., 2010).

Studies had to measure clinical outcomes related to the target condition using validated psychometric measures in at least two time points (before and after the LICBT intervention). Due to the limitations identified in the literature (lack of consensus on the definition, classification, assessment, and reporting of deterioration, and the underreporting of it), studies targeting a specific anxiety disorder were to be included if they had a relevant anxiety outcome measure instead of a specific one (i.e. specific to phobias, panic, or OCD). Special attention will be provided to studies reporting deterioration, defined as worsening symptom progression, using ITT or RCI to calculate prevalence rates and subgroup analyses.

Eligible study designs were peer reviewed publications using randomised controlled trials (RCTs) and cohort study designs, published in English, and set in non-specialised services, mainly in primary care, outpatient, and community settings.

Studies were excluded if they involved children, adolescents, long-term, traditional and high-intensity CBT, non CBT-based interventions, severe mental health conditions or those requiring high-intensity interventions following NICE guidelines, specialised services (step 4 for severe and recurrent conditions, rare or complex presentations, or high risk presentations), purely qualitative or case study designs, and if they were not published in English.

A summary table with more detailed descriptions of the inclusion and exclusion criteria using the population, intervention, comparison, outcomes, and study (PICOS) design framework (Eriksen & Frandsen, 2018) can be found in Table 2:

Table 2.

Eligibility Criteria

	Inclusion	**Exclusion**
Population	Participants aged 18 years-old or over (adults and older adults). Received an LICBT intervention and attended ≥ 1 session.	Participants under 18 years-old (minors).
Intervention	CBT-based delivered in any format (individual, group, or online). Brief interventions with ≤ 9 sessions or modules.	Long-term, traditional or high-intensity CBT (>9 sessions). Interventions that are not primarily based on CBT.

	Targeted towards depression, GAD, specific phobias, panic disorder, OCD, or mixed anxiety and depression presentations	Interventions targeting severe mental health conditions (e.g. eating disorders, psychosis, or addictions) or conditions requiring high-intensity interventions (e.g. social anxiety, PTSD, and body dysmorphia).
Comparator	Not applicable (N/A)	Not applicable (N/A)
Outcome	Clinical outcomes must be identified using validated standardised psychometric measures for depression and/or anxiety in ≥ 2 time points (baseline and post-intervention).	Clinical outcomes measured without validated standardised psychometric measures for depression and/or anxiety in ≥ 2 time points (baseline and post-intervention).
Study	Peer reviewed studies. Randomised controlled trials (RCTs) and cohort study designs. Published in English. Non-specialised services: primary care, outpatient, and community settings.	Grey literature. Qualitative and case-study research designs. Not published in English. Specialised services (e.g. inpatient, and forensic).

Risk of Bias Assessment

A risk of bias assessment was conducted on all studies reporting deterioration to assess the validity and limitations in the research design, methodology, results, and impact using the Critical Appraisal Skills Programme (CASP) checklist for RCTs (CASP, 2023; Appendix B)

and critical appraisal CASP recommendations for studies with quasi-experimental designs (CASP, 2024) by two independent reviewers. RCTs rater inter-reliability was calculated using Cohen's Kappa (Cohen, 1960) suggesting substantial agreement (k= 0.74). The critical appraisal of quasi-experimental studies was developed independently, then discussed and agreed results were summarised in the narrative synthesis (see results section).

The CASP checklist for RCTs includes 11 questions across four sections regarding the design, methodology, results, and impact. As there is no scoring guidelines for this checklists, a point was awarded per met criterion (yes response), and the total was classified as follows: high risk (0-3), medium risk (4-8), and low risk (9-12).

Strategy for Data Extraction and Synthesis

Data extraction captured study details (authors, year of publication, country, design, and setting, sample size), participant characteristics (age), intervention details (target condition, delivery method), clinical outcome measures used (psychometric tools, reported deterioration rates, and the statistical methods used- ITT or RCI), participant's sociodemographic and clinical characteristics reported, and subgroup analyses information in a summary table.

A qualitative narrative synthesis was performed, summarising the number and percentage of studies reporting deterioration, the study's characteristics, outcome measures, statistical methods, deterioration prevalence rates, participant characteristics, subgroup analyses, and risk of bias assessments, with their respective tabulations.

If sufficient data was available with similar variables, these would have been qualitatively integrated through a random-effects meta-analysis. Effect was to be measured by comparing sociodemographic and clinical characteristics between RCI or ITT categories using chi squared for categorical variables and t-test for continuous variables. Heterogeneity subgroup analysis would have used the I^2 statistic. Subgroup or moderator analyses would have

examined high heterogeneity. Potential heterogeneity sources would have been removed and the impact on effect sizes and heterogeneity indices would have been assessed.

Due to the small number of studies reporting deterioration rates, the high levels of variability in participant characteristics and clinical outcome data collection, and the medium-high risk of bias identified, a meta-analysis was not possible to be conducted.

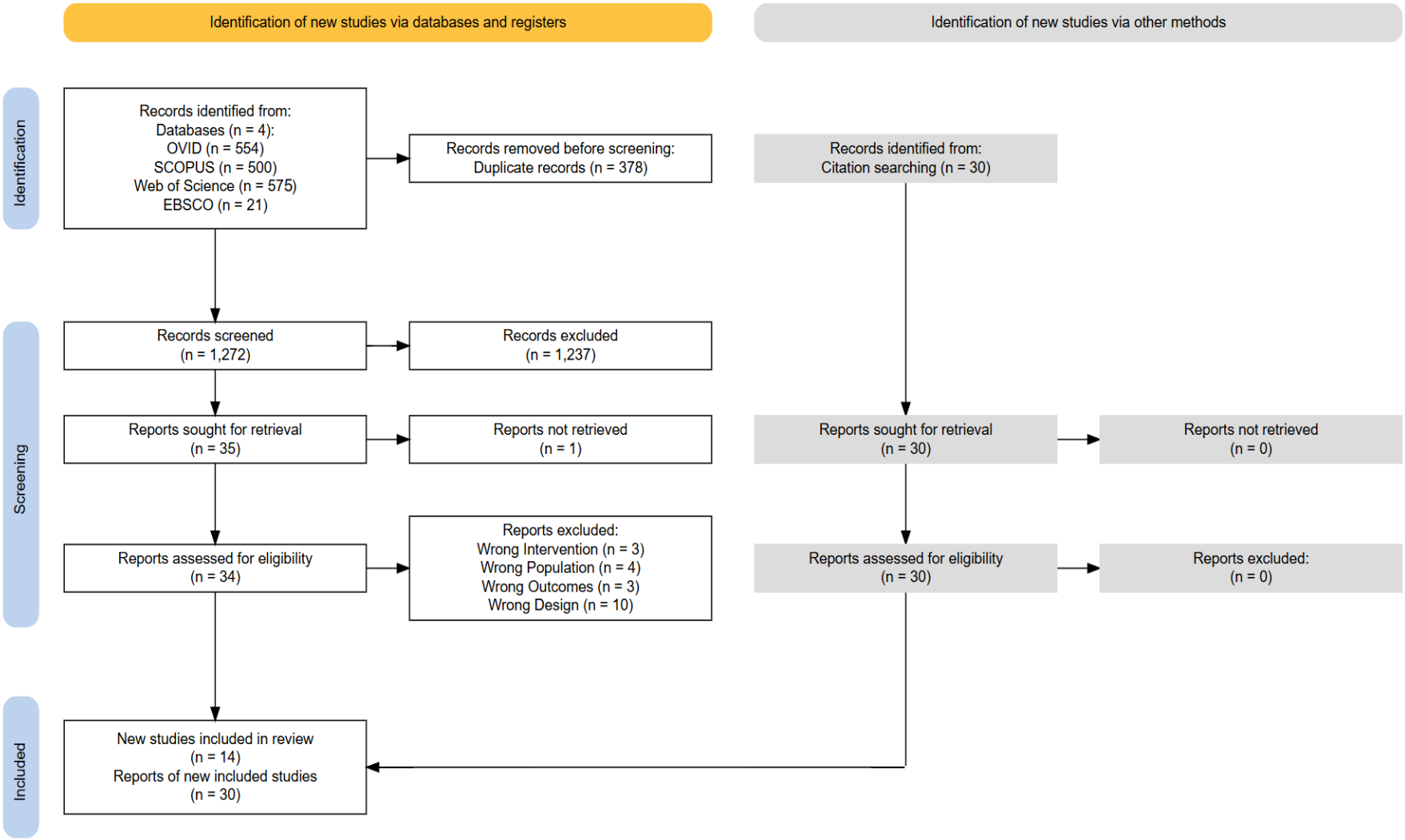
Results

Search Results

The systematic search included four databases: OVID (n= 554), SCOPUS (n= 500), EBSCO (n= 21), and Web of Science (n= 575), for a total of 1,650 studies. After de-duplication (n= 378), 1,272 articles were screened using titles and abstracts and 36 studies were retrieved for full-text screening with only 1 not retrieved. A total of 14 were eligible for inclusion and backward citation searches resulted in an additional 30 eligible studies added. Thus, a total of 44 studies were included in the review as seen in the PRISMA diagram (Page et al., 2021) in Figure 1.

Figure 1.

PRISMA Diagram



Study Characteristics

An overview of the studies characteristics is shown in Table 3. All studies were either RCTs (n= 39, 88.6%) or quasi-experimental designs (n= 5, 11.4%). All studies were published between 1999 and 2023, with 33 studies published after 2010. Most studies were carried out in Europe (n= 33, 75%), followed by North America (n= 9, 20.5%), Oceania (n= 6, 13.6%) and Asia (n= 2, 4.5%). Most of the studies were conducted in Sweden (n= 10, 22.7%) and the United Kingdom (n= 10, 22.7%), the United States (n= 8, 18.2%), Australia (n = 5, 11.4%), and a minority in other countries (n= 12, 27.2%).

The majority of interventions targeted depression (n = 24, 54.5%), while fewer addressed anxiety (n = 11, 25%) or both conditions (n = 9, 20.1%). Most studies conducted in community settings (n = 24, 54.5%) and primary care (n = 11, 25%), with a minority in healthcare (n = 6, 13.6%), and academic settings (n = 4, 9.1%). Of the delivery methods, internet-based interventions were the most common (n = 29, 65.9%), followed by pure or guided self-help (GSH) (n = 16, 36.4%) and group interventions (n = 5, 11.4%).

The studies used a wide variety of clinical outcome measures, the most common ones were the Beck Depression Inventory (BDI; Beck et al., 1962) used in 6 studies and the revised version (BDI-II; Beck, Steer, & Brown, 1996) used in 16, the Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer, & Williams, 2001) in 16, the Generalised Anxiety Disorder-7 (GAD-7; Spitzer et al., 2006) in 9, the Beck Anxiety Inventory (BAI; Beck et al., 1988) in 6, and the Clinical Outcomes in Routine Evaluation–Outcome Measure (CORE-OM; Core System Trust, 2014) in 5. For quality of life and impact scales, variations of the 36-Item Short Form Survey (SF-36; RAND, 1992) in 7, the EuroQol (EG-5D; EuroQol Group, 1990) in 6, and the Work and Social Adjustment Scale (WSAS; Mundt et al., 2002) in 5 were the most common. In total, over 60 different clinical outcome measures were used across the 44 studies, with the majority of them used in ≥ 4 studies, including 35 which were used in one study each.

Table 3.

Data Extraction Summary

Authors	Year	Country	Design and Setting	Population	N	Target Condition	Delivery Method	Outcome Measures	RCI Rates	Characteristics	Sub-Group Analyses
Andersson et al.	2005	Sweden	RCT in community setting	Adults Intervention: 36.4 Control: 36.3 Withdrawal: 35.6	117	Depression	Internet-based	BDI, BAI, MADRS-S, QoLI	None reported	Age, gender, residence, education, diagnosis, treatment history, medication use, symptom severity	None reported
Andersson et al.	2013	Sweden	RCT in community setting	Adults 42.3	213	Depression	Internet-based and Group	BDI, BAI, MADRS-S, HRSD QOLI, CGI-I	None reported.	Age, gender, employment, marital status, diagnosis, treatment history, medication use, symptom severity	None reported
Benton et al.	2016	United States	Quasi-experimental study in academic setting	Adults 21.72	72	Anxiety	Internet-based	BHM-20	None reported	Age, gender, ethnicity, education, sexual orientation.	None reported
Cano-Vindel et al.	2022	Spain	RCT in primary care	Adults 43.6	1,061	Depression and anxiety	Group	PHQ-9, GAD-7, PHQ-15, SDS, WHOQoL-Brief	Intervention: 3% post-treatment and at 12 months. Control: 14%, and 12% (ITT)	Age, gender, marital status, education, employment, income level,	None reported

										symptom severity	
Cavanagh et al.	2006	United Kingdom	Quasi-experimental study in primary care and community settings	Adults 43.6	219	Depression and anxiety	Internet-based	CORE-OM, WSAS	Completers: 5% post-treatment, 6% at 6 months (CORE-OM, RCI); 4% and 4% (WSAS, RCI) Sample: 2% (CORE-OM, ITT); 2% (WSAS, ITT)	Age, gender, main problem, symptom severity and onset	None reported
Christensen et al.	2006	Australia	RCT in community setting	Adults Range 25-44	2,794	Depression	Internet-based	Goldberg Depression Scale, Goldberg Anxiety Scale	None reported	Age, gender, residence, education, history of depression, symptom severity	None reported
Dahne et al.	2018	United States	Quasi-experimental study in community setting	Adults 24.91	11	Depression	Internet-based	BDI-II	None reported	Age, gender, ethnicity, education, occupation, income level	None reported
de Graaf et al.	2009	Netherlands	RCT in primary care	Adults Intervention: 44.3 TAU: 45.1 Intervention + TAU: 45.2	303	Depression	Internet-based	BDI-II, SCL-90, DAS-A. WSAS, SF-36	None reported	Age, gender, education, employment, relationship status, depression diagnosis	None reported
Ebert et al.	2018	Germany	RCT in community setting	Adults 44	204	Depression	Internet-based	QIDS-CR 16, HRSD, HADS-A,	None reported	Age, gender, ethnicity, marital status,	None reported

								CES-D, PSWQ, ISI		employment, education, income level, medication use, symptom severity	
Egede et al.	2015	United States	RCT in community setting	Older Adults 63.9	221	Depression	GSH	GDS, BDI	None reported	Age, gender, ethnicity, marital status, education, employment, income level, health status, insurance	None reported
Forand et al.	2017	United States	RCT in community setting	Adults Intervention: 33.3 Waitlist: 32.4	89	Depression	Internet-based	PHQ-9, HRSD, BAD-SF	None reported	Age, gender, ethnicity, marital status, employment, education, diagnosis, medication use	None reported
Gawrysiak et al.	2009	United States	RCT in academic setting	Adults 18.4	30	Depression	GSH	BDI-II, BAI, EROS, MSPSS	None reported	Age, gender, ethnicity	None reported
Gilbody et al.	2015	United Kingdom	RCT in primary care	Adults 39.86	691	Depression	Internet-based	PHQ-9, CORE-OM, SF-36, EQ-5D	None reported	Age, gender, education, medication use, symptom severity and onset.	None reported
Jonsbu et al.	2011	Norway	RCT in healthcare setting	Adults 52	40	Anxiety	GSH	BDI, BSQ, SF-36	None reported	Age, gender, marital status, education, employment,	None reported

										diagnosis, symptom severity	
Klein et al.	2006	Netherlands	RCT in community setting	Adults Range 18-70	55	Panic Disorder	Internet-based and GSH	PAQ, PDSS, ASP, DASS, BVS	None reported	Age, gender, education, diagnosis, medication use, symptom severity	None reported
Lindhe et al.	2023	Sweden	Pilot RCT in community setting	Adults 40.7	60	Anxiety	Internet-based	BDI-II, GAD-7, PSS, ISI, PHQ-9, BBQ, AUDIT, EAAS-M, CCI-15	None reported	Age, gender, education, employment, psychosocial support, medication use	None reported
Mayou et al.	2002	United Kingdom	RCT in healthcare setting	Adults Intervention: 43.1 Control: 45.2	80	Anxiety	GSH	BDI, STAI	None reported	Age, gender, marital status, diagnosis, disability, symptom severity	None reported
Mourad et al.	2022	Sweden	RCT in community setting	Adults Intervention: 54.3 Control: 56.8	109	Anxiety	Internet-based	PHQ-9, BSQ, EQ-VAS, CAQ	Intervention: 2% Control: 2% (CAQ, ITT)	Age, gender, marital status, economic situation, education, employment, diagnosis, symptom severity, medication use	None reported

Mourad et al.	2016	Sweden	Pilot RCT in community setting	Adults Intervention: 65 Control: 66	16	Anxiety	Internet-based	PHQ-9, BSQ, CAQ	None reported	Age, gender, marital status, economic situation, education, employment, diagnosis, symptom severity, medication use	None reported
Mulder et al.	2019	New Zealand	RCT in healthcare setting	Adults 54.5	424	Anxiety	GSH	HAI, HADS, SF-36, SFQ	None reported	Age, gender, ethnicity	None reported
Newby et al.	2014	Australia	RCT in community setting	Adults 44	99	Depression and anxiety	Internet-based	PHQ-9, GAD-7, RTQ, PBRS-A	None reported	Age, gender, marital status, education, employment, diagnosis, symptom severity	None reported
Ormrod et al.	2010	United Kingdom	Pilot quasi-experimental study in community setting	Adults 44	16	Depression and anxiety	Internet-based	BDI-II, BAI	0% (BDI-II, RCI)	Age, gender, symptom severity	None reported
Perini et al.	2009	Australia	RCT in community setting	Adults 49.29	45	Depression	Internet-based	BDI-II, PHQ-9, PANAS, K10, SDS	None reported	Age, gender, marital status, employment, symptom severity	None reported
Potts et al.	1999	United Kingdom	RCT in healthcare setting	Adults Intervention: 52.8 Control: 55.4	60	Anxiety	Group	HADS, SIP, NHP, NHS	None reported	Age, gender, symptom severity	None reported

Preschl et al.	2011	Switzerland	RCT in community setting	Adults 36.7	53	Depression	Internet-based and GSH	BDI-II	None reported	Age, gender, marital status, employment, education, medication use	None reported
Richards et al.	2012	Ireland	RCT in academic setting	Adults 26.45	80	Depression	Internet-based	BDI-II, CORE-OM, BSI	0% (BDI-II, RCI).	Age, gender	None reported
Richards et al.	2016	Ireland	RCT in community setting	Adults 38.1	281	Depression	Internet-based	BDI-II	None reported	Age, gender, education, employment, medication use, treatment history, symptom severity	None reported
Richards et al.	2015	Ireland	RCT in community setting	Adults 39.86	262	Depression	Internet-based	BDI-II, GAD-7, WSAS	None reported	Age, gender, employment, marital status, dependents, income level, difficulties, medication use, symptom severity and onset	None reported
Richards et al.	2003	United Kingdom	RCT in primary care	Adults 39.2	139	Depression and anxiety	GSH	CORE-OM, EQ-5D	GSH: 3% at 1 month, 6% at 3 months. Control: 2% and 3% (ITT)	Age, gender, symptom severity	None reported
Robinson et al.	2010	Australia	RCT in community setting	Adults 46.96	150	Anxiety (GAD)	Internet-based	GAD-7, PHQ-9, K10, SDS,	None reported	Age, gender, marital status, education,	None reported

								CEQ, PSWQ		employment, medication, symptom severity and onset	
Rollman et al.	2017	United States	RCT in primary care	Adults 42.7	704	Depression and anxiety	Internet-based with and without group	PHQ-9, GAD-7, SF-12, PROMIS	None reported	Age, gender, ethnicity, education, marital status, employment, diagnosis, medication use, symptom severity	None reported
Russell et al.	2019	United Kingdom	RCT in primary care	Adults Intervention: 35.3 Control: 40.2	20	Depression	GSH	PHQ-9, BDI-II, GAD-7, GRID-HAM-D, OCI-R, PANAS, WSAS, EQ-5D, SF-12	None reported	Age, gender, ethnicity, accommodation, education, employment, financial stress, marital status, medication use, diagnosis, symptom severity and onset	None reported
Russell et al.	2019A	United Kingdom	RCT in primary care	Adults 38	70	Depression	GSH	PHQ-9, BDI-II, SIGH-D, GAD-7, GRID-HAM-D, OCI-R, PANAS, WSAS,	None reported	Age, gender, ethnicity, accommodation, education, employment, financial stress, marital status, medication use, diagnosis,	None reported

								RRQ, RBQ-2A, EQ-5D, EuroQuol SF-12		symptom severity and onset	
Schneider et al.	2005	United Kingdom	RCT in community setting	Adults 39	68	Phobia and Panic Disorders	Internet-based	FQ, WSAS	None reported	Age, gender, marital status, education, residence, diagnosis, medication use	None reported
Simon et al.	2004	United States	RCT in primary care	Adults 44.7	600	Depression	GSH	PHQ-9, SCL-90	None reported	Age, gender, marital status, education, ethnicity, medication use, symptom severity	None reported
Thesen et al.	2022	Norway	RCT in healthcare setting	Adults 52	161	Anxiety	Internet-based	PHQ-9, BSQ, CAQ, EQ-VAS	None reported	Age, gender, marital status, employment, comorbidity, symptom severity	None reported
Titov et al.	2010	Australia	RCT in community setting	Adults 43	127	Depression	Internet-based	PHQ-9, BDI-II, K10, SDS	None reported	Age, gender, marital status, education, employment, medication use, symptom severity and onset	None reported
van Beek et al.	2013	Netherlands	RCT in healthcare setting	Adults 48.7	113	Depression and anxiety	GSH	HDRS, HADS, HAM-D,	None reported	Age, gender, marital status, employment,	None reported

						(panic disorder)		STAI, FQ, CGI		education, diagnosis, symptom severity, medication use	
Vazquez et al.	2012	Spain	RCT in academic setting	Adults 23.3	133	Depression	Group	BAI, CES-D	None reported	Age, gender, marital status, education, socioeconomic class, symptom severity	None reported
Vernmark et al.	2010	Sweden	RCT in community setting	Adults 36.8	88	Depression	GSH and internet-based	BDI, BAI, QOLI, CGI, MADRS-S	None reported	Age, gender, marital status, education, employment, medication use, diagnosis, symptom severity, treatment history	None reported
Wagner et al.	2014	Switzerland	RCT in community setting	Adults Internet: 37.99 GSH: 38.73	62	Depression	Internet-based and GSH	BDI-II, SCL-A, ATQ-R, AHS, BSI	None reported	Age, gender, education, marital status, employment, medication use, symptom severity	None reported
Whitfield et al.	2001	United Kingdom	Quasi-experimental study in primary care	Adults 28.68	42	Depression and anxiety	Pure GSH	GHQ-28, DAS, BHS	None reported	Age, gender, marital status	None reported
Williams et al.	2013	United Kingdom	RCT in primary care	Adults 40.4	281	Depression	GSH	BDI-II, CORE-OM	None reported	Age, gender, employment, deprivation	None reported

										score, medication use	
Zagorscak et al.	2018	Germany	RCT in community setting	Adults 45.7	1,089	Depression	Internet-based (guided and unguided)	BDI-II, PHQ-9, GAD-7, WHO-5, BSSS, ESES, PTQ	Guided: 1.7% post-treatment, 2.1% at 3 months, 5% at 6 months, 5.5% at 12 months; Unguided: 2.1%, 3.7%, 5.8%, 4.1% (ITT)	Age, gender, marital status, education, medication use, diagnosis, symptom severity	None reported

Note. CORE-OM = Clinical Outcomes in Routine Evaluation – Outcome Measure; PHQ-9 = Patient Health Questionnaire- 9; PHQ-15 = Patient Health Questionnaire- 15; QIDS-CR16 = Quick Inventory of Depressive Symptomatology- Clinician Rating; GDS = Geriatric Depression Scale; GRID-HAM-D = GRID-Hamilton Rating Scale for Depression; SIGH-D = Structured Interview Guide for the Hamilton Depression Rating Scale; BAD-SF = Behavioural Activation for Depression Scale- Short Form; GAD-7 = Generalised Anxiety Disorder Assessment; BAI = Becks Anxiety Inventory; BDI = Beck Depression Inventory; BDI-II: Beck Depression Inventory Second Edition; MADRS-S= Montgomery-Åsberg Depression Rating Scale- Self-Rated; PROMIS = Patient-Reported Outcomes Measurement Information; CES-D = Centre for Epidemiologic Studies Depression Scale; HADS = Hospital Anxiety and Depression Scale; HDRS = Hamilton Depression Rating Scale; QIDS = Quick Inventory of Depressive Symptomatology; ASP = Anxiety Sensitivity Profile; DASS = Depression, Anxiety, Stress Scales; ACQ = Agoraphobic Cognitions Questionnaire; PTQ = Perseverative Thinking Questionnaire; RTQ = Repetitive Thinking Questionnaire; PBRS-A = Positive Beliefs about Rumination Scale- Adapted Version; ATQ-R = Automatic Thoughts Questionnaire- Revised; PANAS = Positive and Negative Affect Scales; DAS-A = Dysfunctional Attitude Scale; PAQ = Panic Attack Questionnaire; PDSS = Panic Disorder Severity Scale; FQ = Fear Questionnaire; BSQ = Body Sensations Questionnaire; BVS = Body Vigilance Scale; DASS = Depression Anxiety Stress Scales; HADS = Hospital Anxiety and Depression Scale; PSWQ = Penn State Worry Questionnaire; PSS = Perceived Stress Scale; STAI-IT = State-Trait Anxiety Inventory; ASP = Anxiety Sensitivity Profile; CAQ = OCI-R = Obsessive Compulsive Inventory- Revised; Cardiac Anxiety Questionnaire; CCI-15 = Climate Cope Index; ISI = Insomnia Severity Index; ESES = Emotional Self-Efficacy Scale; SCL90 = Symptom Checklist; BSI = Beck Suicide Ideation Scale; AHS= American Hopelessness Scale; SF-36 = Health Survey- 36; HAI = Health Anxiety Inventory; WSAS = Work and Social Adjustment Scale; SFQ = Social Functioning Questionnaire; WHOQoL = World Health Organization Quality of Life Measure; QOLI = Quality of Life Inventory; EQ-VAS = EuroQol Visual Analog Scale; EQ-5D = EuroQol 5 Dimensions; WHO-5 = World Health Organization Wellbeing Index; BBQ = Brunnsviken Brief Quality of Life Scale; CGI-I = Clinical Global Impression- Improvement; BHM-20 = Behavioural Health Measure; NHS = Nijmegen Hyperventilation Scale; SIP = Sickness Impact Scale; NHP = Nottingham Health Profile; SDS = Sheehan Disability Scale; K-10 = Kessler 10; RBQ-2A = Adult Repetitive Behaviour Questionnaire; RRQ = Rumination-Reflection Questionnaire; EROS = Environmental Reward Observation Scale; MSPSS: Multidimensional Scale of Perceived Social Support; BSSS = Berlin Social Support Scale; EAAS-M = Environmental Action Scale; AUDIT = Alcohol Use Disorders Identification Test; CEQ = Credibility/Expectancy Questionnaire

Participant Sociodemographic and Clinical Characteristics

Sample sizes ranged from 11 to 2,794 participants, with a total pooled sample size of 14,716 participants. The populations studied were exclusively adults aged 18 years-old or over, with only one study focusing on older adults aged over 58 years-old (Egede et al., 2015).

Age and gender were the most common sociodemographic characteristics, reported in all studies. Other of the most frequently reported sociodemographic variables were education (n= 30, 68.2%), marital or relationship status (n=29, 65.9%), and employment status (n= 26, 59.1%). Ethnicity was less commonly reported, being included in 11 studies (25%). Other economic characteristics such as income level, socioeconomic class, deprivation scores, and financial stress were reported in 11 studies (25%), Some variables like residence, accommodation, sexual orientation, and psychosocial support were reported in ≤ 4 studies.

Regarding clinical characteristics, symptom severity was a key focus, analysed in 31 studies (70.5%). The other two most common clinical characteristics mentioned included medication use (n= 24, 54.5%), and psychiatric diagnosis (n= 18, 40.1%). The remaining clinical characteristics found, related to clinical history and physical health, were reported in a minority of studies, appearing in only in 7 (15.9%) and 4 (9.1%) studies respectively.

These participant characteristics were mainly used to identify significant group differences at baseline, check for potential confounders, and to calculate effect sizes for positive outcomes.

Deterioration Rates in LICBT Studies

Only 7 of the 44 studies (15.9%) explicitly reported symptomatic deterioration rates. Of these, two approaches to reporting deterioration were identified: ITT analyses and RCI criteria.

ITT analyses include all participants initially randomised into intervention or control groups, regardless of dropout or adherence to treatment, providing a conservative estimate of

deterioration that accounts for real-world variability in treatment implementation. ITT-based deterioration rates were reported in 4 studies: Cano-Vindel et al. (2021), Cavanagh et al. (2006), Mourad et al. (2022), and Richards et al. (2003).

In the ITT group, deterioration rates varied across studies, delivery methods, and timepoints. Cano-Vindel et al. (2021) observed 3% deterioration in the intervention group and 14% in the control group post-treatment, with similar rates at 12-month follow-up (3% and 12%, respectively). Cavanagh et al. (2006) and Mourad et al. (2022) both reported consistent deterioration rates of 2% in both intervention and control groups. Richards et al. (2003) found higher rates in the GSH intervention group (3% at 1 month and 6% at 3 months) compared to the control group (2% and 3% in the same timepoints).

On the other hand, RCI calculations identify deterioration using psychometric criteria, evaluating whether changes in participant scores on validated outcome measures exceed thresholds of reliable measurement error. This approach offers a more precise measure of symptom deterioration but may underestimate rates by excluding participants with incomplete data. Four studies reported deterioration rates using the RCI criteria.

Zagorscak et al. (2018) documented deterioration rates of 1.7% in the guided intervention group post-treatment, increasing to 5.5% at 12 months, which were consistently lower than the unguided group, with 2.1% post-treatment and 5.8% at follow-up. Cavanagh et al (2006) observed higher deterioration rates of 5% post-treatment and 6% at 6-months follow-up in participants that completed the intervention. While Ormrod et al. (2010) and Richards et al. (2012) reporting 0% deterioration rates in their studies.

Studies that used ITT analyses demonstrated rates ranging from 2% to 3% in interventions groups post-treatment, and between 3% and 6% at follow-up, with control groups often exhibiting higher rates (12% and 14%). Those using RCI criteria reported higher

variability post-treatment, with two studies documenting no deterioration and others showing rates between 1.7% and 5%, although follow-up data did not have this variability with rates between 5.5% and 6%. It is also important to highlight that both studies that reported 0% prevalence rates with the RCI criteria should be considered with caution as they had small sample sizes of 16 and 80 participants. Only one study reported both, ITT analyses and RCI criteria, which yielded different post-treatment and follow-up prevalence rates.

Overall, deterioration prevalence rates ranged from 0% to 5% post-treatment and between 3% and 6% at follow-up in the intervention groups across studies. Most of the studies with follow-data, also found significantly higher prevalence rates at follow-up than those at post-treatment.

Characteristics of LICBT Studies Reporting Deterioration Rates

Most identified studies with reported deterioration rates were RCTs (n= 5) with a minority using quasi-experimental designs (n= 2) in community settings (n= 3), primary care (n= 2), or both settings (n= 1). All were conducted in Europe between 2003 and 2022, with three studies conducted in the United Kingdom, and the others in Sweden, Spain, Germany, and Ireland.

Most interventions targeted depression and anxiety (n= 4) with reported rates between 0% and 3% post-treatment and between 3% and 6% at follow-up. Two studies targeted depression with rates between 0% and 1.7% post-treatment, and between 5.5% and 5.8% at follow-up. Only one study's intervention targeted anxiety with reported rates of 2% post-treatment.

Delivery methods were mainly internet-based (n= 5) with the remaining studies focusing on GSH (n= 1) and groups (n= 1). Internet-based interventions reported rates between

0% and 5.5% post-treatment and between 5.5% and 6% at follow-up; group interventions 3% post-treatment and follow-up; and GSH 3% post-treatment and 6% at follow-up.

These studies also had small sample sizes, with two studies including less than 100 participants (Ormrod et al, 2010; Richards et al., 2012), three studies between 100 and 220 participants (Cavanagh et al., 2006; Mourad et al., 2022; Richards et al., 2003), and only two studies including over 1,000 participants (Cano-Vindel et al., 2021; Zagorscak et al., 2018).

Deterioration Sub-Group Analyses and Patient Characteristics

Subgroup analyses were generally limited to treatment effectiveness. Among the seven studies that reported deterioration rates, common sociodemographic variables analysed included age (n = 7), gender (n = 7), marital status (n= 3), education (n= 3), and employment (n= 2). The most common clinical variables were symptom severity (n = 6), psychiatric diagnosis (n= 2) and medication use (n = 2). The clinical outcome measures used in more than one study were the following: BDI-II (n= 3), the PHQ-9 (n= 2), the GAD-7 (n= 2), and the CORE-OM (n=2).

While these variables were gathered and included in baseline subgroup analyses, they were not used in subgroup analyses to investigate effect sizes, possible confounders, predictors or moderators focused on deterioration in any study.

Risk of Bias Assessment

Among RCTs, two studies demonstrated low risk of bias: Cano-Videl. (2021) and Zagorscak et al. (2018), while Mourad et al. (2022), and both Richards et al. studies (2003; 2012) were assessed as having medium risk due to issues with blinding, applications, and generalisability.

All RCTs had clearly formulated research questions, appropriate study designs, adequate participant randomisation, treated participants equally, reported their results

comprehensively including positive and negative treatments effects. On the other hand, common limitations included significant attrition levels, non-blindness, unclear applications to clinical practice due to limited generalisability, and lack of comparison with existing interventions (see Table 4).

Table 4.

Risk of Bias Analysis (RCTs)

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Risk
Cano-Vindel et al. (2021)	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	CT	Low
Mourad et al. (2022)	Y	Y	Y	N	Y	Y	Y	Y	CT	CT	CT	Medium
Richards et al. (2003)	Y	Y	Y	N	Y	Y	Y	Y	CT	CT	CT	Medium
Richards et al. (2012)	Y	Y	Y	N	Y	Y	Y	Y	CT	CT	CT	Medium
Zagorscak et al. (2018)	Y	Y	Y	N	Y	Y	Y	Y	CT	Y	CT	Low

Note: Y = Yes; N = No; CT = Cannot Tell. Questions can be found in Appendix B

Critical appraisals of studies with quasi-experimental designs should assess design suitability, management of confounding variables, data collection, analytic methods, result's validity, generalisability, and ethical considerations (CASP, 2024).

Cavanaugh et al. (2006) and Ormond et al. (2010) used quasi-experimental designs that were suitable for their research aims, used valid psychometric measures, reported results comprehensively, and addressed ethical considerations. Although, common limitations were a lack of power calculations, limited generalisability and unclear clinical implications.

Cavanagh et al. (2006) demonstrated a low risk of bias, while the study included multiple statistical analyses, reported results comprehensively, and stated clear clinical implication, key limitations included high attrition rates, and limited generalisability. On the

other hand, Ormond et al. (2010) had an increased risk of bias in the medium-high range due to significant limitations with a small sample size, reduced statistical power, possible recruitment bias, high attrition rates, unaddressed confounders, and limited generalisability.

Discussion

The majority of studies included in this review, and all of the studies reporting deterioration rates, were conducted in Europe, particularly in the United Kingdom. This geographic concentration raises questions about the generalisability of findings to other regions. North America, Oceania, and Asia were underrepresented, and no eligible studies were identified from Africa or South America, limiting the global applicability of the findings.

Most studies were conducted in community settings, with fewer studies in primary care and healthcare settings, a trend mirrored in the studies reporting deterioration rates. This distribution favours the understanding of LICBT intervention's effects and deterioration rates toward community settings, with limited insights from other services, and contexts.

Delivery methods were dominated by internet-based interventions, while pure or guided self-help and group-based therapies were underrepresented. This bias toward internet-based approaches may overlook effect sizes and deterioration trends in other delivery formats, especially those involving more direct interactions or group dynamics.

The wide array of outcome measures used reflects the multidimensional and complex nature of mental health research and the significant challenges in conducting comparisons between studies and information synthesis for systematic reviews and meta-analyses. In this instance, a meta-analysis was not possible due to the high variability of outcome measures and methodological inconsistencies. Differences in the sensitivity and specificity of outcome measures may result in variable thresholds for identifying deterioration, which could lead to inconsistencies in the calculated rates and effect sizes. Furthermore, some measures used are

rarely seen in the wider literature, limiting their comparability to more established measures like the BDI or PHQ-9. This heterogeneity highlights a critical need for consensus on the most appropriate and standardised measures for evaluating outcomes in LICBT, particularly when assessing adverse effects such as symptomatic deterioration.

A significant finding and limitation in this review is the lack of reporting on deterioration rates and the inconsistent use of reporting frameworks for deterioration, with 37 of the 44 included studies (84.1%) failing to provide deterioration rates. Only seven studies reported deterioration rates using ITT or RCI criteria, although no rationale was provided for their chosen method.

While ITT analyses provide an overall estimate of deterioration that includes data from all participants, RCI offers a more stringent and clinically meaningful criterion, focusing on measurable changes in symptom scores. The use of ITT may overestimate deterioration in cases of treatment non-adherence or dropout, whereas RCI requires validated score changes, potentially underestimating deterioration by excluding participants without complete outcome data. This variability in reporting methods complicates direct comparisons of deterioration rates across studies. This is illustrated in the only study using both ITT and RCI criteria to calculate deterioration rates, finding different post-treatment and follow-up rates. In this review, studies using RCI criteria yielded higher variability post-treatment (from 0% to 5%) than studies using ITT analyses (from 2% to 3%), while those using ITT demonstrated higher variability at follow-up with rates between 3% and 6%. Additionally, a study with high risk of bias and two studies with very small sample sizes were included, reducing reliability.

Interventions targeting depression and anxiety reported 0-3% rates post-treatment and between 3-6% at follow-up. While prevalence rates on studies targeting one condition were identified, findings should be considered with caution as they represent one or two studies

(depression n= 2, anxiety n= 1). Similarly, delivery methods were mainly internet-based (n= 5) with 0-5.5% rates post-treatment and 5.5-6% at follow-up. Yet, GSH and group interventions were only represented in one study each, severely limiting their reliability and generalisability.

Studies in this systematic review that reported deterioration rates found similar rates to those found in previous research. The most common deterioration prevalence rates for adults in psychotherapy are between 5 and 10% (Lambert, 2013). Multiple range variations have been found in the literature between 0 and 17% (Cahill, Barkham, & Stiles, 2010; Cuijpers, 2021; Klatte et al., 2023; Lazar, 2017; Schermuly-Haupt, Linden, & Rush, 2018; Thorpe, 2017).

While there are no systematic reviews or meta-analyses on LICBT deterioration, there are some studies that have used these designs to investigate LICBT effectiveness focusing on a target condition, population, or delivery method. Some systematic reviews and meta-analyses relevant to this study focused on LICBT for depression (Bower et al., 2013), GAD (Powell et al., 2024), and OCD (Hoppen et al., 2021), using GSH (Cuijpers et al., 2010) and cCBT (Grist & Cavanagh, 2013), in specific population like older adults (Cremers et al., 2019). Some of them reported limitations similar to those in this study, including a wide variety of outcome measures, moderate risk of bias, few studies meeting eligibility criteria, and small sample sizes.

Subgroup analyses were notably limited across the studies, with none performing subgroup analyses to identify predictors or moderators of deterioration. While age and gender were reported in all studies, their impact on treatment outcomes was rarely analysed outside baseline comparisons and improvement effect size calculations. Other variables, such as ethnicity, marital status, employment, and other socioeconomic characteristics were inconsistently included and rarely tested for associations with poor treatment outcomes. Similarly, clinical factors, such as diagnosis, symptom severity and medication use, were

analysed more often but not included in subgroup analyses to investigate their association with deterioration.

The lack of deterioration subgroup analyses could be partly due to insufficient statistical power. As deterioration occurs in a minority of participants and a significant proportion of studies had small sample sizes of less than 100 participants (n= 21) with only five studies including over 500 participants. Subgroup analyses with these small numbers could yield results lacking validity, reliability, and precision. Although even the two studies with over 1,000 participants reporting deterioration rates did not report deterioration subgroup analyses

Some sociodemographic characteristics that have been found in the literature to be predictors or moderators for poor treatment outcomes, such as treatment non-response and deterioration, include age, ethnicity, unemployment, disability, long-term conditions, and socioeconomical deprivation (Alegria et al., 2018; Buckman et al., 2021; Delgadillo et al., 2016; Delgadillo, Moreea, & Lutz, 2016; Finegan et al., 2018; Finegan, Firth, & Delgadillo, 2019; Reneflot & Evensen, 2012; Saunders et al., 2021; Saxon et al., 2017; Stochl et al., 2021; Teo, 2021; van Agteren et al., 2021). Although, there are studies that have not found associations between age, gender, ethnicity and treatment outcomes (Bauer-Staeb et al., 2023; Kellet et al., 2016; Saunders et al., 2019; Saunders et al., 2021; Trombello et al., 2020). Studies investigating other predictor or moderation associations with poor outcomes are scarce, including characteristics such as sexual orientation and religion (Bauer-Staeb et al., 2023; Captari et al., 2018).

High levels of effectiveness variability in LICBT have been identified in a CBT meta-review (Fordham et al., 2021) along with gaps in the evidence-base, especially regarding older adults and moderator analyses for cultural factors (particularly ethnicity and country of residence). These gaps were also found in this review, as ethnicity was reported in 11 studies

and residence in 3 of the wider review, with none mentioned in the studies reporting deterioration rates.

Regarding clinical characteristics found as predictors or moderators of poor treatments outcomes in the other research studies, these include psychiatric comorbidity, high baseline severity, and use of psychotropic medication (Bauer-Staeb et al., 2023; Buckman et al., 2021; Kessler et al., 2017; Maj et al., 2020; Saunders et al., 2016; Saunders et al., 2021; Teo, 2021; Weitz et al., 2015).

Overall, multiple studies have found inconsistent evidence regarding age, gender, ethnicity, socioeconomic status, and psychiatric comorbidity to explain the variance in psychotherapy treatment outcomes. Clinical characteristics, such as increased baseline symptom severity and chronicity have been more consistently associated with poorer treatment outcomes.

Strengths and Limitations

The protocol for this systematic review was pre-registered and the search included multiple well-known databases and a wide variation of common terms used to describe the topic of interest (deterioration), population (adults), interventions (LICBT), target conditions (LICBT appropriate depression and anxiety disorders), research design (RCT and quasi-experimental), and clinical setting of interest (primary care, community, and outpatient). This allowed for a wide variety of studies from different countries, settings, delivery methods, and with different target conditions and populations to be screened and included in the review.

This systematic review is the first to investigate deterioration following LICBT interventions, including prevalence rates and subgroup analyses. The review gathered, analysed, and synthesised information regarding study characteristics, common reported patient characteristics (clinical and sociodemographic), and clinical deterioration (analyses

used and prevalence rates found post-treatment and at follow-up, including differences found across delivery modalities and target conditions). This allowed the identification of research gaps and their implications in clinical practice and future research recommendations.

Although, the eligibility criteria included studies focused exclusively on adults, published in English, and with target conditions aligned with the guidelines set forth by NHS Talking Therapies guidelines, restricting the applicability of findings to the broader population requiring LICBT in other countries, settings, populations, and target conditions. Additionally, as the review only included peer reviewed studies, there is the possibility of publication bias.

One of the most significant limitations was the small number of studies identified that reported deterioration and targeted either depression (n= 2) or anxiety (n= 1) and were delivered in a group (n= 1) or GSH (n= 1) format. All studies were also conducted in Europe, and most were delivered in community settings. This lack of representation considerably limits the finding's reliability and generalisability, increasing the risk of bias in favour of internet-based interventions targeting both conditions in European community settings.

Furthermore, the search strategy included search terms regarding deterioration, which limited the number of studies retrieved, screened, and included in the review. To calculate a more precise estimation of the percentage of LICBT studies that report deterioration, a wider selection of studies within the LICBT literature would need to be performed, which could decrease the overrepresentation of internet-based community interventions targeting depression and anxiety, addressing this study's main limitation.

Clinical Practice Implications

This systematic review found that symptomatic deterioration after LICBT is underreported and found in a minority of cases post-treatment and follow-up (0% to 6%). While this adverse outcome might only affect a minority of patients, mental health professionals

should be aware of this phenomenon in routine clinical practice and specific to the models or intervention used. This information should be discussed with patients to improve their understanding of the risks and benefits associated with psychotherapy, enable risk-benefit informed consent and provide a rationale and agreement for the routinely use of problem-specific outcome measures. Practitioners should use validated psychometric measures with reliable improvement and deterioration thresholds to identify progress barriers and adverse events or treatment effects promptly during the intervention to inform their treatment plans and take prompt action to prevent drop out and clinical deterioration post-treatment.

Although this review found limited evidence on specific sociodemographic and clinical predictors or moderators of deterioration in LICBT, it is important to know the evidence found in other research studies in the wider literature. While there is inconsistent evidence that age, gender, ethnicity, socioeconomic status, and psychiatric comorbidity are associated with adverse outcomes, increased baseline symptom severity and chronicity have been more consistently associated with poor treatment outcomes. Clinicians should be aware of the current research on adverse outcomes to inform ethical decision-making and routine clinical practice.

Future Research

This review found significant gaps in the LICBT literature, including a significant lack of reporting deterioration and subgroup analyses. Future research must address this by reporting therapeutic benefits and potential harms, including reporting deterioration rates using ITT analyses or RCI criteria, and exploring sociodemographic and clinical characteristics as moderators or predictors of poor treatment outcomes. To accomplish this, a framework for baseline data collection including variables of interest and psychometric outcomes measures, could be developed and used in psychotherapy research, especially for experimental designs with sufficient statistical power to examine subgroup differences at different timepoints to

explore the clinical and sociodemographic characteristics of those affected by deterioration and their symptom trajectories post-treatment and at different follow-up timepoints.

Other important considerations include expanding geographical and cultural representation, along with diversifying study settings, populations, target conditions, and delivery methods to improve the generalisability of findings to a broader range of contexts and communities.

Conclusion

This systematic review is the first to study deterioration after LICBT interventions. Its main findings include the identification of how many research studies report deterioration rates (15.9%), the prevalence rates found (0%-5.5% post-treatment, and 3%-6% at follow-up), the methods used (ITT analyses and RCI criteria), and the sociodemographic and clinical characteristics associated with deterioration (none due to a lack of subgroup analyses). It is important to consider these findings with caution as some key limitations were identified.

Furthermore, it provides an overview of the included study characteristics and variables of interest in the wider literature and in studies reporting deterioration rates. This allowed the identification of research gaps, common limitations, and future research recommendations.

References

Alegria, M., NeMoyer, A., Falgas, I., Wang, Y., & Alvarez, K. (2018). Social determinants of mental health: Where we are and where we need to go. *Current Psychiatry Reports, 20*, 95. https://doi.org/10.1007/s11920-018-0969-9

Ali, S., Rhodes, L., Moreea, O., McMillan, D., Gilbody, S., Leach, C., Lucock, M., Lutz, W., & Delgadillo, J. (2017). How durable is the effect of low intensity CBT for depression and anxiety? Remission and relapse in a longitudinal cohort study. *Behaviour Research and Therapy, 94*, 1-8. https://doi.org/10.1016/j.brat.2017.04.006

Andersson, G., Bergstrom, J., Hollandare, F., Carlbring, P., Kaldo, V., & Ekselius, L. (2005). Internet-based self-help for depression: Randomised controlled trial. *British Journal of Psychiatry, 187*, 456-461. https://doi.org/10.1192/bjp.187.5.456

Andersson, G., Hesser, H., Veilord, A., Svedling, L., Andersson, F., Sleman, O., Mauritzson, L., Sarkohl, A., Claesson, E., Zetterqvist, V., Lamminen, M., Eriksson, T., & Carlbring, P. (2013). Randomised controlled non-inferiority trial with 3-year follow-up of internet-delivered versus face-to-face group cognitive behavioural therapy for depression. *Journal of Affective Disorders, 151 (3)*, 986-994. https://doi.org/10.1016/j.jad.2013.08.022

Andrews, G., Cuijpers, P., Craske, M. G., McEvoy, P., & Titov, N. (2010). Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: A meta-analysis. *PLoS One, 5 (10)*, e13196. https://doi.org/10.1371/journal.pone.0013196

Bauer-Staeb, C., Griffith, E., Faraway, J. J., & Button, K. S. (2023). Trajectories of depression and generalised anxiety symptoms over the course of cognitive behaviour

therapy in primary care: An observational, retrospective cohort. *Psychological Medicine, 53 (10)*, 4648-4656. https://doi.org/10.1017/S0033291722001556

Beck, A., Epstein, N., Brown, G., & Steer, R. (1988). Beck Anxiety Inventory. *APA PsycTests*. https://doi.org/10.1037/t02025-000

Beck, A., Steer, R., & Brown, G. (1996). Beck Depression Inventory- II. *APA PsycTests*. https://doi.org/10.1037/t00742-000

Beck, A., Ward, C., Mendelson, M., Mock, J., & Erbauch, J. (1961). Beck Depression Inventory. *APA PsycTests*. https://doi.org/10.1037/t00741-000

Bennett-Levy, J., Richards, D. A., Farrand, P., Christensen, H., Griffiths, K., Kavanagh, D., & Williams, C. (2010). *Oxford Guide to Low Intensity CBT Interventions.* Oxford Univerity Press.

Benton, S., Heesacker, M., Snowden, S., & Lee, G. (2016). Therapist-assisted, online (TAO) intervention for anxiety in college students: TAO outperformed treatment as usual. *Professional Psychology: Research and Practice, 47 (5)*, 363-371. https://doi.org/10.1037/pro0000097

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D. A., Gilbody, S., Knowles, S., Cuijpers, P., Andersson, G., Christensen, H., Meyer, B., Huibers, M., Smit, F., van Straten, A., Warmerdam, L., Barkham, M., Bilich, L., Lowell, K., & Liu, E. (2013). Influence of initial severity of depression on effectiveness of low intensity interventions: Meta-analysis of individual patient data. *BMJ, 346*, f540. https://doi.org/10.1136/bmj.f540

Buckman, J. E., Saunders, R., Cohen, Z. D., Barnett, P., Clarke, K., Ambler, G., DeRubeis, R., Gilbody, S., Hollon, S., Kendrick, T., Watkins, E., Wiles, N., Kessler, D., Richards,

D., Sharp, D., Brabyn, S., Littlewood, E., Salisbury, C., White, I., Lewis, G., & Pilling Stephen. (2021). The contribution of depressive 'disorder characteristics' to determinations of prognosis for adults with depression: An individual patient data meta-analysis. *Psychological Medicine, 51 (7)*, 1068-1081. https://doi.org/10.1017/S0033291721001367

Cahill, J., Barkham, M., & Stiles, W. B. (2010). Systematic review of practice-based research on psychological therapies in routine clinic settings. *British Journal of Clinical Psychology, 49 (4)*, 421-53. https://doi.org/10.1348/014466509X470789

Cano-Vindel, A., Munoz-Navarro, R., Moriana, J., Ruiz-Rodriguez, P., Medrano, L., & Gonzalez-Blanch, C. (2022). Transdiagnostic group cognitive behavioural therapy for emotional disorders in primary care: The results of the PsicAP randomized controlled trial. *Psychological Medicine, 52 (15)*, 3336-3348. https://doi.org/10.1017/S0033291720005498

Captari, L. E., Hook, J. N., Hoyt, W., Davis, D. E., McElroy-Heltzel, S. E., & Worthington Jr, E. L. (2018). Integrating clients' religion and spirituality within psychotherapy: A comprehensive meta-analysis. *Journal of Clinical Psychology, 74 (11)*, 1938-1951. https://doi.org/10.1002/jclp.22681

Critical Appraisal Skills Programme. (2023). *Critical Appraisal Checklists*. https://casp-uk.net/casp-tools-checklists/

Critical Appraisal Skills Programme. (2024). *Quasi-Experimental Design.* https://casp-uk.net/news/quasi-experimental-design/

Cavanagh, K., Shapiro, D., Van Den Berg, S., Swain, S., Barkham, M., & Proudfoot, J. (2006). The effectiveness of computerized cognitive behavioural therapy in routine

care. *British Journal of Clinical Psychology, 45 (Pt 4)*, 499-514. https://doi.org/10.1348/014466505X84782.

Christensen, H., Griffiths, K., Mackinnon, A., & Brittliffe, K. (2006). Online randomized controlled trial of brief and full cognitive behaviour therapy for depression. *Psychological Medicine, 36 (12)*, 1736-1746. https://doi.org/10.1017/S0033291706008695

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry, 23 (4)*, 318-327. https://doi.org/10.3109/09540261.2011.606803.

Cohen, J. (1960). A coefficient of agreements for nominal scales. *Educational and Psychological Measurement, 20 (1)*, 37-46. https://doi.org/10.1177/001316446002000104.

Core System Trust. (2014). *The Intruments.* https://www.coresystemtrust.org.uk/home/instruments/

Coull, G., & Morris, P. (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

Cremers, G., Taylor, E., Hodge, L., & Quigley, A. (2019). Effectiveness and acceptability of low-intensity psychological interventions on the well-being of older adults: A systematic review. *Clinical Gerontologist, 45 (2)*, 214–234. https://doi.org/10.1080/07317115.2019.1662867

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

Dahne, J., Kustanowitz, J., & Lejuez, C. (2018). Development and preliminary feasibility study of a brief behavioral activation mobile application (Behavioral Apptivation) to be used in conjunction with ongoing therapy. *Cognitive and Behavavioural Practice, 25 (1)*, 44-56. https://doi.org/10.1016/j.cbpra.2017.05.004

de Graaf, L., Gerhards, S., Arntz, A., Riper, H., Metsemakers, J., Evers, S., Severens, J., Widdershoven, G., & Huibers, M. (2009). Clinical effectiveness of online computerised cognitive–behavioural therapy without support for depression in primary care: Randomised trial. *British Journal of Psychiatry, 195 (1)*, 73-80. https://doi.org/10.1192/bjp.bp.108.054429

Delgadillo, J., Asaria, M., Ali, S., & Gilbody, S. (2016). On poverty, politics and psychology: The socioeconomic gradient of mental healthcare utilisation and outcomes. *British Journal of Psychiatry, 209 (5)*, 429-430. https://doi.org/10.1192/bjp.bp.115.171017

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., 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

Ebert, D., Buntrock, C., Lehr, D., Smit, F., Riper, H., Baumeister, H., & Berking, M. (2018). Effectiveness of web- and mobile-based treatment of subthreshold depression with

adherence-focused guidance: A single-blind randomized controlled trial. *Behaviour Therapy, 49 (1)*, 71-83. https://doi.org/10.1016/j.beth.2017.05.004

Egede, L., Acierno, R., Knapp, R., Lejuez, C., Hernandez-Tejada, M., Payne, E., & Frueh, B. (2015). Psychotherapy for depression in older veterans via telemedicine: A randomised, open-label, non-inferiority trial. *The Lancet, 2 (8)*, 693-701. https://doi.org/10.1016/S2215-0366(15)00122-4

Eriksen, M., & Frandsen, T. (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

Ernst, E. (2001). Assessment of therapeutic safety in systematic reviews: Literature review. *BMJ*, 323. https://doi.org/10.1136/bmj.323.7312.546

Finegan, M., Firth, N., & Delgadillo, J. (2020). Adverse impact of neighbourhood socioeconomic deprivation on psychological treatment outcomes: The role of area-level income and crime. *Psychotherapy Research, 30 (4)*, 546-554. https://doi.org/10.1080/10503307.2019.1649500

Finegan, M., Firth, N., Wojnarowski, C., & Delgadillo, J. (2018). Associations between socioeconomic status and psychological therapy outcomes: A systematic review and meta-analysis. *Depression and Anxiety, 35 (6)*, 560-573. https://doi.org/10.1002/da.22765

Forand, N., Barnett, J., Strunk, D., Hindiyeh, M., Feinberg, J., & Keefe, J. (2018). Efficacy of guided iCBT for depression and mediation of change by cognitive skill acquisition. *Behaviour Therapy, 49 (2)*, 295-307. https://doi.org/10.1016/j.beth.2017.04.004

Fordham, B., Sugavanam, T., Edwards, K., Stallard, P., Howard, R., das Nair, R., Copsey, B., Lee, H., Howick, J., Hemming, K., & Lamb, S. (2021). The evidence for cognitive behavioural therapy in any condition, population or context: A meta-review of systematic reviews and panoramic meta-analysis. *Psychological Medicine, 51 (1)*, 21-29. https://doi.org/10.1017/S0033291720005292

Gawrysiak, M., Nicholas, C., & Hopko, D. (2009). Behavioral activation for moderately depressed university students: Randomized controlled trial. *Journal of Counseling Psychology, 56 (3)*, 468-475. https://doi.org/10.1037/a0016383

Gilbody, S., Littlewood, E., Hewitt, C., Brierley, G., Tharmanathan, P., Araya, R., Barkham, M., Bower, P., Cooper, C., Gask, L., Kessler, D., Lester, H., Lovell, K., Parry, G., Richards, D., Andersen, P., Brabyn, S., Knowles, S., Shepherd, C., Tallon, D., & White, D. (2015). Computerised cognitive behaviour therapy (cCBT). *BMJ, 351*. https://doi.org/10.1136/bmj.h5627

Grist, R., & Cavanagh, K. (2013). Computerised cognitive behavioural therapy for common mental health disorders, what works, for whom under what circumstances? A systematic review and meta-analysis. *Journal of Contemporary Psychotherapy, 43*, 243-251. https://doi.org/10.1007/s10879-013-9243-y

Hoppen, L., Kuck, N., Burkner, P., Karin, E., Wootton, B., & Buhlmann, U. (2021). Low intensity technology-delivered cognitive behavioral therapy for obsessive-compulsive disorder: A meta-analysis. *BMC Psychiatry, 21*, 322. https://doi.org/10.1186/s12888-021-03272-5

Ioannidis, J. P., Evans, S. J., Gotzsche, P. C., O'Neill, R. T., Altman, D. G., Schulz, K., & Moher, D. (2004). Better reporting of harms in randomized trials: An extension of the

CONSORT statement. *Annals of Internal Medicine., 141 (1)*, 781-788. https://doi.org/10.7326/0003-4819-141-10-200411160-00009

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psycholology, 59 (1)*, 12-19. https://doi.org/10.1037//0022-006x.59.1.12

Jonsbu, E., Dammen, T., Morken, G., Moum, T., & Martinsen, E. (2011). Short-term cognitive behavioral therapy for non-cardiac chest pain and benign palpitations: A randomized controlled trial. *Journal of Psychosomatic Research, 70 (2)*, 117-123. https://doi.org/10.1016/j.jpsychores.2010.09.013

Kellett, S., Webb, K., Wilkinson, N., Bliss, P., Ayers, T., & Hardy, G. (2016). Developing services for patients with depression or anxiety in the context of long-term physical health conditions and medically unexplained symptoms: Evaluation of an IAPT pathfinder site. *Behavioural and Cognitive Psychotherapy, 44 (5)*, 553-567. https://doi.org/10.1017/S1352465816000114

Kessler, R. C., van Loo, H., Wardenaar, K., Bossarte, R., Brenner, L., Ebert, D., de Jonge, P., Nierenberg, A., Rosellini, A., Sampson, N., Scoevers, R., Wilcox, M., & Zaslavsky, A. (2017). Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiology and Psychiatric Sciences, 26 (1)*, 22-36. https://doi.org/10.1017/S2045796016000020

Klatte, R., Strauss, B., Fluckiger, C., & Rosendahl, J. (2023). Adverse events in psychotherapy randomized controlled trials: A systematic review. *Psychotherapy Research*, 1-16. https://doi.org/10.1080/10503307.2023.2286992

Klein, B., Richards, J., & Austin, D. (2006). Efficacy of internet therapy for panic disorder. *Journal of Behaviour Therapy and Experimental Psychiatry, 37 (3)*, 213-238. https://doi.org/10.1016/j.jbtep.2005.07.001

Kroenke, K., Spitzer, R., & Williams, J. (2001). The PHQ-9. *Journal of General Internal Medicine, 16,* 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lambert, M. J. (2013). Outcome in psychotherapy: The past and important advances. *Psychotherapy, 50 (1)*, 42-51. https://doi.org/10.1037/a0030682

Lazar, E. (2017). *Client deterioration in individual psychotherapy: A systematic review.* (Publication No. 10605182) [Doctoral dissertation, Indiana University of Pennsylvania]. ProQuest Dissertations & Theses. https://www.proquest.com/openview/ae7c771205476bd30542968850c0637c/1?pq-origsite=gscholar&cbl=18750

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2 (1),* 53-70. https://doi.org/10.1111/j.1745-6916.2007.00029.x

Lindhe, N., Bengtsson, A., Byggeth, E., Engstrom, J., Lundin, M., Ludvigsson, M., Aminoff, V., Berg, M., & Andersson, G. (2023). Tailored internet-delivered cognitive behavioral therapy for individuals experiencing psychological distress associated with climate change: A pilot randomized controlled trial. *Behaviour Research and Therapy, 171*, 104438. https://doi.org/10.1016/j.brat.2023.104438

Maj, M., Stein, D. J., Parker, G., Zimmerman, M., Fava, G. A., De Hert, M., Demyttenaere, K., McIntyre, R., Widiger, T., & Wittchen, H.-U. (2020). The clinical characterization of the adult patient with depression aimed at personalization of management. *World Psychiatry, 19 (3)*, 269-293. https://doi.org/10.1002/wps.20771

Mayou, R., Springings, D., Birkhead, J., & Prince, J. (2002). A randomized controlled trial of a brief educational and psychological intervention for patients presenting to a cardiac clinic with palpitation. *Psychological Medicine, 32*, 699-706. https://doi.org/10.1017}S0033291702005536

McKay, D., & Jensen-Doss, A. (2021). Harmful treatments in psychotherapy. *Clinical Psychology: Science and Practice, 28 (1)*, 2-4. https://doi.org/10.1037/cps0000023

Mourad, G., Eriksson-Liebon, M., Karlstrom, P., & Johansson, P. (2022). The effect of internet-delivered cognitive behavioral therapy versus psychoeducation only on psychological distress in patients with non-cardiac chest pain: Randomized controlled trial. *Journal of Medical Internet Research, 24 (1)*, e31674. https://doi.org/10.2196/31674

Mourad, G., Stromberg, A., Jonsbu, E., Gustafsson, M., Johansson, P., & Jaarsma, T. (2016). Guided internet-delivered cognitive behavioural therapy in patients with non-cardiac chest pain – A pilot randomized controlled study. *BMC Trials, 17*, 352. https://doi.org/10.1186/s13063-016-1491-1

Mulder, R., Zarifeh, J., Boden, J., Lacey, C., Tyrer, P., Tyrer, H., Than, M., & Throughton, R. (2019). An RCT of brief cognitive therapy versus treatment as usual in patients with non-cardiac chest pain. *International Journal of Cardiology, 289*, 6-11. https://doi.org/10.1016/j.ijcard.2019.01.067

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

National Collaborating Centre for Mental Health. (2018). *The Improving Access to Psychological Therapies Manual.* NHS England.

https://www.england.nhs.uk/publication/the-improving-access-to-psychological-therapies-manual/

Newby, J., Mewton, L., Williams, A., & Andrews, G. (2014). Effectiveness of transdiagnostic internet cognitive behavioural treatment for mixed anxiety and depression in primary care. *Journal of Affective Disorders, 165*, 45-52. https://doi.org/10.1016/j.jad.2014.04.037

National Health Service. (2023). *Psychological therapies: Reports on the use of IAPT services, England, March 2023 Final including a report on the IAPT Employment Advisers pilot and Quarter 4 2022-23 data*. https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-report-on-the-use-of-iapt-services/march-2023-final-including-a-report-on-the-iapt-employment-advisers-pilot#

National Institute for Health and Care Excellence. (2014). *Anxiety Disorders*. [NICE Guideline QS53]. https://www.nice.org.uk/guidance/qs53/chapter/Quality-statement-2-Psychological-interventions

National Institute for Health and Care Excellence. (2022). *Depression in adults: Treatment and management*. [NICE Guideline NG222]. https://www.nice.org.uk/guidance/ng222

Ormrod, J., Kennedy, L., Scott, J., & Cavanagh, K. (2010). Computerised cognitive behavioural therapy in an adult mental health service: A pilot study of outcomes and alliance. *Cognitive Behaviour Therapy, 39 (3)*, 188–192. https://doi.org/10.1080/16506071003675614

Page, M., McKenzie, J., Bossuyt, P., Boutron, I., Hoffman, T., Mulrow, D., Shamseer, L., Tetzlaff, J., Aki, E., Brennan, S., Chou, R., Glanville, J., Grimshaw, J., Hrobjartsson, A., Lalu, M., Li, T., Loder, E., Mayo-Wilson, E., McDonald, S., McGuinness, L., Stewards, L., Thomas, J., Tricco, A., Welch, V., Whiting, P., & Moher, D. (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

Perini, S., Titov, N., & Andrews, G. (2009). Clinician-assisted internet-based treatment is effective for depression: Randomized controlled trial. *Australian & New Zealand Journal of Psychiatry, 43 (6)*, 571-578. https://doi.org/10.1080/00048670902873722

Potts, S., Lewin, R., Fox, K., & Johnstone, E. (1999). Group psychological treatment for chest pain with normal. *QJM: An International Journal of Medicine, 92 (2)*, 81-86. https://doi.org/10.1093/qjmed/92.2.81

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

Preschl, B., Maercker, A., & Wagner, B. (2011). The working alliance in a randomized controlled trial comparing online with face-to-face cognitive-behavioral therapy for depression. *BMC Psychiatry, 11*, 189. https://doi.org/10.1186/1471-244X-11-189

RAND. (1992). *36-Item Short Form Survey (SF-36).* https://www.rand.org/health-care/surveys_tools/mos/36-item-short-form.html

Reneflot, A., & Evensen, M. (2012). Unemployment and psychological distress among young adults in the Nordic countries: A review of the literature. *International Journal of Social Welfare, 23 (1)*, 3-15. https://doi.org/10.1111/ijsw.12000

Richards, A., Barkham, M., Cahill, J., Richards, D., Williams, C., & Heywood, P. (2003). PHASE: A randomised, controlled trial of supervised self-help cognitive behavioural therapy in primary care. *British Journal of General Practice, 53 (495)*, 764-770.

Richards, D., Murphy, T., Vigano, N., Timulak, L., Doherty, G., Sharry, J., & Hayes, C. (2016). Acceptability, satisfaction and perceived efficacy of “Space from Depression" an internet-delivered treatment for depression. *Internet Interventions, 5*, 12-22. https://doi.org/10.1016/j.invent.2016.06.007

Richards, D., Timulak, L., & Hevey, D. (2012). A comparison of two online cognitive-behavioural interventions for symtoms of depression in a student population: The role of therapist responsiveness. *Counselling and Psychotherapy Research, 13 (3)*, 184-193. https://doi.org/10.1080/14733145.2012.733715

Richards, D., Timulak, L., O'Brien, E., Hayes, C., Vigano, N., Sharry, J., & Doherty, G. (2015). A randomized controlled trial of an internet-delivered treatment: Its potential as a low-intensity community intervention for adults with symptoms of depression. *Behaviour Research and Therapy, 75*, 20-31. https://doi.org/10.1016/j.brat.2015.10.005

Robinson, E., Titov, N., Andrews, G., McIntyre, K., Schwencke, G., & Solley, K. (2010). Internet treatment for generalized anxiety disorder: A randomized controlled trial comparing clinician vs. technician assistance. *PLoS ONE, 5 (6)*, e10942. https://doi.org/10.1371/journal.pone.0010942

Rollman, B., Herbeck Belnap, B., Mazumdar, S., Abebe, K., Karp, J., Lenze, E., & Schulberg, H. (2017). Telephone-delivered stepped collaborative care for treating anxiety in primary care: A randomized controlled trial. *Journal of General Internal Medicine, 32*, 245-255. https://doi.org/10.1007/s11606-016-3873-1

Rozental, A., Castonguay, L., Dimidjian, S., Lambert, M., Shafran, R., Andersson, G., & Carlbring, P. (2018). Negative effects in psychotherapy: Commentary and

recommendations for future research and clinical practice. *British Journal of Psychiatry Open, 4 (4)*, 307-312. https://doi.org/10.1192/bjo.2018.42

Russell, A., Gaunt, D., Cooper, K., Barton, S., Horwood, J., Kessler, D., Metcalfe, C., Ensum, I., Ingham, B., Parr, J., Raj, D., & Wiles, N. (2020). The feasibility of low-intensity psychological therapy for depression co-occurring with autism in adults: The Autism Depression Trial (ADEPT) – A pilot randomised controlled trial. *Autism, 24 (6)*, 1360-1372. https://doi.org/10.1177/1362361319889272

Russell, A., Gaunt, D., Cooper, K., Horwood, J., Barton, S., Ensum, I., Ingham, B., Parr, J., Metcalfe, C., Raj, D., Kessler, D., & Wiles, N. (2019). Guided self-help for depression in autistic adults: the ADEPT feasibility RCT. *Health Technology Assessment, 23 (68)*, 1-94. https://doi.org/10.3310/hta23680

Saunders, R., Buckman, J. E., Cape, J., Fearon, P., Leibowitz, J., & Pilling, S. (2019). Trajectories of depression and anxiety symptom change during psychological therapy. *Journal of Affective Disorders, 249*, 327-335. https://doi.org/10.1016/j.jad.2019.02.043

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

Saunders, R., Cape, J., Fearon, P., & Pilling, S. (2016). Predicting treatment outcome in psychological treatment services by identifying latent profiles of patients. *Journal of Affective Disorders, 197*, 107-115. https://doi.org/10.1016/j.jad.2016.03.011

Saxon, D., Barkham, M., Foster, A., & Parry, G. (2016). The contribution of therapist effects to patient dropout and deterioration in the psychological therapies. *Clinical Psychology & Psychotherapy, 24 (3)*, 575-588. https://doi.org/10.1002/cpp.2028

Schermuly-Haupt, M.-L., Linden, M., & Rush, A. J. (2018). Unwanted events and side effects in cognitive behaviour therapy. *Cognitive Therapy Research, 42*, 219-229. https://doi.org/10.1007/s10608-018-9904-y

Schneider, A., Mataix-Cols, D., Marks, I., & Bachofen, M. (2005). Internet-guided self-help with or without exposure therapy for phobic and panic disorders. *Psychotherapy and Psychosomatics, 74 (3)*, 154-164. https://doi.org/10.1159/000084000

Sextl-Plotz, T., Steinhoff, M., Baumeister, H., Cuijpers, P., Ebert, D., & Zarski, A. (2024). A systematic review of predictors and moderators of treatment outcomes in internet- and mobile-based interventions for depression. *Internet Interventions, 37*, 100760. https://doi.org/10.1016/j.invent.2024.100760

Simon, G., Ludman, E., Tutty, S., Operskalski, B., & Von Korff, M. (2004). Telephone psychotherapy and telephone care management for primary care patients starting antidepressant treatment. *JAMA, 292 (8)*, 935-942. https://doi.org/10.1001/jama.292.8.935

Spitzer, R. L., Kroenke, K., Williams, J. B., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166 (10)*, 1092-1097. https://doi.org/10.1001/archinte.166.10.1092.

Stochl, J., Soneson, E., Stuart, F., Fritz, J., Walsh, A. E., Croudace, T., Hodgekins, J., Patel, U., Russo D., Knight, C., Jones, P., & Perez, J. (2021). Determinants of patient-reported outcome trajectories and symptomatic recovery in Improving Access to

Psychological Therapies (IAPT) services. *Psychological Medicine, 52 (14)*, 3231-3240. https://doi.org/10.1017/S0033291720005395

Teachman, B., White, B., & Lilienfeld, S. (2021). Identifying harmful therapies: Setting the research agenda. *American Psychological Association, 28 (1)*, 101-106. https://doi.org/10.1037%2Fcps0000002

Teo, C. (2021). *The impact of ethnicity on the trajectory of depression symptom change during psychological interventions.* [Doctoral dissertation, University College London]. UCL Discovery. https://discovery.ucl.ac.uk/id/eprint/10135909/

The EuroQol Group. (1990). EuroQol-a new facility for the measurement of health-related quality of life. *Health Policy, 16 (3)*, 199-208.

Thesen, T., Himle, J., Martinsen, E., Walseth, L., Thorup, F., Gallefoss, F., & Jonsbu, E. (2022). Effectiveness of internet-based cognitive behavioral therapy with telephone support for non-cardiac chest pain: Randomized controlled trial. *Journal of Medical Internet Research, 24 (1)*, e33631. https://doi.org/10.2196/33631

Thorpe, G. (2017). *Identifying and predicting deterioration during psychotherapeutic interventions.* [Doctoral dissertation, University of Sheffield]. White Rose eTheses. https://etheses.whiterose.ac.uk/2160/2/Thorpe,_Germaine.pdf

Titov, N., Andrews, G., Davies, M., McIntyre, K., Robinson, E., & Solley, K. (2010). Internet treatment for depression: A randomized controlled trial comparing clinician vs. technician assistance. *PLoS ONE, 5 (6)*, e10939. https://doi.org/10.1371/journal.pone.0010939

Trombello, J. M., South, C., Sanchez, A., Kahalnik, F., Kennard, B. D., & Trivedi, M. (2020). Two trajectories of depressive symptom reduction throughout behavioral activation

teletherapy among underserved, ethnically diverse, primary care patients: A VitalSign6 report. *Behaviour Therapy, 51 (6)*, 958-971. https://doi.org/10.1016/j.beth.2020.01.002

van Agteren, J., Iasiello, M., Lo, L., Batholomaeus, J., Kopsaftis, Z., Carey, M., & Kyrios, M. (2021). A systematic review and meta-analysis of psychological interventions to improve mental wellbeing. *Nature Human Behavaviour, 5*, 631-652. https://doi.org/10.1038/s41562-021-01093-w

van Beek, M., Oude Voshaar, R., Beek, A., van Zijderveld, G., Visser, S., Speckens, A., Batelaan, N., & van Balkom, A. (2013). A brief cognitive-behavioural intervention for treating depression and panic disorder in patients with non-cardiac chest pain: A 24 week randomized controlled trial. *Depression and Anxiety, 30 (7)*, 670-678. https://doi.org/10.1002/da.22106

Vazquez, F., Torres, A., Blanco, V., Diaz, O., Otero, P., & Hermida, E. (2012). Comparison of relaxation training with a cognitive-behavioural intervention for indicated prevention of depression in university students: A randomized controlled trial. *Journal of Psychiatric Research, 46 (11)*, 1456-1463. https://doi.org/10.1016/j.jpsychires.2012.08.007

Vernmark, K., Lenndin, J., Bjarehed, J., Carlsson, M., Karlsson, J., Oberg, J., & Andersson, G. (2010). Internet administered guided self-help versus individualized e-mail therapy: A randomized trial of two versions of CBT for major depression. *Behaviour Research and Therapy, 48 (5)*, 368-376. https://doi.org/10.1016/j.brat.2010.01.005

Wagner, B., Horn, A., & Maercker, A. (2014). Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial.

Journal of Affective Disorders, 152-154, 113-121. https://doi.org/10.1016/j.jad.2013.06.032

Weitz, E. S., Hollon, S. D., Twisk, J., van Straten, A., Huibers, M. J., David, D., DeRubeis, R., Dimidjian, S., Dunlop, B., Cristea, I., Faramarzi, M., Hegerl U., Jarrett, R., Kheirkhah, R., Kennedy, S., Mergl, R., Miranda, J., Mohr, D., Rush, J., Segal, Z., Siddique, J., Simons, A., Vittengl, J., & Cuijpers, P. (2015). Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs pharmacotherapy: An individual patient data meta-analysis. *JAMA Psychiatry, 72 (11)*, 1102-1109. https://doi.org/10.1001/jamapsychiatry.2015.1516

Whitfield, G., Williams, C., & Shapiro, D. (2001). Assessing the take up and acceptability of a self-help room used by patients awaiting their initial outpatient appointment. *Behavioural and Cognitive Psychotherapy, 29 (3)*, 333-343. https://doi.org/10.1017/S135246580100306X

Williams, C., Wilson, P., Morrison, J., McMahon, A., Andrew, W., Allan, L., McConnachie, A., McNeill, Y., & Tansey, L. (2013). Guided self-help cognitive behavioural therapy for depression in primary care: A randomised controlled trial. *PLoS ONE, 8 (1)*, e52735. https://doi.org/10.1371/journal.pone.0052735

Zagorscak, P., Heinrich, M., Sommer, D., Wagner, B., & Knaevelsrud, C. (2018). Benefits of individualized feedback in internet-based interventions for depression: A randomized controlled trial. *Psychotherapy and Psychosomatics, 87 (1)*, 32-45. https://doi.org/10.1159/000481515

Appendix A: PRISMA 2020 Checklist

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 1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Page 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 4-6
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 6-7
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Page 9-11
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 7
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Page 7-9
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 9-11
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 12
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 12
	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 12
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
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 12-13
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 9-13
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Page 12-13
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Page 12
	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 12-13
	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
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Page 12

PRISMA 2020 Checklist

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 13
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	N/A
Study characteristics	17	Cite each included study and present its characteristics.	Page 17-26
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Page 30-31
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 27-30
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Page 30-31
	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 27-30
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Page 27-30
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Page 32-36
	23b	Discuss any limitations of the evidence included in the review.	Page 36-37
	23c	Discuss any limitations of the review processes used.	Page 36-37
	23d	Discuss implications of the results for practice, policy, and future research.	Page 37-38
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 7
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Page 7
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Page 8
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	ii
Competing interests	26	Declare any competing interests of review authors.	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.	ii

Appendix B: CASP Checklist

CASP Checklist:

For Randomised Controlled Trials (RCTs)

Reviewer Name:	
Paper Title:	
Author:	
Web Link:	
Appraisal Date:	

During critical appraisal, never make assumptions about what the researchers have done. If it is not possible to tell, use the "Can't tell" response box. If you can't tell, at best it means the researchers have not been explicit or transparent, but at worst it could mean the researchers have not undertaken a particular task or process. Once you've finished the critical appraisal, if there are a large number of "Can't tell" responses, consider whether the findings of the study are trustworthy and interpret the results with caution.

Section A Is the basic study design valid for a randomised controlled trial?	
1. Did the study address a clearly formulated research question?	☐ Yes ☐ No ☐ Can't Tell
CONSIDER: *Was the study designed to assess the outcomes of an intervention?* *Is the research question 'formulated' in terms of:* • *Population studied* • *Intervention given* • *Comparator chosen* • *Outcomes measured?*	
2. Was the assignment of participants to interventions randomised?	☐ Yes ☐ No ☐ Can't Tell
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?*	
3. Were all participants who entered the study accounted for at its conclusion?	☐ Yes ☐ No ☐ Can't Tell
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?*	
Section B Was the study methodologically sound?	
4. (a) Were the participants 'blind' to intervention they were given?	☐ Yes ☐ No ☐ Can't Tell
(b) Were the investigators 'blind' to the intervention they were giving to participants?	☐ Yes ☐ No ☐ Can't Tell

(c) 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?	☐ Yes ☐ No ☐ Can't Tell
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?*	
6. Apart from the experimental intervention, did each study group receive the same level of care (that is, were they treated equally)?	☐ Yes ☐ No ☐ Can't Tell
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?*	
Section C: What are the results?	
7. Were the effects of intervention reported comprehensively?	☐ Yes ☐ No ☐ Can't Tell
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?*	

8. Was the precision of the estimate of the intervention or treatment effect reported?	☐ Yes ☐ No ☐ Can't Tell
CONSIDER: • *Were confidence intervals (CIs) reported?*	
9. Do the benefits of the experimental intervention outweigh the harms and costs?	☐ Yes ☐ No ☐ Can't Tell
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.)*	
Section D: Will the results help locally?	
10. Can the results be applied to your local population/in your context?	☐ Yes ☐ No ☐ Can't Tell
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?*	
11. Would the experimental intervention provide greater value to the people in your care than any of the existing interventions?	☐ Yes ☐ No ☐ Can't Tell
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?*	

APPRAISAL SUMMARY: *List key points from your critical appraisal that need to be considered when assessing the validity of the results and their usefulness in decision-making.*		
Positive/Methodologically sound	Negative/Relatively poor methodology	Unknowns

Referencing recommendation:

CASP recommends using the Harvard style referencing, which is an author/date method. Sources are cited within the body of your assignment by giving the name of the author(s) followed by the date of publication. All other details about the publication are given in the list of references or bibliography at the end.

Example:

Critical Appraisal Skills Programme (2024). CASP (insert name of checklist i.e. randomised controlled trials (RCTs) Checklist.) [online] Available at: insert URL. Accessed: insert date accessed.

Creative Commons

Need further training on evidence-based decision making? Our online training courses are helpful for healthcare educational researchers and any other learners who:

- Need to critically appraise and stay abreast of the healthcare research literature as part of their clinical duties.
- Are considering carrying out research & developing their own research projects.
- Make decisions in their role, whether that be policy making or patient facing.

Benefits of CASP Training:

⇒ Affordable – courses start from as little as £6
⇒ Professional training – leading experts in critical appraisal training
⇒ Self-directed study – complete each course in your own time
⇒ 12 months access – revisit areas you aren't sure of and revise
⇒ CPD certification- after each completed module

Scan the QR code below or visit https://casp-uk.net/critical-appraisal-online-training-courses/ for more information and to start learning more.

Chapter Two: Empirical Study

Prevalence and Types of Symptomatic Deterioration after Low Intensity CBT in Primary Care Services

Abstract

Background: Low-intensity cognitive-behavioural therapy (LICBT) is a brief psychological intervention widely used to treat mild to moderate common mental health conditions. Although there is a substantial body of evidence on its effectiveness, there is limited research exploring poor outcomes, such as deterioration, including their prevalence and associated predictors.

Methods: A retrospective cohort study was conducted using data from 16 NHS Talking Therapies services across England. A machine learning approach was used to model counterfactual symptom trajectories in the absence of psychological treatment (synthetic controls), enabling a comparison between actual (observed) and the counterfactual outcome. Using this comparison, two types of deterioration were identified: iatrogenic harm (type 1) and natural illness progression (Type 2). Sociodemographic and clinical characteristics were compared in subgroup analyses using chi-square and t-tests to investigate common factors.

Results: The prevalence rates of deterioration using synthetic controls were between 16.4% and 16.9%, with type 1 deterioration accounting for 10.1%–12.3%. These rates were significantly higher than those found using the RCI criteria (3.6%-6.4%). Participants were more likely to deteriorate if they had a mixed diagnosis, symptomatic comorbidity, functional impairment, were unemployed, and living in socioeconomically deprived neighbourhoods. Other common characteristics associated with specific deteriorating symptoms and deterioration types were also identified.

Conclusions: Although, results should be interpreted with caution as they are limited to a specific setting and treatment, findings suggests that deterioration may be more prevalent than previously recognised and more likely to be due to iatrogenic reactions to therapy. Future research should validate these findings and explore mechanisms underpinning deterioration.

Keywords: Low-intensity cognitive behavioural therapy, machine learning, deterioration

Practitioner Points

- Adverse outcomes after LICBT, though uncommon, are clinically significant and warrant closer monitoring.
- Clear information about the potential benefits and risks of therapy could support patients to make informed decisions.
- Patients with mixed depression and anxiety diagnoses, symptomatic comorbidities, high functional impairment, and who are unemployed and living in socioeconomically deprived neighbourhoods may be at higher risk of adverse outcomes.
- Use of routine outcome monitoring could aid in the identification of early signs of deterioration and prevent poor treatment outcomes.
- Feedback-informed treatment methods could increase the intervention's effectiveness and reduce the likelihood of therapeutic iatrogenic effects.

Introduction

Historically, clinical trials and studies of medical interventions are required to follow bioethical and regulatory procedures that include the reporting of adverse incidents and negative reactions to treatment (Shorter, 2008). The mandatory monitoring of adverse effects in mental health research is justified based on the argument that all interventions that have the potential to be beneficial also have the potential to be harmful (Berk & Parker, 2009).

There is evidence that specific psychotherapeutic interventions, misapplication of evidence-based practice and treatment deviation have the potential to be harmful (McKay & Jensen-Doss, 2021). Some examples of potentially harmful psychological interventions include the 'Scared Straight' programme that exposed high-risk youth to prison life and resulted in increased offending (Petrosino, Turpin-Petrosino, & Buehler, 2002), and Critical Incident Stress Debriefing (CISD) which had the goal to reduce PTSD symptoms and resulted in worsening symptoms (Berk & Parker, 2009), with a meta-analysis finding evidence for harm as moderate in the Scared Straight programme and as strong in CISD (Williams et al., 2020).

Psychotherapy research often focuses on the benefits, as is the case of non-inferiority and superiority trials (Berk & Parker, 2009), and the reporting of adverse effects is often overlooked (Lilienfeld, 2007). Systematic reviews have found incidence reports of adverse events in 21% (Linden & Schermuly-Haupt, 2014) and 60% of psychotherapy trials (Vaughan et al., 2014). This limitation is partly due to a lack of consensus or guidelines on how define, classify and assess adverse outcomes in psychotherapy (Linden, 2012; Rozental et al., 2018), even though the Consolidated Standards for Reporting Trials (CONSORT) recommends all researchers to define, monitor, measure and report adverse outcomes (Ioannidis et al., 2004).

Common adverse outcome definitions used in the wider literature include treatment non-response (no improvement) and deterioration (worsening symptom progression) (Linden,

2012). Other common definitions focused on adverse events include unwanted events (adverse events occurring parallel with treatment), adverse treatment reaction (adverse event caused by the correct treatment), malpractice reactions (caused by misapplied treatment) and unethical conduct (Klatte, et al., 2018; Linden et al., 2018; Schermuly-Haupt, Liden, & Rush, 2018).

Current evidence-based interventions have been found to have the potential to be harmful with studies finding adverse effects prevalence rates of 5.2% (Crawford et al, 2016), 41% (Lorenz, 2020), and 57.8% (Strauss et al, 2021). Those focusing on CBT have found rates of 33% (Balder, Linden, & Rose, 2024), and 43% (Schermuly-Haupt, Linden, & Rush., 2018). These wide variations are also related to the type of adverse effects, as most psychological interventions can cause adverse treatment reactions (i.e. anxiety, discomfort, or distress), and while this can be unavoidable, it is important to acknowledge and differentiate from other adverse effects (i.e. malpractice, treatment non-response, and deterioration).

Regarding adverse treatment outcomes, research suggests that between 40% to 60% of patients experience treatment non-response, by not reaching a clinically significant recovery criterion (Curran, et al., 2019), and between 5% and 20% of patients experience clinical deterioration (Klatte, et al., 2018). These estimates vary depending on the measurements used and differences in clinical populations (Curran, et al., 2019), alongside the inconsistent definitions and reporting of adverse effects in the psychological literature (Klatte, et al., 2018).

It is also important to determine whether or not adverse outcomes may be attributable to an intervention or whether they may be related to wider contextual or biopsychosocial factors (Duggan et al., 2014). Thus, it is plausible that there might be two possible broad types of adverse outcomes: those that occur as a result of iatrogenic harm from an intervention (type 1) and those that reflect a natural progression towards chronic symptoms (type 2). However, the limited research on adverse events in psychotherapy has not yet yielded clear criteria on how

to define, operationalise, and identify these adverse outcomes. Hence, their likely prevalence is relatively unclear, and the specific prevalence of type 1 and 2 adverse outcomes is unknown.

The development of methods to define such adverse outcomes and to understand the common characteristics of patients who are likely to have type 1 or type 2 adverse outcomes, as research also suggests that patient characteristics could affect the likelihood of adverse effects, yet there is limited consistency regarding specific associations (Barkham, Lutz, & Castonguay, 2021).

Research on adverse treatment outcomes in psychotherapy could help inform treatment allocation, advice service planning, and support clinicians to respond in a proactive way in order to monitor progress and adjust treatment plans to prevent deterioration. Ethical implications include the promotion of patient choice and informed consent, as safety information on the potential benefits and risks is equally as important as knowledge on clinical effectiveness to make informed decisions for any proposed intervention (Ernst, 2001).

Aims

The aim of this empirical study is to investigate the prevalence of deterioration, defined as worsening symptom progression, after low intensity cognitive-behavioural therapy (LICBT) delivered in routine psychological care. Specific objective are as follows:

1. To identify two classes of cases using a routine-practice dataset: those with type 1 and type 2 deterioration of symptoms. Where type 1 iatrogenic deterioration is worsening symptom progression most likely caused by the LICBT intervention, and type 2 chronic deterioration is most likely natural illness progression towards chronicity.
2. To investigate if patients with type 1 versus type 2 deterioration have different sociodemographic and clinical characteristics.

Research Questions

- What is the prevalence of adverse outcomes after LICBT delivered in primary care?
- Of these, what is the prevalence of Type 1 and Type 2 deterioration?
- What are the sociodemographic and clinical characteristics of cases with Type 1 and Type 2 deterioration?

Methods

Design

This was a retrospective, observational cohort study using archival data from patients who accessed NHS Talking Therapies (NHS-TT) services in England[1].

Setting

Data were gathered from 16 NHS-TT services managed by 8 National Health Service (NHS) trusts across England, covering East Riding, Barnsley, Trafford, Tameside and Glossop, Stockport, Oldham, Rochdale, Middleton, Heywood, Bury, Cheshire and Wirral, Cambridge, and London. This dataset includes anonymised records of all the patients who accessed these services between 2014 and 2017. Original data collection and analysis was approved by the London City & East NHS Research Ethics Committee (Ref: 15/LO/2200) and this current study was approved by the University of Sheffield Research Ethics Committee (Appendix A).

Low Intensity Cognitive-Behavioural Therapy

NHS-TT services deliver psychological interventions for common mental health problems in adults following a stepped care model. Access to these services can be gained via a professional referral from voluntary, community, primary or secondary care, or by self-referral. Then, a person-centred assessment (step 1) is used to identify a potential diagnosis and

[1] Formerly known as Improving Access to Psychological Therapies (IAPT) services.

treatment plan (Clark, 2011). Available treatments are organised into different levels of intensity, referred to as low intensity (step 2) and high intensity (step 3) interventions.

Most patients with mild-to-moderate symptoms of depression or anxiety are offered a step 2 low intensity intervention based on principles of cognitive behavioural therapy (CBT) delivered by a Psychological Wellbeing Practitioner (PWP). This type of intervention includes psychoeducational groups, computerized CBT (cCBT), and guided self-help (GSH; Clark, 2011). Guided self-help is the most common intervention, it is highly structured, guided by self-help booklets, and based on national protocols in up to 8 sessions (Thompson, Parker & Cave, 2021). If this level of intervention is unsuccessful, patients can be stepped-up and access step 3 high intensity interventions delivered by a trainee or qualified high intensity therapist within the service (National Collaborating Centre for Mental Health, 2018).

Following clinical guidelines, LICBT interventions are aimed towards mild to moderate depression, GAD, panic disorder, phobias, and OCD. High intensity interventions are offered to those who remain symptomatic after LICBT, and also offered as a first-line treatment for patients with conditions for which clinical guidelines only recommend high intensity interventions, such as social anxiety disorder, post-traumatic stress disorder (PTSD), and body dysmorphic disorder (BDD) (National Collaborating Centre for Mental Health, 2018).

LICBT focuses on psychoeducation, aiming to support patients in learning about maintaining factors common to all mental health disorders through a CBT framework. Patients are encouraged to practice cognitive behavioural coping skills using treatment protocols such as behavioural activation for depression, worry management for generalised anxiety disorder (GAD), graded exposure for phobias and panic disorder, exposure and response prevention (ERP) for obsessive-compulsive disorder (OCD), along with other skills such as applied relaxation, cognitive restructuring, sleep hygiene, problem solving, and relapse prevention

(Bennett-Levy et al., 2010). Protocols are evidence-based, supported by evidence of brief CBT interventions (Shafran et al., 2021), and meta-analyses suggest that LICBT interventions are effective in treating depression and anxiety (Andrews et al., 2010; Cuijpers et al., 2010).

There is evidence that NHS-TT services are effective, for example a systematic review and meta-analysis associated NHS-TT interventions with moderate pre-post treatment effect sizes in the work and social adjustment scales (WSAS), and large effect sizes on anxiety and depression measurements (Wakefield, 2021). Meta-analyses focusing have also found low intensity interventions effective in the treatment of depression and anxiety (Bower et al., 2013; Gellatly et al, 2007; Powell, et al., 2024). Last year, NHS Talking Therapies reported improvement, recovery, and deterioration rates in LICBT, including GSH (54.8%, 42%, and 7.3%), groups (43%, 32.8%, and 7.1%), and cCBT (45.1%, 36.7%, and 7.7%) (NHS, 2023).

Participants

The source database includes clinical records for 146,078 referrals, of whom 102,026 accessed low and/or high intensity CBT interventions, with 76,962 who accessed LICBT.

The present study sample only selected data for patients that accessed LICBT and attended at least two sessions of this treatment. Applying this selection criteria resulted in the inclusion of 63,481 LICBT cases. The rationale for including cases with at least two sessions is to ensure that at least one additional outcome measure was available. Since session 1 questionnaires capture pre-treatment symptom severity information during the two weeks prior to the appointment, the last available measure would hence be used as the final treatment outcome.

Data Sources and Outcome Measures

Demographic data routinely collected by NHS-TT includes the following: age, gender, ethnicity, disability, employment status, use of psychopharmacological medication, primary

diagnosis, and index of multiple deprivation (IMD) linked to the patient's neighbourhood (derived from home postcodes). The clinical outcome measures include the Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer & Williams, 2001), the Generalised Anxiety Disorder-7 (GAD-7; Spitzer, Kroenke, & Williams, 2006), and the Work and Social Adjustment Scale (WSAS; Mundt et al., 2002) (see Appendix B).

The PHQ-9 is a self-reported 9-item questionnaire developed to screen and measure depression severity with good internal consistency of 0.86, specificity of 0.88, sensitivity of 0.88 and high reliability (Kroenke et al., 2001). It has a recommended clinical cut-off score of ≥10 (Levis, Benedetti & Thombs, 2019) and a reliable change index (RCI) of ≥6 (Gyani et al., 2013).

The GAD-7 is a self-administered 7-item questionnaire developed to screen and assess GAD severity with a specificity of 0.82, a sensitivity of 0.89 (Spitzer et al., 2006), a reliability of 0.88, with good internal consistency, validity, and reliability (Johnson et al, 2019). It has a recommended clinical cut-off score of ≥8 and a RCI of ≥4 (Gyani et al., 2013).

The WSAS is an 8-item scale of functional impairment, which assesses difficulties in functioning across several areas of daily activities and relationships. The scale has acceptable internal consistency ranging from 0.70 to 0.94, test-retest correlation of 0.73, with correlations with depression (0.76) and obsessive-compulsive (0.61) symptom severity. Scores <10 indicating subclinical symptoms, from 10-20 significant functional impairment and >20 moderately severe to severe functional impairment (Mundt et al., 2002).

Sample Characteristics

A summary of the demographic and clinical characteristics of the 63,481 participants included in the analysis can be found in the following table:

Table 1.

Sample Characteristics

Characteristics	Descriptive Statistics
Age (mean, SD)	40.50 (15.19)
Gender	65.0% Female, 35.0% Male
Ethnicity	84.7% White British, 15.3% Other
Employment	78.2% Active, 21.8% Unemployed
Medication	45.8% NP, 45.2% PT, 9.0% PNT
Self-reported LTC	69.3% None, 30.7% Yes
Self-reported Disability	85.3% None, 14.7% Yes
IMD (mean, SD)	4.95 (2.78)
Primary Diagnosis	33.7% Affective Disorder, 27% anxiety disorder, 17% Anxiety and Depression (mixed), 12.3% Other.
Baseline PHQ-9 (mean, SD)	14.7 (6.14)
Baseline GAD-7 (mean, SD)	13.4 (4.99)
Baseline WSAS (mean, SD)	19.36 (9.7)
Intervention	89.3% GSH, 7.5% PG, 2.1% cCBT, 1.1% Other
Sessions Taken (mean, SD)	7.4 (4.84)
Reason for End of Care	5.6% Referred (external), 0.5% Stepped Up, 25.1% Dropped Out, 67.4% Completed Treatment.

Note: LTC= long-term condition. NP= not prescribed, PT= prescribed and taking, PNT= prescribed and not taking. IMD= index of multiple deprivation. GSH= guided self-help, PG= psychoeducational group

Data Analysis

The data analysis plan proceeded in three stages. First, demographic characteristics were reported using descriptive statistics. Second, the classification of type 1 and 2 deterioration was achieved using a supervised machine learning approach called extreme

gradient boosting (XGBoost). Third, demographic characteristics were summarised and compared statistically between cases with type 1 vs type 2 deterioration. The statistical methods applied in stages 2 and 3 are explained below.

Stage 2: Machine Learning Approach

Machine learning is a data mining process that enables the development of statistical algorithms that help to perform prediction and classification tasks (Delgadillo & Atzil-Slonim, 2023). This process aims to discover relationship patterns observed in available "training data" to generate outputs in the form of a prediction or classification (Breiman, 2001).

There are two broad types of machine learning approaches, *supervised* and *unsupervised* models. Supervised machine learning is used to analyse datasets that have both *features* (sample characteristics) and *labels* (a dependent variable of interest, such as a diagnostic label or clinical outcome). This methodology trains a prediction model to recognise the pattern of features associated with the label (Hastie, Friedman & Tibshirani, 2009). On the other hand, unsupervised machine learning is used to discover subgroups or clusters of cases with similar features, without pre-specifying any labels, thus developing a classification scheme through a data mining process (Hastie, Friedman & Tibshirani, 2009).

Prediction and Classification Methodology

The aim of the study involved identifying cases who had symptomatic deterioration, which can be framed as a classification task. Furthermore, we expected that two classes of deterioration may be present in routine care samples, type 1 and type 2. In order to identify and differentiate these classes of deterioration, we developed machine learning classifiers.

The independent variables (features) used as predictors in this model included the PHQ-9, GAD-7 and WSAS questionnaires, as well as the time (in days) spent on the waiting list prior to starting LICBT treatment. The dependent variable (label) represents the level of

symptom severity measured at the first LICBT session, after a waitlist period. Given that the duration of time on waiting list varies considerably from case-to-case, this approach enabled us to train an algorithm that models the symptom trajectory changes in the absence of psychological treatment (i.e., while on waiting list), as a function of the initial pattern of symptoms and waiting time. The trained model was then used to predict a counterfactual outcome, which refers to the level of symptom severity that we might have expected to observe at the end of each patient's treatment if they had remained on waiting list rather than having treatment. This approach is also called a synthetic control, as it is a simulation of a waitlist control outcome for a patient who in fact received treatment, thus enabling a counterfactual comparison in the absence of a randomised control group (Bouttell et al., 2018).

For example, consider a patient initially assessed on day 0, who was on waiting list and started LICBT on day 20, and who accessed 6 appointments, finishing their last session on day 62. This patient's data at timepoints 0 and 20 will be used (pooled among data from other patients) to train a prediction model. Next, the trained model is used to predict this patient's counterfactual symptom score (e.g., synthetic waitlist control) at day 62. This counterfactual prediction enables us to estimate what kind of score this patient would have had if they had not received treatment for all that time, thus modelling the natural progression of the condition by learning from patterns of data observed in other cases with variable waiting durations. If the patient's actual (observed) post-treatment score (day 62) is worse than their baseline measure (day 0), we would classify this as an adverse outcome. Furthermore, if the actual score is significantly worse that the counterfactual score, we would classify this as type 1 deterioration (the patient is worse off after treatment than they would have been without treatment). Otherwise, if the actual score is not significantly worse than the counterfactual score but worse than baseline, we would classify this as a type 2 deterioration (the patient's condition followed a natural progression towards chronicity, and treatment was ineffective to reverse this).

We used the reliable change index (RCI; Jacobson & Truax, 1991) threshold to differentiate between classes of outcomes, as shown in the table below. The RCI is a well-established methodology to assess treatment response in psychotherapy (Jacobson & Truax, 1991), and it is recommended for use in research and clinical practice (Barkham, Lutz, & Castonguay, 2021). In this study, we were particularly interested in quantifying type 1 and type 2 deterioration, as operationalised below, but we nevertheless reported the prevalence of other outcomes (e.g., improvement, no reliable change) for context. In order to determine if the actual post-treatment score is significantly worse than the counterfactual score, the post-treatment score should be higher than the counterfactual score by a magnitude greater than 1.96 * SE, where SE = the standard error of the counterfactual predictions calculated in the test sample (the machine learning cross-validation process is described in more detail below).

Table 2.

Operational definition of clinical outcomes.

Outcome Class	**Change in symptom severity between baseline and post-treatment score**	**Severity of post-treatment score compared to counterfactual score**
Improved	Decrease > RCI	
No reliable change	Change < RCI	
Deterioration- type 1	Increase > RCI	Significantly worse 1.96 * SE
Deterioration- type 2	Increase > RCI	Not significantly worse

Note: Deterioration type 1 (iatrogenic- worse symptoms after treatment than without treatment) and type 2 (chronic- natural progression towards chronicity, and treatment was ineffective to reverse this).

Decision Trees and Extreme Gradient Boost Algorithm

Decision trees are supervised machine learning algorithms, which are capable of modelling nonlinear relationships between independent-dependent variables and complex interactions between independent variables (Golino & Gomes, 2016). Their general purpose

lies with prediction and classification, and its main characteristic is recursive partitioning of a dataset to organise cases into more granular subsets with similar features (de Ville, 2013). This is advantageous compared to traditional regression models which only model linear relationships, unless interaction terms (e.g., exponential terms, or interactions between different predictors) are explicitly entered by the data analyst. Since, in theory, many possible interactions, linear and nonlinear relationships could exist, traditional regression modelling requires a pre-specification of such relationships a priori. An advantage of decision trees, like other forms of machine learning, is that complex relationships do not need to be specified a priori, because the data mining process seeks to discover such relationships in a data-driven way. One key limitation in decision trees, however, is that they are prone to overfitting, especially when trees include subsets of cases (e.g., child nodes) with small samples. Ensemble techniques can be used to minimize overfitting, by pooling information across multiple decision trees (Golino & Gomes, 2016). Gradient boosting is a common ensemble technique where trees are generated sequentially, with each tree identifying and correcting the prediction error of its predecessor, collectively creating a model that pools predictions across multiple decision trees. Extreme gradient boost (XGBoost) is a widely used model that uses gradient boosting in large and complex datasets with highly effective predictive performance (Chen & Guestrin, 2016)

Once a model is trained, it stores a fixed set of decision and classification rules, which can be applied to new cases to predict an outcome of interest (Probst, Wright & Boulesteix, 2019).

Machine Learning Pipeline

A machine learning pipeline refers to a series of interconnected processes that enable the training and evaluation of an algorithm used to solve prediction of classification problems. A development pipeline includes: (1) sample size calculation, (2) data pre-processing methods,

(3) hyper-parameter selection, (4) model training, (5) model testing, and (6) clinical evaluation (Delgadillo & Atzil-Slonim, 2022). All of these steps of the pipeline are described below.

(1) Sample Size Calculation

A sample size calculation was performed following the methodology proposed by Riley et al (2018) for the development of clinical prediction models using continuous outcome variables. This is a specific form of sample size calculation for machine learning models, since it includes an adjustment for cross-validation shrinkage (e.g., the expected level of generalizability error when applying a trained model in a statistically independent test dataset). The planned statistical model introduced 4 variables as predictors (baseline depression, anxiety, functional impairment, time in waiting list), with post-treatment symptom severity as a dependent variable. Using 4 parameters, with an expected effect size of r^2= 0.30 informed by a prior study in this setting (Delgadillo et al., 2020) resulted in a minimum sample size of 54 participants. As this study used a machine learning approach, we required two statistically independent partitions of the dataset to train and cross-validate the prediction model, thus 108 participants were needed overall (randomly split between training-testing partitions). The available database includes 63,481 participants, well above the required sample size.

(2) Data Pre-Processing

Cases for this analysis were selected using the inclusion criteria described above. The dataset was randomly split (50:50) into training and test partitions, in order to apply an external cross-validation approach to evaluate the model's prediction accuracy. All variables were inspected to identify missing data-points. Missing data were imputed separately in the training and test partitions. Imputation was carried out by aggregating 25 iterations into a single imputed dataset using the Markov Chain Monte Carlo method (Schunk, 2008).

(3) Hyperparameter Tuning

Hyperparameters refer to different "settings" that can be applied to train a machine learning algorithm. Different configurations of these settings can affect the performance of the training process, for instance resulting in overfitting and poor generalisability. Hence, the selection of an optimal configuration of hyperparameters is desirable, which is a process referred to as "hyperparameter tuning". Tuning can be automated, using a grid search process that test multiple parameter configurations systematically on a grid, while manual tuning is unsystematic and can be bias-prone (Quemy, 2019). In this study, we applied a grid searching method, which automatically explores a search space of potential hyperparameters (e.g., number of observations drawn randomly for each tree, number of variables drawn randomly for each split, the splitting rule, etc.), building a series of models and comparing the models using out-of-bag sampling to derive optimal settings (Costa et al., 2018).

(4) Model Training

XGBoost was applied in the training partition, entering the following predictors: baseline PHQ-9, GAD-7, WSAS and time (in days) on waiting list prior to the first LICBT session. The dependent variables were the symptom scores measured at the first LICBT session. Separate models were trained to predict depression (PHQ-9) and anxiety (GAD-7) outcomes. We evaluated the performance of the model within the training sample by comparing the predicted vs. observed scores statistically (r correlation coefficient) and graphically (calibration plot).

(5) Model Validation

External cross-validation was performed in a statistically independent test sample. The trained algorithm was applied to predict therapy session 1 scores and to compare these with observed scores statistically (r correlation coefficient; standard error of prediction) and graphically (calibration plot). This enabled us to assess the accuracy of the prediction algorithm

and its generalizability in a new sample. It also enabled us to compute the prediction standard error (SE), which was applied to the classification rules (see table 2) to identify type 1 vs. type 2 deterioration.

(6) Clinical Evaluation

The trained model was applied in the test sample to predict the counterfactual post-treatment score for each patient (e.g. synthetic waitlist controls). We then applied the classification rules in Table 1 to enable us to report the overall proportions of cases in each of these adverse outcome classes.

Stage 3: Characteristics of Type 1 and Type 2 Deterioration Cases

We investigated prevalence rate for type 1 and type 2 deterioration in each of the measures of depression (PHQ-9) and anxiety (GAD-7). Then compared the clinical (diagnosis, baseline severity across all measures, use of medication, number of treatment sessions attended, and reason for end of care) and demographic characteristics (age, gender, ethnicity, disability, long-term conditions, and IMD) between cases classed as type 1 and type 2 deterioration through parametric statistical tests, using chi-square analysis for categorical variables, and independent t-tests for continuous variables, as all data was normally distributed.

Results

Deterioration in Depression (PHQ-9)

Overall pre-post treatment deterioration rates for depression simply based on the RCI for the PHQ-9 RCI measure were 3.7% in the training sample and 3.6% in the test sample. Both, the training and test samples, showed no reliable change in 49% of cases, while 47.3% showed reliable improvement.

The overall prevalence rates of deterioration based on the synthetic control method were 16.9% in the training sample and 16.8% in the test sample. Type 1 deterioration was found in 10.1% of the training sample and 10.2% of the test sample, while type 2 deterioration was found in 6.8% of the training sample and 6.6% of the test sample.

In the training sample, there were statistically significant differences in age, employment, medication, disability, IMD, diagnosis, baseline GAD-7 and WSAS scores, and reason for end of care. Participants with type 1 deterioration were older, unemployed, prescribed and taking medication, had a lower IMD, diagnosed with an affective disorder or mixed diagnosis, had higher GAD-7 and WSAS baseline scores, and had higher incidences of dropping out of treatment than participants with type 2 deterioration, as seen in table 3.

Table 3.

PHQ-9 Training Sample

Characteristic	**Type 1- Iatrogenic**	**Type 2- Chronic**	**P-value**
Age	mean=40.01 (14.42)	mean= 38.88 (9.55)	0.024
Gender	36% male, 64% female	34.9% male, 65.1% female	0.593
Ethnicity	82.5% White British, 17.5% Other	84.9% White British, 15.1% Other	0.132
Employment	37.1% Unemployed, 62.9% Active	25% Unemployed, 75% Active	<0.001
Medication	42.1% NP, 8.5% PNT, 49.4% PT	44.7% NP, 9.2% PNT, 40.5% PT	<0.001
LTC	65% No, 35% Yes	70.6% No, 29.4% Yes	0.236
Disability	78.4% No, 21.5% Yes	87.3% No, 12.7% Yes	0.026
IMD	mean=4.33 (2.71)	mean=5.07 (2.92)	<0.001
Diagnosis	41.8% affective disorder, 23.6% anxiety disorder,	35% affective disorder, 34.6% anxiety disorder,	<0.001

	30.4% mixed diagnosis, 0.8% other, 3.5% NOS.	24.5% mixed diagnosis, 1.3% other, 4.6% NOS.	
PHQ-9	mean=15.76 (5.09)	mean= 9.55 (4.96)	0.537
GAD-7	mean=13.36 (4.36)	mean=9.47 (4.66)	0.002
WSAS	mean=21.48 (9.41)	mean=17.22 (9.91)	0.016
Sessions	mean= 4.26 (2.62)	mean= 4.59 (2.58)	0.050
End of Care	13.6% referred, 40.3% dropped out, 44.9% completed, 1.2% stepped up	8.7% referred, 31.1% dropped out, 56.3% completed, 1% stepped up	<0.001

On the other hand, there were statistically significant differences in age, ethnicity, employment, medication, LTC, disability, IMD, diagnosis, baseline PHQ-9 scores, baseline GAD-7 and WSAS scores, and reason for end of care in the test sample. Participants with type 1 deterioration were younger, other non-White British ethnicity, unemployed, prescribed and taking medication, had a long-term condition, a disability, affective or mixed diagnosis, lower IMD, higher baseline scores in the PHQ-9, GAD-7, and WSAS, and higher incidence rates of dropping out of treatment than participants with type 2 deterioration, as seen below in table 4.

Table 4.

PHQ-9 Test Sample

Characteristic	**Type 1- Iatrogenic**	**Type 2- Chronic**	**P-Value**
Age	mean=39.21 (14.59)	mean= 39.58 (15.84)	0.002
Gender	35.4% male, 64.6% female	34.7% male, 65.3% female	0.582
Ethnicity	81.8% White British, 18.2% Other	87.2% White British, 12.8% Other	<0.001
Employment	36.2% Unemployed, 63.8% Active	18.1% Unemployed, 81.9% Active	<0.001

Medication	41.4% NP, 7.2% PNT, 51.3% PT	49.8% NP, 9.3% PNT, 40.9% PT	<0.001
LTC	65.3% No, 34.7% Yes	72.2% No, 27.8% Yes	<0.001
Disability	74.74% No, 25.26% Yes	88.31% No, 11.69% Yes	<0.001
IMD	mean=4.43 (2.71)	mean=5.22 (2.81)	<0.001
Diagnosis	40% affective disorder, 25.7% anxiety disorder, 29.5 mixed diagnosis, 1% other, 3.8 NOS	28.7% affective disorder, 41.3% anxiety disorder, 24.4% mixed diagnosis, 1% other, 4.7% NOS	<0.001
PHQ-9	mean=15.80 (5.10)	mean= 8.48 (4.59)	0.002
GAD-7	mean=14.50 (4.53)	mean=10.86 (4.89)	<0.001
WSAS	mean=21.79 (9.47)	mean=16.26 (9.88)	0.004
Sessions	mean= 4.19 (2.51)	mean= 4.42 (2.41)	0.124
End of Care	13.1% referred, 40.7% dropped out, 44.4% completed, 1.9% stepped up	6% referred, 34% dropped out, 59.3% completed, 0.7% stepped up	<0.001

Overall, there were statistically significant differences in age, employment, medication, disability, IMD, diagnosis, baseline GAD-7 and WSAS scores, and reason for end of care in both samples. While age differences varied, participants with type 1 deterioration on depression had higher rates of unemployment, prescribed and taking medication, disability, lower IMD, affective or mixed disorder diagnosis, higher baseline GAD-7 and WSAS scores, and higher treatment drop out incidence rates than participants with type 2 deterioration.

Deterioration in Anxiety (GAD-7)

Overall pre-post treatment deterioration rates for anxiety simply based on the RCI for the GAD-7 measure were 6.4% in the training sample and 6% in the test sample. No reliable

change was found in 36% of the training sample and in 35.8% of the test sample, while 57.5% of the training sample and 58.2% of the test sample showed reliable improvement.

The overall prevalence rates of deterioration based on the synthetic control methos were 16.9% in the training sample and 16.4% in the test sample. Type 1 deterioration was found in 11.6% of the training sample and 12.3% of the test sample, while type 2 deterioration was found in 5.3% of the training sample and 4.1% of the test sample.

In the training sample, statistically significant differences were found regarding age, gender, ethnicity, employment, long-term conditions, IMD, diagnosis, PHQ-9 and WSAS baseline scores, and reason for end of care. Participants with type 1 deteriorations were younger, female, other non-White British ethnicity, unemployed, had a long-term condition, lower IMD, mixed diagnosis, higher PHQ-9 and WSAS baseline scores, and higher incidences of dropping out of treatment than participants with type 2 deterioration, as seen below.

Table 5.

GAD-7 Training Sample

Characteristic	**Type 1- Iatrogenic**	**Type 2- Chronic**	**P-Value**
Age	mean=39.53 (14.33)	mean= 41.77 (16.19)	<0.001
Gender	34.4% male, 65.6% female	40.8% male, 59.3% female	0.005
Ethnicity	82.4% White British, 17.6% Other	86.3% White British, 13.7% Other	0.034
Employment	34.6% Unemployed, 65.4% Active	28.5% Unemployed, 71.5% Active	0.008
Medication	44.1% NP, 8.8% PNT, 47.1% PT	46.7% NP, 8.4% PNT, 44.9% PT	0.593
LTC	66.7% No, 33.3% Yes	62.1% No, 37.9% Yes	0.076
Disability	81.4% No, 18.6% Yes	81.3% No, 18.8% Yes	0.980

IMD	mean=4.42 (2.74)	mean=5.11 (2.95)	0.002
Diagnosis	40.4% affective disorder, 25.4% anxiety disorder, 29.8 mixed diagnosis, 1% other, 3.4 NOS	44.8% affective disorder, 23.8% anxiety disorder, 24.4% mixed diagnosis, 1% other, 6% NOS	<0.001
PHQ-9	mean=14.71 (5.45)	mean= 10.66 (6.05)	<0.001
GAD-7	mean=13.30 (4.28)	mean=7.73 (4.13)	0.179
WSAS	mean=20.87 (9.47)	mean=17.50 (10.41)	<0.001
Sessions	mean= 4.23 (2.39)	mean= 4.58 (2.55)	0.441
End of Care	12.4% referred, 40.7% dropped out, 45.7% completed, 1.2% stepped up	8.2% referred, 32.7% dropped out, 58.4% completed, 0.7% stepped up	<0.001

Results from the test sample show statistically significant differences in age, gender, ethnicity, employment, medication, disability, IMD, diagnosis, attended sessions, baseline scores in the PHQ-9. GAD-7 and WSAS, and reason for end of care. Participants with type 1 deterioration were younger, female, other non-White British ethnicity, unemployed, prescribed and taking medication, had a disability, mixed diagnosis, less attended sessions, higher baseline scores in the PHQ-9, GAD-7 and WSAS, and higher treatment drop out incidence rates than those with type 2 deterioration, as seen in the table below.

Table 6

GAD-7 Test Sample

Characteristic	**Type 1- Iatrogenic**	**Type 2- Chronic**	**P-Value**
Age	mean=39.41 (14.55)	mean= 41.04 (15.91)	<0.001
Gender	33.7% male, 66.3% female	38.4% male, 61.6% female	0.003

Ethnicity	82.1% White British, 17.9% Other	85.9% White British, 14.1% Other	0.003
Employment	33.1% Unemployed, 66.9% Active	20.6% Unemployed, 79.4% Active	<0.001
Medication	43.7% NP, 8.6% PNT, 47.6% PT	47.9% NP, 9.9% PNT, 42.2% PT	0.011
LTC	66.9% No, 33.1% Yes	69.3% No, 30.7% Yes	0.190
Disability	77.7% No, 22.3% Yes	84.8% No, 15.2% Yes	0.035
IMD	mean=4.56 (2.73)	mean=5.27 (2.76)	0.042
Diagnosis	39.1% affective disorder, 27.7% anxiety disorder, 29.1 mixed diagnosis, 0.9% other, 3.2 NOS	43.5% affective disorder, 25.9% anxiety disorder, 23.5% mixed diagnosis, 1% other, 6.1% NOS	<0.001
PHQ-9	mean=15.71 (5.74)	mean= 11.99 (6.21)	0.002
GAD-7	mean=13.44 (4.25)	mean=7.32 (3.96)	<0.001
WSAS	mean=21.10 (9.47)	mean=17.35 (10.31)	<0.001
Sessions	mean= 4.19 (2.38)	mean= 4.49 (2.57)	0.014
End of Care	11.3% referred, 41% dropped out, 46.3% completed, 1.4% stepped up	6.7% referred, 35.6% dropped out, 56.9% completed, 0.9% stepped up	<0.001

Overall, in both samples there were statistically significant differences regarding age, gender, ethnicity, employment, IMD, diagnosis, PHQ-9 and WSAS baseline scores, and reason for end of care. Participants with type 1 deterioration in anxiety symptoms tended to be younger, female, other non-White British ethnicity, unemployed, had lower IMD, mixed diagnosis, higher baseline scores in the PHQ-9 and WSAS, and had higher treatment drop out incidence rates than participants with type 2 deterioration.

Discussion

While adult psychotherapy deterioration prevalence rates vary widely between studies, common prevalence rates cited in the literature are between 5% and 10% (Lambert, 2013). Recent literature and systematic reviews have found variations to this range, including prevalence rates between 0% and 10% (Lazar, 2017), 0% and 13% with a mean of 4% (Cahill, Barkham, & Stiles, 2010), 9% and 17% (Thorpe, 2017), and of 3.79% (Klatte et al., 2023). Systematic reviews and meta-analyses focused on specific settings, diagnoses, or psychotherapy models have also found low deterioration rates, such as 1.5% in primary care (Cahill, Barkham, & Stiles, 2010), 5% in psychotherapy for depression (Cuijpers, 2021), and 9% in CBT (Schermuly-Haupt, Linden, & Rush, 2018).

In the present sample, simple calculations of deterioration rates using the RCI method were between 3.6% and 3.7% for depression and between 6% and 6.4% for anxiety, generally consistent with prior studies. Such calculations give the appearance that deterioration rates after psychological therapy are generally low.

However, these prevalence rates are likely to be inaccurate due to known issues related to the RCI method. This statistical approach uses group averages to estimate singular response curve trajectories resulting in a ceiling effect, where participants with severe symptomatology have an increased possibility to fall within the reliable change threshold and a decreased possibility to fall within the reliable deteriorated threshold (Lambert & Ogles, 2004). Because most psychometric measures have an upper limit in terms of severity, patients with high baseline severity scores cannot report past that upper limit in future sessions or post-treatment, for example: a patient with a score of 25 out of 27 in the PHQ-9 at the start of treatment and a score of 27 after treatment would not fall within the reliable deterioration threshold even if there is significant clinical deterioration. Other criticisms of the RCI method also include the

use of common standard error values and reliability coefficients instead of using those specific to the population, target condition, measure, setting, and context (Blampied, 2022).

In view of the methodological problems with the RCI to precisely identify deterioration cases, the application of a synthetic control method yielded significantly higher rates, between 16.8% and 16.9% for depression and between 16.4% and 16.9% for anxiety. These deterioration calculations are likely to be more precise as they use sample specific parameters that yielded similar results in both training and validation models with considerable sample sizes.

Furthermore, this study's method of classifying deterioration includes a statistical criterion (see Table 1) that requires deterioration symptoms to be more severe than those expected by a margin that denotes statistical significance (e.g., not likely to be explained by chance or random events). This latter criterion is advantageous as it uses a classification threshold personalised to each individual (based on their synthetic control), rather than imposing an arbitrary threshold (the RCI magnitude) to all cases irrespective of their characteristics and intake level of severity.

Regarding patient characteristics, there were significant differences in some demographic characteristics between participants with type 1 and type 2 deterioration for depression and anxiety. Participants with type 1 deterioration had higher rates of unemployment and lower IMD scores (e.g. living in more socioeconomically deprived neighbourhoods). These demographic characteristics have been found in the literature to be predictors or moderators for treatment non-response and deterioration with studies finding unemployment (Bauer-Staeb et al., 2023; Saunders et al., 2021; Alegria et al., 2018; Saxon et al., 2017; Pelzer et al., 2014; Reneflot & Evensen, 2012), and socioeconomic deprivation

(Stochl et al., 2021; Buckman et al., 2021; Finegan, Firth, & Delgadillo, 2019; Finegan et al., 2018; Alegria et al., 2018; Delgadillo et al., 2016).

The clinical characteristics with significant differences between groups included primary diagnosis, routine outcome measures scores, and end of care reason for both, depression and anxiety. Participants with type 1 deterioration had a higher incidence of mixed diagnosis, comorbidity (higher baseline scores in the opposite measure to the deterioration category, i.e. higher PHQ-9 baseline scores with anxiety deterioration), and functional impairment (higher baseline WSAS scores). Prior studies have found similar results, where comorbidity (Buckman et al., 2021; Kessler et al., 2017; Maj et al., 2020; Saunders et al,. 2016) and poorer functional impairment predicted treatment non-response and clinical deterioration (Amati et al., 2018; Bauer-Staeb et al., 2023; Delgadillo et al., 2018; Stochl et al., 2021; Zimmerman et al., 2016).

The demographic and clinical characteristics found to be significantly different between type 1 and type 2 deterioration in depressive symptoms was mirrored by recent NHS-TT studies, which found that being prescribed and taking medication, age, ethnic minority groups, and having a disability were predictors or moderators of negative treatments outcomes in patients with depression (Bauer-Staeb et al., 2023, Saunders et al., 2021; Teo, 2021), although they had contradictory information on age, with two studies finding younger and the other older people as having increased rates of treatment non-response or deterioration. A study by Delgadillo, Moreea, & Lutz (2016) also found associations between disability, younger age and persistent post-treatment depressive symptoms. On the other hand, there are studies in the literature that have not found medication status to be associated with depressive symptom trajectory (Lin & Farber, 2021; Saunders, 2019; Trombello et al,. 2020).

Significant differences on age, gender and ethnicity were found between type 1 and type 2 deterioration for anxiety in this study. Recent NHS-TT studies have found mixed results on the association between gender and anxiety symptom trajectory, with one not finding significant associations and the other associated females with lower clinical improvement; both studies also found younger people had increased odds of treatment non-response (Stochl et al., 2021; Teo, 2021). Ethnicity has also been associated with clinical outcomes in NHS-TT (Delgadillo, Moreea, & Lutz., 2016; Teo, 2021) and other settings (Cabral & Smith, 2011), while others have found the contrary (Clark et al., 2009; Mercer et al., 2019), including studies in other settings where no association between age, gender, ethnicity and clinical outcomes was found (Bauer-Staeb et al., 2023; Saunders et al., 2019; Trombello et al., 2020). Although, a limitation in this study was that non-White ethnicities lacked representation and could not be subdivided.

Additionally, participants with type 1 deterioration in both samples were more likely to drop out than those with type 2 deterioration. Drop out prevalence was also higher in either deterioration group in both samples than those found in the literature. Recent meta-analyses have found varying dropout rates, including 19.7% in psychotherapy RCTs (Swift & Greenberg, 2012), 26.2% in CBT (Fernandez et al., 2015), 17.5% in RCTs for depression (Cooper & Conklin, 2015), 24.63% in routine clinical practice for depression (Hans & Hiller, 2013), and 16.99% in psychotherapy for GAD (Gersh et al., 2017). Some of these studies identified possible moderators, as dropout rates were higher in studies with a higher number of participants of ethnic minorities (Cooper & Conklin, 2015) and associated with a depression diagnosis (Fernandez et al., 2015), younger age, lower educational level, and therapist experience (Swift & Greenberg, 2012). Other studies have also found increased functional impairment, male sex, lower educational level, personality styles, treatment expectations, and therapist effects as possible predictors (Zimmerman et al, 2016). Some of these possible

moderators or predictors are similar to those of clinical deterioration as seen in this study, although any specific drop out investigations were out of the scope of the study.

Limitations

This study has several limitations. Although the prevalence of deterioration types were quantified, the results do not shed light on the mechanisms that may be implicated in these outcomes. For example, no process measures were available to provide triangulated evidence in support of type 1 deterioration (e.g., alliance, adherence to treatment protocol on part of the therapists, adherence to homework assignments on part of the patients, etc.).

Likewise, type 2 deterioration could be explained by mechanisms such as intersessional adverse life events, concurrent illnesses, or other factors implicated in the progression of chronic health problems (e.g., biomarkers of disease severity), but no such data sources were available. Overall, this is an exploratory study limited to routine quantitative outcome data that requires replication and further research, which could focus on specific demographic and clinical characteristics instead of the broad categories used in the study. Other therapeutic modalities, levels of care, and settings could also be explored and studied to assess generalisability. Future studies applying the synthetic control method could additionally validate the deterioration subtypes by triangulating the classifications with theoretically plausible mechanistic variables such as those listed above. For example, we might expect to see lower average alliance scores in type 1 cases relative to type 2 cases, and we might expect to see higher average inflammation markers in type 2 cases relative to type 1 cases. Such investigations may enable clinical practitioners and researchers to refine and validate their understanding of deterioration subtypes in the future.

Another important limitation is the lack of longer-term follow-up measures of depression and anxiety after the end of treatment. This limits the extent to which we could

validate the type 2 class of deterioration – as we might expect to see a persistent trend of deterioration in this subtype over time, relative to type 1 cases. Studies that have longer-term follow-up measures could be useful to validate the differentiation between subtypes in this way (e.g., expecting that the mean level of severity would be higher in type 2, at a follow-up measurement). Moreover, this analysis is limited to cases accessing brief and low-intensity CBT in the NHS-TT system in England, and hence we cannot make generalisations to other types of treatment or settings.

Clinical Implications

This study suggests that the prevalence of deterioration after LICBT is likely to be higher than previously thought, in the range of 16%, with 10% to 12% likely due to iatrogenic reactions to therapy (type 1), and 4% to 7% due to a natural illness progression (type 2). These results indicate that deterioration might not occur in a small minority of cases, and that it could be more likely attributed to the LICBT intervention itself instead of other biopsychosocial factors.

Additionally, patients who seemingly responded adversely to LICBT (type 1) were more likely to drop out of treatment, an association that should be considered by clinical practitioners.

The fact that one in ten patients has an adverse reaction to LICBT is an important ethical and clinical dilemma, which warrants careful monitoring of treatment response and close clinical supervision to prevent deterioration in such cases, which presumably can be corrected if it is related to the treatment itself. Wider evidence in the field indicates that the use of routine outcome monitoring and feedback-informed treatment can effectively help to prevent deterioration in cases at risk of poor treatment response (de Jong et al., 2021).

Conclusion

The present study applied a methodologically robust method to identify deterioration in LICBT, including its subtypes of type 1 and type 2 cases. The naturalistic and multi-service dataset enabled the examination of deterioration rates in a diverse and representative clinical sample. Machine learning models were used to generated synthetic controls, a form of causal inference methodology applied to observational data that calculates more reliable estimations of treatment effects relative to traditional statistical methods such as the RCI. Additionally, this study conducted sub-group analyses to identify common characteristics between participants without deterioration, and with type 1 and 2 of deterioration. Findings suggests that deterioration after LICBT may be more prevalent and more likely to be due to iatrogenic reactions to therapy than previously reported in the wider literature. Furthermore, participants were more likely to deteriorate if they had a mixed diagnosis, symptomatic comorbidity, functional impairment, were unemployed, and living in socioeconomically deprived neighbourhoods. Other common characteristics such as age, gender, ethnicity, medication use, and disability status were also associated with different deterioration subtypes. Although, future research is needed validate these findings and explore mechanisms underpinning deterioration as this study's results are limited to a specific setting, treatment, and methodology,

References

Alegria, M., NeMoyer, A., Falgas, I., Wang, Y., & Alvarez, K. (2018). Social determinants of mental health: Where we are and where we need to go. *Current Psychiatry Reports, 20*, 95. https://doi.org/10.1007/s11920-018-0969-9

Amati, F., Banks, C., Greenfield, G., & Green, J. (2018). Predictors of outcomes for patients with common mental health disorders receiving psychological therapies in community settings: A systematic review. *Journal of Public Health, 40 (3)*, e375-e387. https://doi.org/10.1093/pubmed/fdx168

Andrews, G., Cuijpers, P., Craske, M., McEvoy, P., & Titov, N. (2010). Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: A meta-analysis. *PLoS One, 5 (10)*, e13196. https://doi.org/10.1371/journal.pone.0013196.

Balder, T., Linden, M., & Rose, M. (2024). Side effects in psychodynamic and cognitive behavior therapy. *Journal of Contemporary Psychotherapy. 54 (3)*, 235-244. https://doi.org/https://doi.org/10.1007/s10879-023-09615-5

Barkham, M., Lutz, W., & Castonguay, L. (2021). *Bergin and Garfield's handbook of psychotherapy and behavior change* (50th ed.). Wiley.

Bauer-Staeb, C., Griffith, E., Faraway, J., & Button, K. (2023). Trajectories of depression and generalised anxiety symptoms over the course of cognitive behaviour therapy in primary care: An observational, retrospective cohort. *Psychological Medicine, 53 (10)*, 4648-4656. https://doi.org/10.1017/S0033291722001556

Bennett-Levy, J., Richards, D., Farrand, P., Christensen, H., Griffiths, K., Kavanagh, D., & Williams, C. (2010). *Oxford Guide to Low Intensity CBT Interventions.* Oxford University Press.

Berk, M., & Parker, G. (2009). The elephant on the couch: Side-effects of psychotherapy. *Austraulian & New Zealand Journal of Psychiatry, 43 (9),* 787-794. https://doi.org/10.1080/00048670903107559

Blampied, N. M. (2022). Reliable change and the reliable change index: Still useful after all these years? *The Cognitive Behaviour Therapist, 15*, e50. https://doi.org/10.1017/S1754470X22000484

Bouttell, J., Craig, P., Lewsey, J., Robinson, M., & Popham, F. (2018). Synthetic control methodology as a tool for evaluating population-level health interventions. *Journal of Epidemiology and Community Health, 72(8)*, 673-678. https://doi.org/10.1136/jech-2017-210106

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D., Gilbody, S., Knowles, S., Cuijpers, P., Andersson, G., Christensen, H., Meyer, B., Huibers, M., Smit, F., van Straten, A., Warmerdam, L., Barkham, M., Bilich, L., Lovell, K., & Liu, E. (2013). Influence of initial severity of depression on effectiveness of low intensity interventions: Meta-analysis of individual patient data. *BMJ, 346*, f540. https://doi.org/10.1136/bmj.f540

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

Buckman, J., Saunders, R., Cohen, Z., Barnett, P., Clarke, K., Ambler, G., & Pilling, S. (2021). The contribution of depressive 'disorder characteristics' to determinations of

prognosis for adults with depression: An individual patient data meta-analysis. *Psychological Medicine, 51 (7)*, 1068-1081. https://doi.org/10.1017/S0033291721001367

Cabral, R., & Smith, T. (2011). Racial/ethnic matching of clients and therapists in mental health services: A meta-analytic review of preferences, perceptions, and outcomes. *Journal of Counselling Psychology, 58 (4)*, 537-554. https://doi.org/10.1037/a0025266

Cahill, J., Barkham, M., & Stiles, W. (2010). Systematic review of practice-based research on psychological therapies in routine clinic settings. *British Journal of Clinical Psychology, 49 (4)*, 421-453. https://doi.org/10.1348/014466509X470789

Captari, L., Hook, J., Hoyt, W., Davis, D., McElroy-Heltzel, S., & Worthington Jr, E. (2018). Integrating clients' religion and spirituality within psychotherapy: A comprehensive meta-analysis. *Journal of Clinical Psychology, 74 (11)*, 1938-1951. https://doi.org/10.1002/jclp.22681

Chen, T., & Gestrin, C. (2016). XGBoost: A scalable tree boosting system. [Conference presentation] *ACM 2016 Conference Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* San Fransisco, CA, United States. https://ceur-ws.org/Vol-2324/Paper19-AQuemy.pdf https://doi.org/10.1145/2939672.2939785

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *Internation Review of Psychiatry, 23 (4)*, 318-327. https://doi.org/10.3109/09540261.2011.606803

Clark, D., Layard, R., Smithies, R., Richards, D., Suckling, R., & Wright, B. (2009). Improving access to psychological therapy: Initial evaluation of two UK

demonstration sites. *Behaviour Research and Therapy, 47 (11)*, 910-920. https://doi.org/10.1016/j.brat.2009.07.010

Cooper, A., & Conklin, L. (2015). Dropout from individual psychotherapy for major depression: A meta-analysis of randomized clinical trials. *Clinical Psychology Review, 40*, 57-65. https://doi.org/10.1016/j.cpr.2015.05.001

Costa, A., Di Buccio, E., Melucci, M., & Nannicini, G. (2018). Efficient parameter estimation for information retrieval using black-box optimization. *IEEE Transactions on Knowledge and Data Engineering, 30 (7)*, 1240-1253. https://doi.org/10.1109/TKDE.2017.2761749

Crawford, M. J., Thana, L., Farquharson, L., Palmer, L., Hancock, E., Bassett, P., Clarke, J., & Parry, G. D. (2018). Patient experience of negative effects of psychological treatment: Results of a national survey. *British Journal of Psychiatry, 208 (3)*, 260-265. https://doi.org/10.1192/bjp.bp.114.162628

Cuijpers, P., Quero, S., Noma, H., Ciharova, M., Miguel, C., Karyotaki, E., Cipriani, A., Cristea, I., & Furukawa, T. A. (2021). Psychotherapies for depression: A network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. *World Psychiatry, 20 (2)*, 283-293. https://doi.org/10.1002/wps.20860

Cuijpers, P., Reijnders, M., Karyotaki, E., de Wit, L., & Ebert, D. (2018). Negative effects of psychotherapies for adult depression: A meta-analysis of deterioration rates. *Journal of Affective Disorders, 239*, 138-145. https://doi.org/10.1016/j.jad.2018.05.050

Curran, J., Parry, G., Hardy, G., Darling, J., Mason, A., & Chambers, E. (2019). How does therapy harm? A model of adverse process using task analysis in the meta-synthesis of service users' experience. *Frontiers in Psychology, 10,* 347. https://doi.org/10.3389/fpsyg.2019.00347

de Jong, K., Conijn, J., Gallagher, R., Reshetnikova, A., Heij, M., & Lutz, M. (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review, 85,* 102002. https://doi.org/10.1016/j.cpr.2021.102002

de Ville, B. (2013). Decision trees. *WIREs Computational Statistics, 5 (6)*, 448-455. https://doi.org/10.1002/wics.1278

Delgadillo, J., & Atzil-Slonim, D. (2023). Artificial intelligence, machine learning and mental health. *Encyclopedia of Mental Health (3rd ed)*, 132-142 https://doi.org/10.1016/B978-0-323-91497-0.00177-6

Delgadillo, J., Asaria, M., Ali, S., & Gilbody, S. (2016). On poverty, politics and psychology: the socioeconomic gradient of mental healthcare utilisation and outcomes. *British Journal of Psychiatry, 209 (5)*, 429-430. https://doi.org/10.1192/bjp.bp.115.171017

Delgadillo, J., de Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., & McMillan, D. (2018). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: A multisite, open-label, cluster randomised controlled trial. *The Lancet, 5 (7)*, 564-572. https://doi.org/10.1016/S2215-0366(18)30162-7

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., Rubel, J., & Barkham, M. (2020). Towards personalized allocation of patients to therapists. *Journal of Consulting and Clinical Psychology, 88*, 799-808. https://doi.org/10.1037/ccp0000507

Duggan, C., Parry, G., McMurran, M., Davidson, K., & Dennis, J. (2014). The recording of adverse events from psychological treatments in clinical trials: Evidence from a review of NIHR-funded trials. *BMC Trials, 15 (1)*, 335. https://doi.org/10.1186/1745-6215-15-335

Ernst, E. (2001). Assessment of therapeutic safety in systematic reviews: Literature review. *BMJ*, *323*, 546. https://doi.org/10.1136/bmj.323.7312.546

Fernandez, E., Salem, D., Swift, J., & Ramtahal, N. (2015). Meta-analysis of dropout from cognitive behavioral therapy: Magnitude, timing, and moderators. *Journal of Consulting and Clinical Psychology, 83 (6)*, 1108-1122. https://doi.org/10.1037/ccp0000044

Finegan, M., Firth, N., & Delgadillo, J. (2020). Adverse impact of neighbourhood socioeconomic deprivation on psychological treatment outcomes: The role of area-level income and crime. *Psychotherapy Research, 30 (4)*, 546-554. https://doi.org/10.1080/10503307.2019.1649500

Finegan, M., Firth, N., Wojnarowski, C., & Delgadillo, J. (2018). Associations between socioeconomic status and psychological therapy outcomes: A systematic review and meta-analysis. *Depression and Anxiety, 35 (6)*, 560-573. https://doi.org/10.1002/da.22765

Gellatly, J., Bower, P., Hennessy, S., Richards, D., Gilbody, S., & Lovell, K. (2007). What makes self-help interventions effective in the management of depressive symptoms? Meta-analysis and meta-regression. *Psychological Medicine, 37 (9)*, 1217-1228. https://doi.org/10.1017/S0033291707000062

Gersh, E., Hallford, D., Rice, S., Kazantzis, N., Gersh, H., Gersh, B., & McCarty, C. (2017). Systematic review and meta-analysis of dropout rates in individual psychotherapy for

generalized anxiety disorder. *Journal of Anxiety Disorders, 52*, 25-33. https://doi.org/10.1016/j.janxdis.2017.10.001

Golino, H., & Gomes, C. (2016). Random forest as an imputation method for education and psychology research: Its impact on item fit and difficulty of the Rasch model. *International Journal of Research & Methods in Education, 39 (4)*, 401-421. https://doi.org/10.1080/1743727X.2016.1168798

Gyani, A., Shafran, R., Layard, R., & Clark, D. (2013). Enhancing recovery rates: Lessons from year one of IAPT. *Behaviour Research and Therapy, 51 (9)*, 597-606. https://doi.org/10.1016/j.brat.2013.06.004

Hans, E., & Hiller, W. (2013). Effectiveness of and dropout from outpatient cognitive behavioral therapy for adult unipolar depression: A meta-analysis of non-randomized effectiveness studies. *Journal of Consulting and Clinical Psychology, 81 (1)*, 75-88. https://doi.org/10.1037/a0031080

Hastie, T., Friedman, J., & Tibshirani, R. (2009). *The Elements of Statistical Learning.* Springer.

Ioannidis, J., Evans, S., Gotzsche, P., O'Neill, R., Altman, D., Schulz, K., & Moher, D. (2004). Better reporting of harms in randomized trials: An extension of the CONSORT statement. *Annals of Internal Medicine, 141 (1)*, 781-788. https://doi.org/10.7326/0003-4819-141-10-200411160-00009

Jacobson, N., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59 (1)*, 12-19. https://doi.org/10.1037//0022-006x.59.1.12

Kellett, S., Webb, K., Wilkinson, N., Bliss, P., Ayers, T., & Hardy, G. (2016). Developing services for patients with depression or anxiety in the context of long-term physical health conditions and medically unexplained symptoms: Evaluation of an IAPT pathfinder site. *Behavioural and Cognitive Psychotherapy, 44 (5)*, 553-567. https://doi.org/10.1017/S1352465816000114

Kessler, R., van Loo, H., Wardenaar, K., Bossarte, R., Brenner, L., Ebert, D., de Jonge, P., Nierenberg, A., Rosellini, A., Sampson, N., Schoevers, R., Wilcox, M., & Zaslavsky, A. (2017). Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiology and Psychiatric Sciences, 26 (1)*, 22-36. https://doi.org/10.1017/S2045796016000020

Klatte, R., Strauss, B., Fluckiger, C., & Rosendahl, J. (2023). Adverse events in psychotherapy randomized controlled trials: A systematic review. *Psychotherapy Research*, *35 (1)*, 1-16. https://doi.org/10.1080/10503307.2023.2286992

Klatte, R., Strauss, B., Fluckiger, C., & Rosendhal, J. (2018). Adverse effects of psychotherapy: Protocol for a systematic review and meta-analysis. *BMC Systematic Reviews, 7*, 135. https://doi.org/10.1186/s13643-018-0802-x

Kroenke, K., Spitzer, R., & Williams, J. (2001). The PHQ-9. *Journal of General Internal Medicine, 16*, 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lambert, M. J. (2004). *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change.* (5th ed.). Wiley.

Lambert, M. J. (2013). Outcome in psychotherapy: The past and important advances. *Psychotherapy, 50 (1)*, 42-51. https://doi.org/10.1037/a0030682

Lazar, E. (2017). *Client Deterioration in individual psychotherapy: A systematic review.* (Publication No. 10605182) [Doctoral dissertation, Indiana University of Pennsylvania]. ProQuest Dissertations & Theses. https://www.proquest.com/openview/ae7c771205476bd30542968850c0637c/1?pq-origsite=gscholar&cbl=18750

Levis, B., Benedetti, A., & Thombs, B. (2019). Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *BMJ, 365*, I1476. https://doi.org/10.1136/bmj.l1476

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2 (1)*. https://doi.org/10.1111/j.1745-6916.2007.00029.x

Lin, T., & Farber, B. (2021). Trajectories of depression in psychotherapy: How client characteristics predict clinical improvement. *Journal of Clinical Psychology, 77 (6)*, 1354-1370. https://doi.org/10.1002/jclp.23119

Linden, M. (2012). How to define, find and classify side effects in psychotherapy: From unwanted events to adverse treatment reactions. *Clinical Psychology & Psychotherapy, 20 (4)*, 236-296. https://doi.org/10.1002/cpp.1765

Linden, M., & Schermuly-Haupt, M. (2014). Definition, assessment and rate of psychotherapy side effects. *World Psychiatry, 13 (3)*, 306-309. https://doi.org/10.1002/wps.20153

Linden, M., Strauss, B., Scholten, S., Nestoriuc, Y., Brakemeier, E.-L., & Wasilewski, J. (2018). Definition und Entscheidungsschritte in der Bestimmung und Erfassung von Nebenwirkungen von Psychotherapie. *Psychotherapy, Psychosomatic Medicine and Psychology*, *68*, 377-382. https://doi.org/10.1055/a-0619-5949

Lorenz, T. K. (2020). Predictors and impact of psychotherapy side effects in young adults. *Counselling and Psychotherapy Research, 21 (1)*, 237-243. https://doi.org/10.1002/capr.12356

Maj, M., Stein, D., Parker, G., Zimmerman, M., Fava, G., De Hert, M., & Wittchen, H. (2020). The clinical characterization of the adult patient with depression aimed at personalization of management. *World Psychiatry, 19 (3)*, 269-293. https://doi.org/10.1055/a-0619-594910.1002/wps.20771

McKay, D., & Jensen-Doss, A. (2021). Harmful treatments in psychotherapy. *Clinical Psychology: Science and Practice, 28 (1)*, 2-4. https://doi.org/10.1037/cps0000023

Mercer, L., Evans, L. J., Turton, R., & Beck, A. (2019). Psychological therapy in secondary mental health care: Access and outcomes by ethnic group. *Journal of Racial and Ethnic Health Disparities, 6*, 419-426. https://doi.org/10.1055/a-0619-594910.1007/s40615-018-00539-8

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *British Journal of Psychiatry, 180*, 461-464. https://doi.org/10.1055/a-0619-594910.1192/bjp.180.5.461

National Collaborating Centre for Mental Health. (2018). *The Improving Access to Psychological Therapies Manual.* NHS England. https://www.england.nhs.uk/publication/the-improving-access-to-psychological-therapies-manual/

National Health Service. (2023). *Psychological therapies: Reports on the use of IAPT services, England, March 2023 Final including a report on the IAPT Employment Advisers pilot and Quarter 4 2022-23 data*. https://digital.nhs.uk/data-and-

information/publications/statistical/psychological-therapies-report-on-the-use-of-iapt-services/march-2023-final-including-a-report-on-the-iapt-employment-advisers-pilot#

Petrosino, A., Turpin-Petrosino, C., & Buehler, J. (2002). 'Scared Straight' and other juvenile awareness programs for preventing juvenile delinquency. *Cochrane Database of Systematic Reviews, 30 (4),* CD002796. https://doi.org/10.1002/14651858.CD002796.pub2

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

Probst, P., Wright, M. N., & Boulesteix, A. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery, 9 (3)*, e1301. https://doi.org/10.1002/widm.1301

Quemy, A. (2019). Data pipeline selection and optimization. *Workshop Proceedings of the EDBT/ICDT 2019 Joint Conference.* Lisbon, Portugal: CEUR-WS. https://ceur-ws.org/Vol-2324/Paper19-AQuemy.pdf

Riley, R., Snell, K., Ensor, J., Burke, D., Harrell Jr, F., Moons, K., & Collins, G. (2018). Minimum sample size for developing a multivariable prediction model: Part I – Continuous outcomes. *Statistics in Medicine, 38 (7)*, 1262-1275. https://doi.org/10.1002/sim.7993

Rozental, A., Castonguay, L., Dimidjian, S., Lambert, M., Shafran, R., Andersson, G., & Carlbring, P. (2018). Negative effects in psychotherapy: Commentary and recommendations for future research and clinical practice. *British Journal of Psychiatry Open, 4 (4)*, 307-312. https://doi.org/10.1192/bjo.2018.42

Rozental, A., Kottorp, A., Forsstrom, D., Mansson, K., Boettcher, J., & Andersson, G. (2019). The negative effects questionnaire: Psychometric properties of an instrument for assessing adverse and unwanted events in psychological treatments. *Behavavioural and Cognitive Psychotherapy, 47 (5)*, 307–312. https://doi.org/10.1017/S1352465819000018

Saunders, R., Buckman, J. E., Cape, J., Fearon, P., Leibowitz, J., & Pilling Stephen. (2019). Trajectories of depression and anxiety symptom change during psychological therapy. *Journal of Affective Disorders, 249*, 327-335. https://doi.org/10.1016/j.jad.2019.02.043

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

Saunders, R., Cape, J., Fearon, P., & Pilling, S. (2016). Predicting treatment outcome in psychological treatment services by identifying latent profiles of patients. *Journal of Affective Disorders, 197*, 107-115. https://doi.org/10.1016/j.jad.2016.03.011

Schermuly-Haupt, M.-L., Linden, M., & Rush, A. J. (2018). Unwanted events and side effects in cognitive behaviour therapy. *Cognitive Therapy Research, 42*, 219-229. https://doi.org/10.1007/s10608-018-9904-y

Schunk, D. (2008). A Markov chain Monte Carlo algorithm for multiple imputation in large surveys. *AStA Advances in Statistical Analysis, 92*, 101-114. https://doi.org/10.1007/s10182-008-0053-6

Shafran, R., Myles-Hooton, P., Bennett, S., & Ost, L. (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

Shorter, E. (2008). History of psychiatry. *Current Opinion in Psychiatry, 21* (6), 593-597. https://doi.org/10.1097/YCO.0b013e32830aba12

Spitzer, R. L., Kroenke, K., Williams, J. B., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166 (10)*, 1092-1097. https://doi.org/10.1001/archinte.166.10.1092.

Stochl, J., Soneson, E., Stuart, F., Fritz, J., Walsh, A. E., Croudace, T., Hodgekins, J., Patel, U., Russo, D., Knight, C., Jones, P., & Perez, J. (2021). Determinants of patient-reported outcome trajectories and symptomatic recovery in Improving Access to Psychological Therapies (IAPT) services. *Psychological Medicine, 52 (14)*, 3231-3240. https://doi.org/10.1017/S0033291720005395

Strauss, B., Gawlytta, R., Schleu, A., & Frenzl, D. (2021). Negative effects of psychotherapy: estimating the prevalence in a random national sample. *British Journal of Psychiatry Open, 7 (6)*, e186. https://doi.org/10.1192/bjo.2021.1025

Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80 (4)*, 547-559. https://doi.org/10.1037/a0028226

Teo, C. (2021). *The impact of ethnicity on the trajectory of depression symptom change during psychological interventions.* [Doctoral dissertation, University College London]. UCL Discovery. https://discovery.ucl.ac.uk/id/eprint/10135909/

Thompson, M., Parker, H., & Cave, J. (2021). Exploring which aspects of a low-intensity CBT intervention were found to contribute to a successful outcome from the service user point of view: A mixed methods study. *Counselling and Psychotherapy Research, 22 (2)*, 279-291. https://doi.org/10.1002/capr.12433

Thorpe, G. (2017). *Identifying and predicting deterioration during psychotherapeutic interventions.* [Doctoral dissertation, University of Sheffield]. White Rose eTheses Online. https://etheses.whiterose.ac.uk/2160/2/Thorpe,_Germaine.pdf

Trombello, J., South, C., Sanchez, A., Kahalnik, F., Kennard, B., & Trivedi, M. (2020). Two trajectories of depressive symptom reduction throughout behavioral activation teletherapy among underserved, ethnically diverse, primary care patients: A VitalSign6 Report. *Behaviour Therapy, 51 (6)*, 958-971. https://doi.org/10.1016/j.beth.2020.01.002

van Gent, E. M., & Zwart, F. M. (1991). Psychoeducation of partners of bipolar-manic patients. *Journal of Affective Disorders, 21*(1), 15-18. https://doi.org/10.1016/0165-0327(91)90013-I

Vaughan, B., Goldstein, M. H., Alikakos, M., Cohen, L. J., & Serby, M. J. (2014). Frequency of reporting of adverse events in randomized controlled trials of psychotherapy vs. psychopharmacotherapy. *Comprehensive Psychiatry, 55 (4)*, 849-855. https://doi.org/10.1016/j.comppsych.2014.01.001

Wakerfield, S., Kellett, S., Simmonds-Buckley, M., Stockton, D., Bradbury, A., & Delgadillo, J. (2021). Improving Access to Psychological Therapies (IAPT) in the United Kingdom: A systematic review and meta-analysis of 10-years of practice-based evidence. *British Journal of Clinical Psychology, 60 (1)*, 1-37. https://doi.org/10.1111/bjc.12259

Weitz, E. S., Hollon, S. D., Twisk, J., van Straten, A., Huibers, M. J., David, D., DeRubeis, R., Dimidjian, S., Dunlop, B., Cristea, I., Faramarzi, M., Hegerl, U., Jarret, R., Kheirkhah, F., Kennedy, S., Mergl, R., Miranda, J., Mohr, D., Rush, A., Segal, Z., Siddique, J., Simons, A., Vittengl, J., & Cuijpers, P. (2015). Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs pharmacotherapy: An individual patient data meta-analysis. *JAMA Psychiatry, 72 (11)*, 1102-1109. https://doi.org/10.1001/jamapsychiatry.2015.1516

Werch, C. E., & Own, D. M. (2002). Iatrogenic effects of alcohol and drug prevention programs. *Journal of Studies on Alcohol, 63*, 581-590. https://doi.org/10.15288/jsa.2002.63.581

Williams, A. J., Botanov, Y., Kilshaw, R. E., Wong, R. E., & Sakaluk, J. K. (2020). Potentially harmful therapies: A meta-scientific review of evidential value. *Clinical Psychology: Science and Practice*. https://doi.org/10.1111/cpsp.12331

Zimmermann, D., Rubel, J., Page, A. C., & Lutz, W. (2016). Therapist effects on and predictors of non-consensual dropout in psychotherapy. *Clinical Psychology and Psychotherapy, 24 (2)*, 312-321. https://doi.org/10.1002/cpp.2022

Appendix A: Ethical Approval Confirmation Letter

Approved: 03/03/2023

Itzia Perez Morales
Registration number: 210154962
Psychology
Programme: Doctorate in Clinical Psychology

Dear Itzia

PROJECT TITLE: Prevalence and Types of Symptomatic Deterioration after Low Intensity CBT in Primary Care services
APPLICATION: Reference Number 051168

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: Routine Outcome Measure Questionnaires

Patient Health Questionnaire and General Anxiety Disorder
(PHQ-9 and GAD-7)

Date_____ Patient Name:_____ Date of Birth: _____

Over the last 2 weeks, how often have you been bothered by any of the following problems?
Please circle your answers.

PHQ-9	**Not at all**	**Several days**	**More than half the days**	**Nearly every day**
1. Little interest or pleasure in doing things.	0	1	2	3
2. Feeling down, depressed, or hopeless.	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much.	0	1	2	3
4. Feeling tired or having little energy.	0	1	2	3
5. Poor appetite or overeating.	0	1	2	3
6. Feeling bad about yourself – or that you are a failure or have let yourself or your family down.	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television.	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed. Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual.	0	1	2	3
9. Thoughts that you would be better off dead, or of hurting yourself in some way.	0	1	2	3
Add the score for each column				

Total Score (add your column scores): _____

If you checked off any problems, how difficult have these made it for you to do your work, take care of things at home, or get along with other people? (Circle one)

Not difficult at all **Somewhat difficult** **Very Difficult** **Extremely Difficult**

Over the last 2 weeks, how often have you been bothered by any of the following problems?
Please circle your answers.

GAD-7	**Not at all sure**	**Several days**	**Over half the days**	**Nearly every day**
1. Feeling nervous, anxious, or on edge.	0	1	2	3
2. Not being able to stop or control worrying.	0	1	2	3
3. Worrying too much about different things.	0	1	2	3
4. Trouble relaxing.	0	1	2	3
5. Being so restless that it's hard to sit still.	0	1	2	3
6. Becoming easily annoyed or irritable.	0	1	2	3
7. Feeling afraid as if something awful might happen.	0	1	2	3
Add the score for each column				

Total Score (add your column scores): _____

If you checked off any problems, how difficult have these made it for you to do your work, take care of things at home, or get along with other people? (Circle one)

Not difficult at all **Somewhat difficult** **Very Difficult** **Extremely Difficult**

UHS Rev 4/2020

Developed by Drs. Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke and colleagues, with an educational grant from Pfizer Inc.
No permission required to reproduce, translate, display or distribute, 1999.

Figure 1: PHQ-9 and GAD-7

WORK AND SOCIAL ADJUSTMENT SCALE

NAME:.....DATE:.....HOSPITAL No.:.....PRE/MID/POST/ MFU

INSTRUCTIONS FOR FILLING THIS FORM

People's problems sometimes affect their ability to do certain day-to-day tasks in their lives. To rate your problems look at each section and determine on the scale provided how much your problem affects your ability to carry out the activity. Once you have decided on a number, circle it. Then proceed to the next stage.

WORK

0 ---------- 1 ---------- 2 ---------- 3 ---------- 4 ---------- 5 ---------- 6 ---------- 7 ---------- 8

not at all slightly definitely markedly very severely I cannot work

==========

HOME MANAGEMENT

Cleaning, tidying, shopping, cooking, looking after home/children, paying bills etc.

0 ---------- 1 ---------- 2 ---------- 3 ---------- 4 ---------- 5 ---------- 6 ---------- 7 ---------- 8

not at all slightly definitely markedly very severely

==========

SOCIAL LEISURE ACTIVITIES

With other people, e.g. parties, pubs, outings, entertaining etc.

0 ---------- 1 ---------- 2 ---------- 3 ---------- 4 ---------- 5 ---------- 6 ---------- 7 ---------- 8

not at all slightly definitely markedly very severely

==========

PRIVATE LEISURE ACTIVITIES

Done alone, e.g. reading, gardening, sewing, hobbies, walking etc.

0 ---------- 1 ---------- 2 ---------- 3 ---------- 4 ---------- 5 ---------- 6 ---------- 7 ---------- 8

not at all slightly definitely markedly very severely

==========

FAMILY AND RELATIONSHIPS

Form and maintain close relationships with others including the people that I live with.

0 ---------- 1 ---------- 2 ---------- 3 ---------- 4 ---------- 5 ---------- 6 ---------- 7 ---------- 8

not at all slightly definitely markedly very severely

Figure 2: WSAS