








ARTICLE

Comparison of outcomes across low-intensity psychological interventions for depression and anxiety within a stepped-care setting: A naturalistic cohort study using propensity score modelling

Jorge Palacios^{1,2}  | Adedeji Adegoke¹  | Rebecca Wogan¹  |
 Daniel Duffy^{1,2}  | Caroline Earley^{1,2}  | Nora Eilert^{1,2} |
 Angel Enrique^{1,2}  | Sarah Sollesse³ | Judith Chapman³ |
 Derek Richards^{1,2} 

¹SilverCloud Research, SilverCloud Health, Dublin, Ireland

²E-mental Health Research Group, School of Psychology, Trinity College Dublin, Dublin, Ireland

³Berkshire Healthcare NHS Foundation Trust, Bracknell, UK

Correspondence

Jorge Palacios, School of Psychology, Aras an Phiarsaigh, Trinity College Dublin, Dublin, Ireland.
 Email: jorge.palacios@tcd.ie

Funding information

SilverCloud Health; Berkshire Healthcare Foundation Trust

Abstract

Low-intensity interventions for common mental disorders (CMD) address issues such as clinician shortages and barriers to accessing care. However, there is a lack of research into their comparative effectiveness in routine care. We aimed to compare treatment effects of three such interventions, utilizing four years' worth of routine clinical data. Users completing a course of guided self-help bibliotherapy (GSH), internet-delivered cognitive behavioural therapy (iCBT) or psychoeducational group therapy (PGT) from a stepped-care service within the NHS in England were included. Propensity score models (stratification and weighting) were used to control for allocation bias and determine average treatment effect (ATE) between the interventions. 21,215 users comprised the study sample (GSH = 12,896, iCBT = 6862, PGT = 1457). Adherence-to-treatment rates were higher in iCBT. All interventions showed significant improvements in depression (PHQ-9), anxiety (GAD-7) and functioning (WSAS) scores, with largest effect sizes for iCBT. Both propensity score models showed a significant ATE in favour of iCBT versus GSH and PGT, and in favour of GSH versus PGT. Discernible differences in effectiveness were seen for iCBT in comparison with GSH and PGT. Given variance in delivery mode and human resources between different low-intensity interventions, building on these findings

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *British Journal of Psychology* published by John Wiley & Sons Ltd on behalf of The British Psychological Society.

would be valuable for future service provision and policy decision making.

KEYWORDS

anxiety, depression, guided self-help, internet-delivered cognitive behavioural therapy, low-intensity interventions

BACKGROUND

Common mental disorders (CMD), such as depression and anxiety, account for a sizeable portion of the global burden of disease (GBD). The WHO estimates place depressive disorder as the single largest contributor to years lived with disability, whilst anxiety disorders are ranked 6th (World Health Statistics, 2017). Alongside significant disability and reduced quality of life, these CMDs incur substantial economic costs, not only directly but also indirectly through unemployment, sickness benefit and loss of productivity (König et al., 2019; Konnopka & König, 2020). In the United Kingdom, there has been a general trend over time towards increasing prevalence of moderate to severe symptoms of CMD (McManus et al., 2016). With the ongoing COVID-19 pandemic, prevalence of CMD has increased even further due to direct virus-related health concerns as well as the impact of measures implemented to slow the virus spread (Brooks et al., 2020; Gallagher et al., 2020). Furthermore, the number of adults experiencing some form of depression during the pandemic (June 2020) was approximately one in five, double the amount seen before the pandemic (Vizard et al., 2020), with reports acknowledging an increase in mental distress greater than would have been predicted by the existing upward trends (Pierce et al., 2020).

Overcoming barriers to access to care

One major limitation for people with depression and anxiety symptoms is gaining access to evidence-based treatments (Alonso et al., 2018; Thornicroft et al., 2017). There are several important reasons for this limited access, but primary amongst them are a scarcity of mental health services, a low perceived need for treatment and stigma associated with these conditions (Thornicroft et al., 2017). In fact, the 2014 Adult Psychiatry Morbidity Survey conducted nationally across England reported that approximately 59% of people with depression and only 48% of those with anxiety received an evidence-based treatment (McManus et al., 2016). Overcoming significant barriers is difficult, but innovation in how care can be delivered has been proposed. One example of an innovative, stepped-care approach to treating CMD is the Improving Access to Psychological Therapy (IAPT) programme in England. Originally proposed by Lord Layard and Professor Clark (Layard, 2006; Layard & Clark, 2015), the rationale for developing IAPT was centred around addressing issues such as access to evidence-based mental healthcare interventions, clinician shortages, facilitating patient preference and ameliorating the economic burden of mental illness. IAPT manages Steps 2 and 3 of a national stepped-care programme within the National Health Service (NHS) where Step 1 involves screening and monitoring ('watchful waiting') of minimal symptoms at the primary care level. Step 2 delivers low-intensity interventions via trained psychological well-being practitioners (PWP), intended for service users with mild to moderate symptomatology, whereas for more severe symptoms requiring high-intensity therapies, users are transferred to Step 3 and seen by more experienced licenced therapists for individual face-to-face treatment.

Low-intensity interventions within stepped care

The empirical literature for the low-intensity interventions within this programme has been evaluated by the National Institute for Health and Care Excellence (NICE) and guided self-help bibliotherapy

(GSH), computerized or internet-delivered cognitive behaviour therapy (iCBT), as well as group-based peer support are recommended in their clinical guidelines for the treatment of CMD (NICE, 2009, 2011a, 2011b). Interventions follow CBT best practice, including components such as mood monitoring, behavioural activation, cognitive restructuring, relaxation training and challenging core beliefs. These low-intensity treatments require less ‘therapist time’ and can therefore increase access to evidence-based treatment manuals and protocols. A study analysing the service outcomes in IAPT showed that the overall effects produced by the evidence-based interventions offered within this model were in line with expectation from clinical trials (Clark, 2018). The results are supportive of the IAPT stepped-care model for addressing the ever-growing prevalence of CMD. Additionally, routine data collection is a key feature of IAPT services, supporting the expected outcomes such as recovery rates and symptom improvement and likely return on investments (Richards et al., 2020).

Recently, following the publication of the ‘Five Year Forward View for Mental Health’ additional funding was secured for the continuation of IAPT services but with larger expectations on services to reach greater capacity [IAPT access targets increased from 15% to 25%] whilst maintaining quality and outcomes (Independent Mental Health Taskforce, 2016). Given the differences in cost and human resources to deliver different low-intensity interventions in IAPT (NICE, 2020), the current context raises an important question as to the comparative effectiveness of the low-intensity treatments being delivered. This study therefore seeks to analyse and contrast clinical outcomes across three main low-intensity interventions that are typically available at Step 2 of the IAPT programme, GSH, iCBT and psychoeducational group therapy (PGT), which all contain similar clinical content but differ in their modes of delivery, in order to determine their comparative effectiveness and aid in future policy and clinical decision making.

METHODS

Study design

A naturalistic, observational cohort study design was conducted, using propensity score modelling techniques. The study cohort was drawn from service users of a well-established IAPT service provider who completed a course of treatment at Step 2 within a four-year period (1 April 2016 to 31 March 2020). The authors assert that all procedures contributing to this work follow the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. Approval for users to not provide informed consent was given, as consent had been given for treatment and for therapists to gather and to examine clinical outcome measures as part of normal service evaluation procedures. Service users who refused consent to having their data passed on for these purposes were not included in this study. Furthermore, steps were taken to fully anonymize all retained data, including deletion of patient identifiers, location information and date of birth. As an added precaution, an assessment of k-anonymity was carried out following de-identification to verify the risk of subject re-identifiability was sufficiently low. All procedures involving human patients, the collection and use of data were approved by the National Health Service (NHS) Health Research Authority (HRA) Research Ethics Committee (REC Reference: 18/LO/0385).

Setting and participants

Talking Therapies is an NHS IAPT provider within Berkshire Healthcare Foundation Trust, serving a population of 900,000 across six demographically and economically diverse localities. Individuals who contact the Berkshire NHS Foundation Trust Talking Therapies service are given an initial assessment by phone or in-person, and complete the minimum data set (MDS), which contains validated self-report measures such as the patient health questionnaire (PHQ-9), generalized anxiety disorder

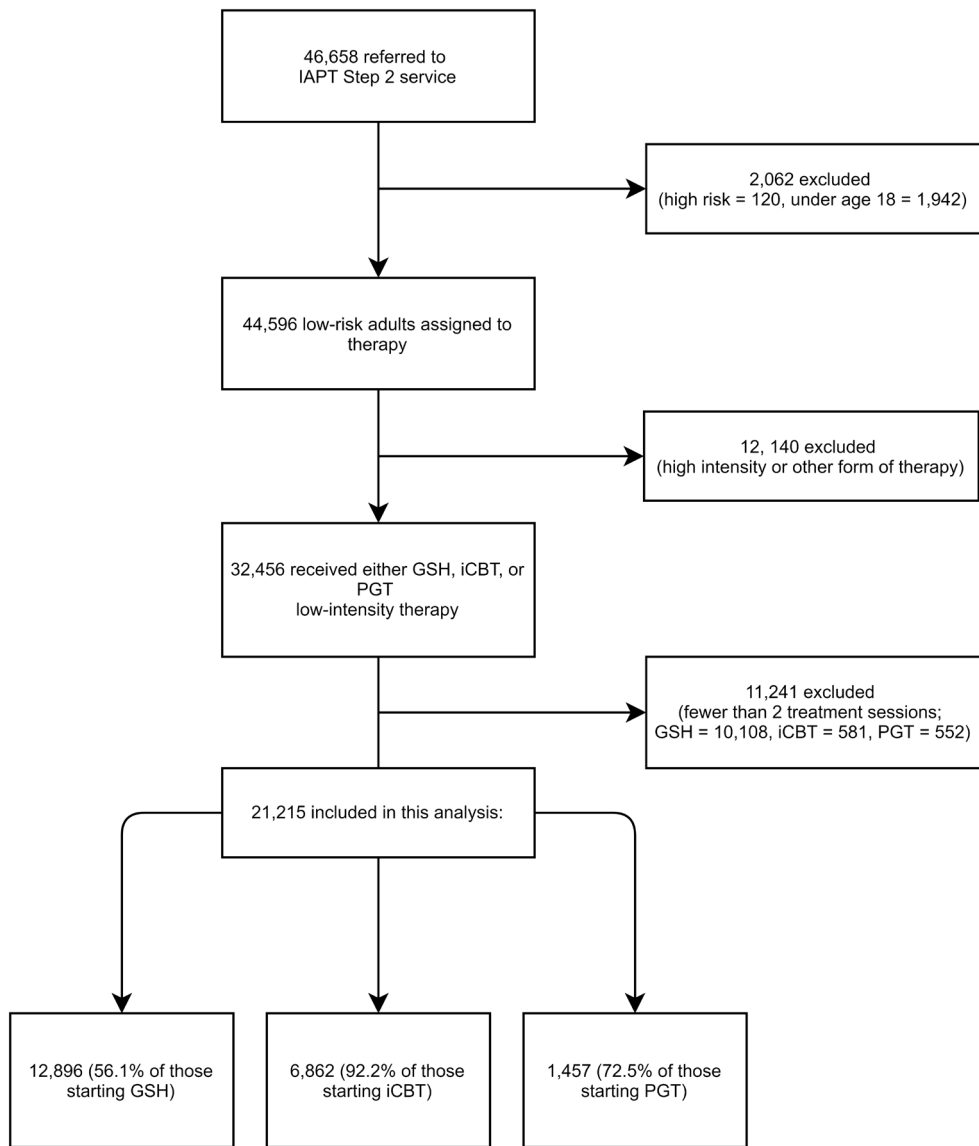


FIGURE 1 Flow and selection of users for inclusion. GSH, guided self-help; IAPT, improving access to psychological treatment; iCBT, internet-delivered cognitive behavioural therapy; PGT, psychoeducational group therapy.

(GAD-7), phobia scale, and Work and Social Adjustment Scale (WSAS), as part of a comprehensive screening procedure. The assessment determines the level of symptomatology and the optimal allocation within the stepped-care model. When discussing treatment options with service users, the PWP provide information about the characteristics of the interventions (i.e. nature, content and duration). The PWP and service user then arrive at a collaborative decision regarding treatment, whilst considering the scores from the MDS, the clinical assessment and the service user's own preference.

Registered IAPT service users who received an initial assessment appointment at Berkshire Talking Therapies between 1 April 2016 and 31 March 2020 and subsequently started and completed a course of treatment at IAPT Step 2 in either Guided Self-Help (GSH), SilverCloud iCBT or Psychoeducational Group Therapy (PGT) were included in the analysis (see Figure 1). Following IAPT guidelines, a completed course of treatment is defined as attendance at two or more treatment appointments (or

receiving two or more online reviews). Those who did not complete a course of treatment (having attended less than 2 appointments) and service users younger than 18 years at the initial assessment appointment were excluded from the analysis.

Outcome measures

The well-established and validated measures which form part of the MDS are routinely collected throughout IAPT services, and include the following:

Patient health questionnaire-9 item (PHQ-9)

The PHQ-9 is a self-report measure of depressive symptoms experienced over the past two weeks, widely used in research and a regular screening measure utilized in primary care and hospital settings (Kroenke et al., 2001). The nine items reflect the diagnostic criteria for depression outlined by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV).

Generalized anxiety disorder-7 item (GAD-7)

GAD-7 comprises seven items measuring symptoms and severity of anxiety based on the DSM-IV diagnostic criteria for generalized anxiety disorder (Spitzer et al., 2006) and is increasingly used in large-scale studies and service provision as a generic measure of change in anxiety symptomatology, using a cut-off score of 8 (Clark, 2011).

Work and Social Adjustment Scale (WSAS)

This is a reliable, valid measure of impaired functioning and its sensitivity to treatment change has been demonstrated (Zahra et al., 2014). Five questions concern how the disorder impairs the service user's ability to function day to day across five dimensions: work, social life, home life, private life and close relationships.

Interventions

The three low-intensity treatment options included in this analysis have all been developed following NICE guidelines (NICE, 2011a, 2011b) and are discussed further below:

Guided Self-Help (GSH)

Guided Self-Help begins with one initial face-to-face treatment planning session with a PWP lasting around 45 min. The treatment plan is based on CBT strategies given in the form of written self-help materials. These materials include information about the specific condition, CBT techniques such as behavioural activation and cognitive restructuring (Baguley et al., 2010), along with related exercises that service users can complete. Support is provided through 4–6 telephone calls usually scheduled every two weeks, each lasting between 20 and 25 min.

Internet-delivered Cognitive-Behavioural therapy (iCBT)

The SilverCloud 'Space from Depression' and 'Space from Anxiety' programmes are 7-module online CBT-based interventions targeting anxiety and depression symptoms. Programme content is delivered on a web 2.0 platform which includes several forms of rich media content (videos, animations and audio) to facilitate the delivery of the intervention. Treatment content consists of cognitive and

behavioural strategies common to CBT protocols, behavioural activation, mood monitoring, cognitive restructuring and relapse prevention (Richards et al., 2020). Support is provided by trained PWP's within the Berkshire NHS Trust. The assigned supporter provides motivation and encouragement, using their clinical skills to provide weekly online asynchronous feedback (usually taking between 15 and 20 min), which the service user can reply to if they wish. The recommended duration of supported treatment is 6–8 weeks, after which the user can still access the content of the programmes for up to twelve months.

Psychoeducational group therapy (PGT)

This well-being course is a psychoeducation intervention taught in groups and typically facilitated by two PWP's, delivering CBT-based materials and content to cope with depression and anxiety symptoms. Clients are encouraged to share their experiences amongst peers in the group, and how these are relevant to the material being discussed, to increase awareness of individual issues in a collective manner. Clients are also asked to complete small homework tasks to support their learning and recovery, these take 15–20 min daily. Typically, up to 15 people can attend, and it has a duration of four weekly sessions of around 90 min each. This PGT serves two main purposes: It helps meet service needs without increasing waiting times and offers CBT tools to users who want to share experiences, learn from others and normalize their difficulties within a group setting.

Data analysis

Adherence rate for each intervention was calculated based on the minimal number of recommended sessions for each intervention (6 for iCBT and 4 for both GSH and PGT). Chi-square tests for independence were used to test for significance between the treatment groups and the adherence rate.

To test the overall efficacy of the three low-intensity interventions in reducing depression, anxiety, and functional impairment scores, a repeated measures Wilcoxon signed-rank test was conducted for significance. Within-group Cohen's *d* and between-group effect sizes were calculated utilizing pooled standard deviations and one way between-group ANOVA utilizing the post-treatment outcome measures. The test-retest reliability was analysed using intraclass correlation coefficient (ICC) with two-way mixed average measure (Trevethan, 2017) and percent agreement, classified as follows: excellent (>0.80), substantial (>0.60 to ≤ 0.80), moderate (>0.40 to 0.60) and poor (≤ 0.40 ; Landis & Koch, 1977). Categories of caseness, reliable improvement and recovery were defined according to IAPT reporting criteria (National Collaborating Centre for Mental Health, 2019). Service users scoring above the clinical threshold at referral on measures of depression (PHQ-9 ≥ 10), anxiety (GAD-7 ≥ 8) or both were at 'caseness'. A service user at caseness prior to treatment and below the clinical threshold on both the PHQ-9 and GAD-7 at the end of treatment was deemed to have recovered. Standard IAPT Reliable change indices (RCIs) were used as cut-offs to measure reliable change on the PHQ-9 (RCI = 6) and GAD-7 (RCI = 4). Reliable improvement was defined as a decrease in either the PHQ-9 or the GAD-7 which was greater than the RCI and no increase in either score larger than the RCI. A service user moving from caseness to non-caseness (recovery) and showing reliable improvement on either the PHQ-9 or the GAD-7 post-treatment was categorized as reliable recovery.

To compare the effectiveness of the three treatments, we used propensity scores analysis, often utilized in non-experimental research to counter the effect of covariates on different intervention groups. The use of propensity scores allows for the control of imbalances on observed variables in non-randomized or observational studies examining the causal effects of treatments or interventions. In this case, treatment allocation may be influenced by users' baseline demographics, clinical symptoms and other social factors, which themselves may affect the outcome, and thus we aimed to mitigate this by balancing all covariates across the three interventions. Once the propensity score has been estimated in a given data set, a data 'pre-processing' procedure is performed to create comparability between study groups which typically involves matching, stratification and weighting. There is no consensus on

which method may outperform the other, as the treatment effect may show some variation depending on the method used (Austin, 2011). Propensity score stratification and weighting were used in this study before conducting treatment effect estimate analysis for both in order to obtain robust findings and add validity to the conclusions.

For stratification, propensity scores were estimated with logistic regression, using the R package 'glmnet'. The covariates included for the propensity score calculation were socio-demographics (gender, age-group, ethnicity, nationality, employment status and locality), service characteristics (referral source and PWP experience) and baseline clinical characteristics (risk-rating, presence of a long-term condition, use of psychotropic medication, PHQ-9 score, GAD-7 score and WSAS score). Missing covariate data, ranging from 0.06% in gender to 3.1% in psychotropic medication, was accounted for via multiple imputation and sensitivity analyses were conducted using chi-square test of independence to compare the missing variables and the imputed variables to understand the significant difference between both samples (Table S1). Propensity scores were then divided into strata, categorizing individuals into homogeneous groups and thus reducing bias. Individuals with similar propensity scores (and thus similar observed baseline characteristics) were categorized into the same stratum. Kolmogorov–Smirnov's test of equivalence (for continuous variables) and Fisher's exact test (for categorical variables) were used to assess the balance of the baseline characteristics within each stratum (Tables S3, S5, and S7). Three pairwise comparisons were conducted to determine the average treatment effects (ATE) across the three treatment groups. For each pairwise comparison and for each clinical measure (PHQ-9, GAD-7 and WSAS), an overall ATE was estimated through the following steps. First, within each stratum the pre- to post-treatment score difference (treatment effect) was calculated for each respective treatment. Then, the difference between the mean treatment effects was calculated for each stratum. Lastly, these differences were pooled to estimate the overall ATE, that is the difference in mean score reduction between compared treatments.

For propensity score weighting, the same covariates were used to estimate the propensity score using the 'twang' package in R, a toolkit for non-equivalent groups which performs propensity score weighting for multiple treatment groups at once. It estimates propensity scores and weighting of treatment cases to estimate the population ATE using a tree-based generalized boosted regression model. The propensity score was estimated using the 'mnps' function in the 'twang' package, which is centred on boosted logistic regression, estimating the probability of an individual to fall into one of the treatment groups. The balance across the baseline characteristics on the multiple treatment groups was calculated according to the absolute standardized mean difference which, according to Harder et al. (2010), must be less than 0.25 to achieve balance across the covariates. After balance was achieved, propensity scores were converted to weights using the 'survey' package in R. The standard error for the propensity score weight estimate was obtained using a resampling method (bootstrapping), using the function 'as.svydesign' to add 1000 replication weights to the weight design object. Propensity score weighting may result in extreme weights which inflate the standard error of the treatment effect estimates. We therefore used weight truncation at the 1st and 99th percentile, as proposed by Cole and Hernán (2008), to decrease bias. Finally, as with the stratification method, three pairwise comparisons were conducted to compare the three treatment groups, with the difference between the weighted mean outcomes being the estimate of the ATE.

A subgroup, follow-up analysis was also undertaken to test whether results were similar in users who had scored at caseness for depression only, for anxiety only or for both depression and anxiety, this was done with the intention to further understand possible differences in these treatments for one particular condition over another and aid in clinical decision making.

The R code that forms the analytical basis of this study is available via the Open Science Framework: DOI [10.17605/OSF.IO/F237T](https://doi.org/10.17605/OSF.IO/F237T)

RESULTS

In total 21,215 service users completed treatment and met the criteria for inclusion, thereby comprising the overall study sample (GSH $n = 12,896$, iCBT $n = 6862$, PGT $n = 1457$; Figure 1). Demographics of

TABLE 1 Baseline demographics of the sample by intervention type

	N (%)				Test statistic
	Total	GSH	iCBT	PGT	
Age					
18–29	7339 (34.6%)	4391 (34%)	2599 (37.9%)	349 (24%)	
30–49	9282 (43.8%)	5406 (41.9%)	3225 (47%)	651 (44.7%)	
50–64	3315 (15.6%)	2109 (16.4%)	887 (12.9%)	319 (21.9%)	
65+	1279 (6%)	990 (7.7%)	151 (2.2%)	138 (9.5%)	$\chi^2(6) = 423.43^*$
Gender					
Female	14,019 (66.1%)	8544 (66.3%)	4588 (66.9%)	887 (60.9%)	
Male	7196 (33.9%)	4352 (33.7%)	2274 (33.1%)	570 (39.1%)	$\chi^2(2) = 19.63^*$
Referral					
GP	1776 (8.4%)	1178 (9.1%)	459 (6.7%)	139 (9.5%)	
Self	18,107 (85.3%)	10,791 (83.7%)	6079 (88.6%)	1237 (84.9%)	
Other	1332 (6.3%)	927 (7.2%)	324 (4.7%)	81 (5.6%)	$\chi^2(4) = 91.94^*$
Employment status					
In employment	15,256 (71.9%)	8865 (68.7%)	5344 (77.9%)	1047 (71.9%)	
Unemployed & actively seeking	1910 (9%)	1296 (10%)	466 (6.8%)	148 (10.2%)	
Receiving benefits	361 (1.7%)	280 (2.2%)	58 (0.8%)	23 (1.6%)	
Retired/homemaker	3688 (17.4%)	2455 (19%)	994 (14.5%)	239 (16.4%)	$\chi^2(6) = 207.78^*$
Psychoactive medication					
Not prescribed	11,473 (54.1%)	6726 (52.2%)	3985 (58.1%)	762 (52.3%)	
Prescribed & taking	8903 (42%)	5619 (43.6%)	2648 (38.6%)	636 (43.7%)	
Prescribed & not taking	839 (4%)	551 (4.3%)	229 (3.3%)	59 (4%)	$\chi^2(4) = 67.42^*$
Long-term condition					
No	17,094 (80.6%)	10,191 (79%)	5789 (84.4%)	1114 (76.5%)	
Yes	4121 (19.4%)	2705 (21%)	1073 (15.6%)	343 (23.5%)	$\chi^2(2) = 98.5^*$

Note: * $p < .0001$.

Abbreviations: GP, general medical practitioner; GSH, guided self-help; iCBT, internet-delivered cognitive behavioural therapy; PGT, psychoeducational group therapy.

the study sample in the overall population and each of the treatment groups are shown in Table 1. Adherence rates between the interventions differed significantly: 8109 (62.9%) GSH users completed 4+ GSH sessions with a mean of 4.9 sessions, 5554 (80.9%) iCBT users completed 6+ iCBT sessions with an mean of 5.6 sessions and 1061 (72.8%) PGT users completed 4+ PGT sessions with a mean of 4.7 sessions. Chi-square tests for independence indicated significant difference between the treatment groups and adherence status, ($\chi^2_{(2)} = 696.4, p < .001, \phi = .18$).

Wilcoxon signed-ranked tests revealed significant pre-post reductions in depression, anxiety and functioning impairment for the overall sample, and for each individual treatment ($p < .0001$; Table S2). Test and retest administration for depression, anxiety and functioning impairment revealed ICCs of 0.50, 0.42 and 0.57, respectively, showing moderate reliability and credibility of the scales. Within-group effects showed large effect sizes for depression and anxiety symptom reduction, and medium effects for

TABLE 2 Descriptive statistics for pre-post and effect sizes for each of the interventions

Outcome	Mean		Within-groups Cohen's <i>d</i> [95% CI]	Between-groups eta-squared [95% CI]
	Baseline (<i>SD</i>)	Post-treatment (<i>SD</i>)		
Total				
PHQ-9	13.8 (5.9)	8.0 (6.2)	0.95 [0.93, 0.97]	.023 [.020, .028]
GAD-7	12.9 (4.9)	7.4 (5.6)	0.98 [0.97, 1.00]	.021 [.017, .025]
WSAS	16.9 (9.1)	11.3 (9.2)	0.60 [0.59, 0.62]	.011 [.008, .014]
iCBT				
PHQ-9	12.8 (5.6)	6.6 (5.4)	1.06 [1.03, 1.09]	
GAD-7	12.4 (4.9)	6.2 (5.0)	1.12 [1.09, 1.15]	
WSAS	15.6 (8.8)	9.9 (8.5)	0.61 [0.58, 0.64]	
GSH				
PHQ-9	14.4 (5.9)	8.7 (6.5)	0.92 [0.89, 0.94]	
GAD-7	13.4 (4.9)	7.9 (5.8)	0.95 [0.92, 0.97]	
WSAS	17.7 (9.1)	11.9 (9.5)	0.61 [0.59, 0.63]	
PGT				
PHQ-9	13.3 (6.2)	8.5 (6.1)	0.80 [0.75, 0.86]	
GAD-7	11.4 (5.3)	7.6 (5.5)	0.75 [0.69, 0.81]	
WSAS	16.8 (9.2)	12.3 (8.9)	0.52 [0.46, 0.57]	

Note. Level of significance $p < .0001$ in all interventions.

Abbreviations: GAD-7, generalized anxiety disorder-7 item questionnaire; GSH, guided self-help; iCBT, internet-delivered cognitive behavioural therapy; PGT, psychoeducational group therapy; PHQ-9, patient health questionnaire-9 item; WSAS, work and social adjustment scale.

functioning score improvement across all three interventions. Between-group effects showed small effect sizes across all post-treatment outcome measures (Table 2).

Overall rates of reliable improvement, recovery and reliable recovery per treatment were calculated for each intervention. The reliable improvement rate was higher in iCBT (67%) compared to GSH (59%) and PGT (49%; $\chi^2_{(2)} = 209.2, p < .0001$). When compared to recovery rates in GSH (50%) and PGT (41%), a higher recovery rate of 65% was observed in service users who completed iCBT ($\chi^2_{(2)} = 411.3, p < .0001$). Similarly, reliable recovery rates were higher in users who completed iCBT treatment (50%) compared to those who completed GSH (41%) and PGT (30%; $\chi^2_{(2)} = 359.3, p < .0001$). Reliable improvement for the entire sample was 61%, recovery was 52%, and total rate of reliable recovery was 46%.

For propensity score stratification, estimated propensity scores were stratified into $K = 10$ strata based on the decile cut-off scores. Outcome analysis was conducted on the clinical measures using the balanced strata which generated each stratum-specific mean (Tables S4, S6, and S8), the stratified mean of all strata and the overall estimate of the ATE (Table 3). With regard to depression as measured by the PHQ-9, iCBT had a statistically significant ATE of 1.26 (SE 0.09; 95% CI 1.08–1.44) above GSH and a statistically significant ATE of 1.71 above PGT (SE 0.22; 95% CI 1.27–2.15). GSH had a statistically significant ATE of 0.46 above PGT (SE 0.22; 95% CI 0.02–0.91). iCBT also had a statistically significant higher ATE for anxiety symptoms, as measured by the GAD-7. Compared to GSH, the ATE was 1.17 (SE 0.09; 95% CI 1.00–1.34), whilst compared with PGT, the ATE was 1.90 (SE 0.21; 95% CI 1.50–2.31). GSH had a statistically significant higher ATE of 0.74 versus PGT (SE 0.19; 95% CI 0.36–1.11). Finally, regarding WSAS scores, a similar pattern was found, all with statistically significant results. The ATE for iCBT was 0.86 versus GSH (SE 0.14; 95% CI 0.58–1.14) and 1.96 versus PGT (SE 0.31; 95% CI 1.35–2.57), whilst for GSH versus PGT, ATE was 1.13 (SE 0.33; 95% CI 0.49–1.77).

TABLE 3 Average treatment effects for the propensity score stratification and weighted method

	Stratified mean		ATE (SE)	95% CI
iCBT vs. GSH	iCBT	GSH		
PHQ-9	6.76	5.50	1.26* (0.09)	[1.08–1.44]
GAD-7	6.48	5.31	1.17* (0.09)	[1.00–1.34]
WSAS	6.33	5.47	0.86* (0.14)	[0.58–1.14]
iCBT vs. PGT	iCBT	PGT		
PHQ-9	6.31	4.60	1.71* (0.22)	[1.27–2.15]
GAD-7	6.08	4.18	1.90* (0.21)	[1.50–2.31]
WSAS	5.79	3.83	1.96* (0.31)	[1.35–2.57]
GSH vs. PGT	GSH	PGT		
PHQ-9	5.71	5.25	0.46* (0.22)	[0.02–0.91]
GAD-7	5.35	4.61	0.74* (0.19)	[0.36–1.11]
WSAS	5.76	4.63	1.13* (0.33)	[0.49–1.77]
	Weighted mean			
iCBT vs. GSH	iCBT	GSH		
PHQ-9	26.78	25.50	1.28* (0.09)	[1.08, 1.47]
GAD-7	23.47	22.27	1.20* (0.08)	[1.02, 1.38]
WSAS	46.40	45.46	0.94* (0.29)	[0.66, 1.23]
iCBT vs. PGT	iCBT	PGT		
PHQ-9	26.78	24.92	1.86* (0.19)	[1.27, 2.15]
GAD-7	23.47	21.33	2.14* (0.17)	[1.80, 2.47]
WSAS	46.40	44.45	1.94* (0.29)	[1.37, 2.51]
GSH vs. PGT	GSH	PGT		
PHQ-9	25.50	24.92	0.58* (0.18)	[0.22, 0.94]
GAD-7	22.27	21.33	0.94* (0.16)	[0.62, 1.25]
WSAS	45.46	44.45	1.01* (0.28)	[0.45, 1.54]

Note: * $p < .0001$.

Abbreviations: ATE, average treatment effect; CI, confidence interval; GSH, guided self-help; iCBT, internet-delivered cognitive behavioural therapy; PGT, psychoeducational group therapy; SE, standard error.

Propensity score weighting involved a generalized boosted regression model to create balance for multiple treatment groups at once. All covariates were confirmed as achieving balance using the absolute standardized mean difference (Table S9). The issue of extreme weight was encountered and was solved using truncation of weight at the 1st and 99th percentile. After weighting, overall rates of reliable improvement, recovery and reliable recovery per treatment were calculated for each intervention. The reliable improvement rate was higher in iCBT (67%) compared with GSH (59%) and PGT (49%); $\chi^2_{(2)} = 209.2, p < .0001$). When compared to recovery rates in GSH (46%) and PGT (42%), a higher recovery rate of 59% was observed in service users who completed iCBT ($\chi^2_{(2)} = 245.2, p < .0001$). Similarly, reliable recovery rates were higher in users who completed iCBT treatment (55%) compared with those who completed GSH (44%) and PGT (38%); $\chi^2_{(2)} = 359.3, p < .0001$). Reliable improvement for the entire sample was 60%, recovery was 50%, and total rate of reliable recovery was 47%. Pairwise approach was

then used to compare the results. Findings were similar to comparisons seen through the stratification method (Table 3), with statistically significant results seen at all comparisons. For depression, iCBT had a higher ATE versus GSH (1.28; CI 95% 1.08–1.47) and PGT (1.86; 1.48–2.24), whilst the GSH ATE was higher than the PGT ATE (0.58; 0.22–0.94). Likewise for anxiety, the ATE in iCBT was higher (ATE 1.20; 1.02–1.38) and PGT (ATE 2.14; 1.80–2.47) whilst the ATE was higher in GSH versus PGT (ATE 0.94; 0.62–1.25). Meanwhile, iCBT also had a higher ATE in terms of functioning scores versus GSH (0.94; 0.66–1.23) and PGT (1.94; 1.37–2.51), whilst GSH had a higher ATE versus PGT (1.01; 0.45–1.54).

A final, subgroup analysis was undertaken to gauge whether results were similar in those users who scored above the threshold for caseness only on depression, only on anxiety or on both measures. In total 1431 service users were at caseness on measure of depression only, 3305 on anxiety only and 14,543 on both (comorbid). With regard to depression as measured by the PHQ-9, iCBT had a statistically significant ATE of 0.79 (*SE* 0.30; 95% CI 0.21–1.37) above GSH and a statistically significant ATE of 1.44 above PGT (*SE* 0.47; 95% CI 0.52–2.36), whilst GSH had a statistically significant ATE of 0.65 above PGT (*SE* 0.45; 95% CI -0.23–1.54). iCBT also had a statistically significantly higher ATE for anxiety symptoms. Compared with GSH, the ATE was 0.69 (*SE* 0.17; 95% CI 0.35–1.03), whilst compared to PGT, the ATE was 1.65 (*SE* 0.43; 95% CI 0.82–2.49). GSH had a statistically significant ATE of 0.96 versus PGT (*SE* 0.42; 95% CI 0.14–1.79). Finally, regarding the comorbid group, a similar pattern was found. Depression ATE for iCBT was 1.64 versus GSH (*SE* 0.12; 95% CI 1.40–1.88) and 2.36 versus PGT (*SE* 0.24; 95% CI 1.88–2.85), whilst for GSH versus PGT, ATE was 0.72 (*SE* 0.23; 95% CI 0.26–1.18) and anxiety ATE for iCBT was 1.47 versus GSH (*SE* 0.11; 95% CI 1.25–1.68) and 2.49 versus PGT (*SE* 0.21; 95% CI 2.07–2.91), whilst for GSH versus PGT, ATE was 1.03 (*SE* 0.20; 95% CI 0.63–1.43). All the results in the comorbid group were also statistically significant.

DISCUSSION

We undertook the objective of comparing three low-intensity interventions offered within a national stepped-care service, similar in therapeutic content but differing in their mode of delivery. The demographic characteristics, overall reliable improvement, recovery and reliable recovery rates were comparable to a recent national report on IAPT service data from 2019 to 2020 (Community and Mental Health Team, 2020). Therefore, this was a sample of service users representative of the overall UK IAPT population. The overall results obtained reach the UK government target, which is a 50% recovery rate for all referrals (NHS, 2019). In line with established research across all three interventions (Wakefield et al., 2020), large effects in terms of improvement in depression and anxiety, and medium effects in terms of functioning scores, were observed. However, we found that there are significant differences in the comparative effectiveness of these treatments.

The observed reliable improvement, recovery rates and reliable recovery rates, prior to undertaking propensity score analysis, were higher for iCBT in comparison with GSH and PGT, and within-group pre-post effects were the largest for iCBT. It is worth pointing out that some baseline sociodemographic differences exist that may partly account for the descriptive differences in reliable improvement and recovery rates. iCBT users have a lower rate of psychoactive medication and long-term conditions, suggesting less 'complex' cases. For this reason, we calculated the reliable improvement and recovery rates post-propensity score to account for this. Additionally, adherence rates differed significantly between treatments, with highest adherence seen for iCBT. Furthermore, both propensity score models, which account for potential allocation bias and control for baseline covariate imbalances across interventions, showed that allocation to iCBT resulted in larger improvements in depression, anxiety and functioning post-treatment scores across 4 years' worth of data, reflected in higher reliable improvement and recovery rates. Therefore, although scores for depression, anxiety and impaired functioning decreased for all treatments, the decrease was larger and translated to a higher percentage change from caseness to non-caseness, in iCBT compared with GSH or PGT.

Low-intensity intervention comparisons

Despite the large amount of evidence backing the effectiveness of low-intensity interventions individually (Andrews et al., 2018; Etzelmüller et al., 2020; Gualano et al., 2017), comparisons between these low-intensity psychological therapies are scarce. A recent meta-analysis found no significant differences in terms of efficacy between iCBT and GSH, as well as in terms of client adherence to the interventions (Andrews et al., 2018). However, the data were taken from a relatively small sample of users, from three studies on individuals with depression, panic disorder and social phobia, respectively. Furthermore, there appears to be a lack of direct comparisons of group therapy to other evidence-based low-intensity treatments. Specifically, there is still uncertainty regarding the efficacy of iCBT and GSH compared to psychoeducational group therapy.

Implications of the findings

It is important to theorize the reasons as to why we have found a difference in treatment effect between interventions similar in content but differing in delivery within a service. There are unmeasured variables related to service implementation which may be playing a part. Decisions taken by clinicians to allocate certain users to one treatment over another could play a role, for example a user diagnosed with other comorbidities or a certain type of anxiety may be allocated to group psychotherapy where peer support could help in other ways, and this may not be reflected in changes in PHQ-9 and GAD-7 scores. In addition, the bibliographic content given via guided self-help is certainly not as vast nor interactive as that within an online platform. iCBT, with its high emphasis on user experience facilitating reading and understanding of the exercises and content, may offer more flexibility with the programme at a schedule convenient for the user, and it also offers perceived anonymity given the interactions are not face to face at the clinic, which may help users engage more via the platform.

Whilst supported iCBT has been previously evaluated as a rapid and effective treatment option within the NHS IAPT service (Learmonth et al., 2008), the current study shows its greater effectiveness compared with other low-intensity interventions within the same service over a four-year period. During this time, over 90% of iCBT users completed a second assessment (and are thus measurable for outcome analysis as per IAPT standards) versus 56% for GSH. Higher completion of assessments would increase overall treatment coverage, which is relevant given that despite broad improvements in IAPT since its inception, only 60% of referrals are treated (Clark, 2018). Gathering further information on usage, adherence, and long-term engagement of the intervention, which iCBT facilitates (Enrique et al., 2019), could also be key in understanding service users' patterns of behaviour whilst undergoing these treatments, and how best to utilize these patterns to deliver a positive clinical outcome to a greater number of patients. Further studies that analyse iCBT and its effect on different population types are warranted to help maximize the advantages this relatively new delivery method offers at an individual and system-wide level.

Limitations

This study is not without important limitations. As in any retrospective naturalistic analysis, there may exist unobserved variables which could be affecting the results. Decisions are taken by clinicians and service users regarding their treatment which could influence recovery rates. There may be a bias within the service as to which types of service users get placed on each treatment. However, propensity score modelling was implemented to reduce the selection bias that exists in the absence of randomization. By addressing this bias, the internal validity of the findings increases by isolating the effect of the treatment on the outcome. It must be stated, however, that propensity score techniques are not without their limitations themselves. Stratification is not a substitute for randomization and only ensures balance in

measured, not unmeasured confounders. Likewise, it can control for only a limited number of covariates, since stratifying on too many covariates creates groups which are too sparse and cannot reliably estimate treatment effects (Sainani, 2012). Weighting, on the other hand, can be problematic when the issue of extreme weights inflates the treatment effects, increasing bias (Harder et al., 2010). However, we aimed to minimize this effect in our sample using a truncation method (Lee et al., 2011).

Further, these data include information on the treatment, which was ultimately delivered to these service users, but cannot account for the fact that a different treatment may have been offered initially, as that information was not available to the researchers. However, this happens rarely within the service, and users get moved to another treatment usually only if scores remain the same after multiple sessions. The baseline clinical scores included in the analysis do indeed reflect those users had upon starting the current treatment taken into account for the study.

Another limitation of our study is the reliance on self-report questionnaires for our outcome data without a direct psychological assessment; however, these questionnaires are indeed the measurement method used across IAPT services and as described, have good diagnosis specificity and rates of follow-up. Our study also lacks data on engagement and adherence measures. This could provide further information as to why recovery rates are higher in certain treatments and population subgroups, and additionally, may provide clues to help increase these recovery rates even more. In addition, although our sample is representative of the IAPT population, it is from a single service, and the way low-intensity interventions are implemented in other services may indeed affect outcomes (Clark, 2018). Finally, our sample is taken from before the COVID-19 pandemic, and we look forward to further research that analyses similar data to gauge for any differences in effectiveness in these interventions during and after the pandemic, or whether other patterns emerge.

CONCLUSION

This study demonstrates that low-intensity interventions, requiring less intensive clinician inputs, are effective interventions for service users with mild to moderate presentations of anxiety and depression in stepped-care settings. Additionally, this study shows that the delivery format of low-intensity treatments matters and can be related to outcomes, with iCBT being the intervention that shows more positive results. Our study provides further evidence for services to continue and increase the use of low-intensity interventions at Step 2. This may contribute to the decrease of waiting times, increase in service users attended, with higher rates of recovery and successful discharges overall. The current COVID-19 pandemic is in no small measure contributing to an increment in mental health problems through secondary effects such as isolation, stress and economic losses, at a time where services cannot function as normal and capacity is at a tipping point. Future research should continue to utilize data on these and other measures, such as therapist time spent, to identify which interventions work best for whom, and in what setting, as growing the evidence base for effective, innovative treatments is key at this critical juncture, where demand is outpacing what is currently available. In addition, our analyses should be replicated across other services and implementation methods should be considered to identify the gold standard in achieving the best possible outcomes.

AUTHOR CONTRIBUTIONS

Jorge Palacios: Conceptualization; data curation; investigation; methodology; supervision; validation; writing – original draft. **Adedeji Adegoke:** Data curation; formal analysis; methodology; writing – original draft. **Rebecca Wogan:** Investigation; visualization; writing – original draft. **Daniel Duffy:** Investigation; writing – review and editing. **Caroline Earley:** Project administration; writing – review and editing. **Nora Eilert:** Methodology; writing – review and editing. **Angel Enrique:** Conceptualization; writing – review and editing. **Sarah Sollesse:** Validation; writing – review and editing. **Judith Chapman:** Resources; writing – review and editing. **Derek Richards:** Conceptualization; funding acquisition; methodology; supervision; writing – review and editing.

ACKNOWLEDGEMENTS

Resources from Berkshire NHS Foundation Trust have been generously given in kind for the purpose of the execution of the study: we primarily wish to thank the R&D and clinical team members for their support in bringing this study to completion. Special thanks to Seema Dihar, Data Analyst at Berkshire NHS Foundation Trust, for her enormously valuable contribution in providing the anonymous dataset which provided the basis for the study cohort. The study was funded by SilverCloud Health and Berkshire Healthcare Foundation Trust. Open access funding provided by IReL.

CONFLICT OF INTEREST

Authors Palacios, Adegoke, Wogan, Duffy, Earley, Enrique and Richards are employees of SilverCloud Health. SilverCloud Health is a commercial entity that develops computerized psychological interventions for depression, anxiety, stress and comorbid long-term conditions and sells these to Health Services globally. In England, the SilverCloud service is delivered free to patients through the Improving Access to Psychological Therapies (IAPT) programme. Commercial departments within SilverCloud Health played no role in the analysis or interpretation of data.

DATA AVAILABILITY STATEMENT

The data for this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions, being data that relates to third party service provision and evaluation, which was shared solely for the intended purpose of the objectives outlined in this paper.

ORCID

Jorge Palacios  <https://orcid.org/0000-0002-2103-5507>
 Adedeji Adegoke  <https://orcid.org/0000-0003-2870-9532>
 Rebecca Wogan  <https://orcid.org/0000-0001-6863-3064>
 Daniel Duffy  <https://orcid.org/0000-0001-9722-0437>
 Caroline Earley  <https://orcid.org/0000-0001-5303-0893>
 Angel Enrique  <https://orcid.org/0000-0003-0585-4008>
 Derek Richards  <https://orcid.org/0000-0003-0871-4078>

REFERENCES

- Alonso, J., Liu, Z., Evans-Lacko, S., Sadikova, E., Sampson, N., Chatterji, S., Abdulmalik, J., Aguilar-Gaxiola, S., Al-Hamzawi, A., Andrade, L. H., Bruffaerts, R., Cardoso, G., Cia, A., Florescu, S., de Girolamo, G., Gureje, O., Haro, J. M., He, Y., de Jonge, P., ... Thornicroft, G. (2018). Treatment gap for anxiety disorders is global: Results of the world mental health surveys in 21 countries. *Depression and Anxiety*, 35(3), 195–208. <https://doi.org/10.1002/da.22711>
- Andrews, G., Basu, A., Cuijpers, P., Craske, M. G., McEvoy, P., English, C. L., & Newby, J. M. (2018). Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. *Journal of Anxiety Disorders*, 55(4), 70–78. <https://doi.org/10.1016/j.janxdis.2018.01.001>
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424. <https://doi.org/10.1080/00273171.2011.5687>
- Baguley, C., Farrand, P., Hope, R., Leibowitz, J., Lovell, K., Lucock, M., O'Neill, C., Pxtton, R., Pilling, S., Richardson, D., Turpin, G., White, J., & Williams, C. (2010). *Good practice guidance on the use of self-help materials within IAPT services* (Issue March).
- Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. *The Lancet*, 395(10227), 912–920. [https://doi.org/10.1016/S0140-6736\(20\)30460-8](https://doi.org/10.1016/S0140-6736(20)30460-8)
- 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>
- Clark, D. M. (2018). Realizing the mass public benefit of evidence-based psychological therapies: The IAPT program. *Annual Review of Clinical Psychology*, 14(1), 159–183. <https://doi.org/10.1146/annurev-clinpsy-050817-084833>
- Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, 168(6), 656–664. <https://doi.org/10.1093/aje/kwn164>
- Community and Mental Health Team. (2020). Psychological Therapies, Annual Report on the use of IAPT Services. In *NHS Digital* (Issue February).

- Enrique, A., Palacios, J. E., Ryan, H., & Richards, D. (2019). Exploring the relationship between usage and outcomes of an internet-based intervention for individuals with depressive symptoms: Secondary analysis of data from a randomized controlled trial. *Journal of Medical Internet Research*, *21*(8), 1–13. <https://doi.org/10.2196/12775>
- Etzelmüller, A., Vis, C., Karyotaki, E., Baumeister, H., Titov, N., Berking, M., Cuijpers, P., Riper, H., & Ebert, D. (2020). Effects of internet-based cognitive behavioral therapy in routine care for adults in treatment for depression and anxiety: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *22*(8), e18100. <https://doi.org/10.2196/18100>
- Gallagher, M. W., Zvolensky, M. J., Long, L. J., Rogers, A. H., & Garey, L. (2020). The impact of Covid-19 experiences and associated stress on anxiety, depression, and functional impairment in American adults. *Cognitive Therapy and Research*, *44*(6), 1043–1051. <https://doi.org/10.1007/s10608-020-10143-y>
- Gualano, M. R., Bert, F., Martorana, M., Voglino, G., Andriolo, V., Thomas, R., Gramaglia, C., Zeppego, P., & Siliquini, R. (2017). The long-term effects of bibliotherapy in depression treatment: Systematic review of randomized clinical trials. *Clinical Psychology Review*, *58*, 49–58. [10.1016/j.cpr.2017.09.006](https://doi.org/10.1016/j.cpr.2017.09.006)
- Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, *15*(3), 234–249. <https://doi.org/10.1037/a0019623>
- Independent Mental Health Taskforce. (2016). The five year forward view for mental health. In *The Mental Health Taskforce* (Issue February).
- König, H., König, H. H., & Konnopka, A. (2019). The excess costs of depression: A systematic review and meta-analysis. *Epidemiology and Psychiatric Sciences*, *29*(30), 1–16. <https://doi.org/10.1017/S2045796019000180>
- Konnopka, A., & König, H. (2020). Economic burden of anxiety disorders: A systematic review and meta-analysis. *Pharmacoeconomics*, *38*(1), 25–37. <https://doi.org/10.1007/s40273-019-00849-7>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159. <https://doi.org/10.2307/2529310>
- Layard, R. (2006). The case for psychological treatment centres. *British Medical Journal*, *332*(7548), 1030–1032. <https://doi.org/10.1136/bmj.332.7548.1030>
- Layard, R., & Clark, D. M. (2015). *Thrive: How better mental health care transforms lives and saves money*. Princeton University Press. <https://doi.org/10.2307/j.ctvc77595>
- Learmonth, D., Trosh, J., Rai, S., Sewell, J., & Cavanagh, K. (2008). The role of computer-aided psychotherapy within an NHS CBT specialist service. *Counselling and Psychotherapy Research*, *8*(2), 117–123. <https://doi.org/10.1080/14733140801976290>
- Lee, B. K., Lessler, J., & Stuart, E. A. (2011). Weight trimming and propensity score weighting. *PLoS One*, *6*(3), e18174.
- McManus, S., Bebbington, P., Jenkins, R., & Brugha, T. (2016). *Mental health and wellbeing in England: the adult psychiatric morbidity survey 2014*. NHS Digital.
- NHS. (2019). *Improving Access to Psychological Therapies Manual*. NHS Digital.
- NICE. (2009). Depression in adults: recognition and management. In *National Institute for Health and Care Excellence*. NICE.
- NICE. (2011a). Generalised anxiety disorder and panic disorder (with or without agoraphobia) in adults: management in primary, secondary and community care (partial update). In *National Clinical Guideline* (Issue 113).
- NICE. (2011b). Common mental health problems: identification and pathways to care. In *Clinical Guideline (CG123)*. Published 25 May 2011.
- NICE. (2020). Space from Depression for treating adults with depression. In *National Institute for Health and Care Excellence* (Issue May).
- Pierce, M., Hope, H., Ford, T., Hatch, S., Hotopf, M., John, A., Kontopantelis, E., Webb, R., Wessely, S., McManus, S., & Abel, K. M. (2020). Mental health before and during the COVID-19 pandemic: A longitudinal probability sample survey of the UK population. *The Lancet Psychiatry*, *7*(10), 883–892. [https://doi.org/10.1016/S2215-0366\(20\)30308-4](https://doi.org/10.1016/S2215-0366(20)30308-4)
- Richards, D., Enrique, A., Eilert, N., Franklin, M., Palacios, J., Duffy, D., Earley, C., Chapman, J., Jell, G., Sollesse, S., Timulak, L., Palacios, J., Duffy, D., Earley, C., Chapman, J., Jell, G., Sollesse, S., & Timulak, L. (2020). A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. *NPJ Digital Medicine*, *3*(1), 85. <https://doi.org/10.1038/s41746-020-0293-8>
- Sainani, K. L. (2012). Propensity scores: uses and limitations. *PM R*, *4*(9), 693–697.
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>
- Thornicroft, G., Chatterji, S., Evans-Lacko, S., Gruber, M., Sampson, N., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Andrade, L., Borges, G., Bruffaerts, R., Bunting, B., De Almeida, J. M. C., Florescu, S., De Girolamo, G., Gureje, O., Haro, J. M., He, Y., Hinkov, H., ... Kessler, R. C. (2017). Undertreatment of people with major depressive disorder in 21 countries. *British Journal of Psychiatry*, *210*(2), 119–124. <https://doi.org/10.1192/bjp.bp.116.188078>
- Trethvan, R. (2017). Intraclass correlation coefficients: Clearing the air, extending some cautions, and making some requests. *Health Services & Outcomes Research Methodology*, *17*(2), 127–143. <https://doi.org/10.1007/s10742-016-0156-6>
- Vizard, T., Davis, J., White, E., & Beynon, B. (2020). *Coronavirus and depression in adults, Great Britain*. Office for National Statistics.
- Wakefield, S., Kellett, S., Simmonds-Buckley, M., Stockton, D., Bradbury, A., & Delgado, J. (2020). 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>

- World Health Statistics. (2017). World Health Statistics 2017: Monitoring health for the SDGs. In *WHO*. World Health Organization.
- Zahra, D., Qureshi, A., Henley, W., Taylor, R., Quinn, C., Pooler, J., Hardy, G., Newbold, A., & Byng, R. (2014). The work and social adjustment scale: Reliability, sensitivity and value. *International Journal of Psychiatry in Clinical Practice*, *18*(2), 131-138. <https://doi.org/10.3109/13651501.2014.894072>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Palacios, J., Adegoke, A., Wogan, R., Duffy, D., Earley, C., Eilert, N., Enrique, A., Sollesse, S., Chapman, J., & Richards, D. (2022). Comparison of outcomes across low-intensity psychological interventions for depression and anxiety within a stepped-care setting: A naturalistic cohort study using propensity score modelling. *British Journal of Psychology*, *00*, 1-16. <https://doi.org/10.1111/bjop.12614>