Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

Predicting early dropout in online versus face-to-face guided self-help: A machine learning approach

Paulina Gonzalez Salas Duhne [a,*], Jaime Delgadillo [a], Wolfgang Lutz [b]

[a] Clinical and Applied Psychology Unit, Department of Psychology, University of Sheffield, Cathedral Court Floor F, 1 Vicar Lane, Sheffield, S1 2LT, United Kingdom
[b] Clinical Psychology and Psychotherapy, Department of Psychology, University of Trier, D - 54286 Trier, Trier, Germany

ARTICLE INFO

Keywords:
Dropout prediction
Machine learning
Guided self-help
Computerized CBT
Precision mental healthcare

ABSTRACT

Background: Early dropout hinders the effective adoption of brief psychological interventions and is associated with poor treatment outcomes. This study examined if attendance and depression treatment outcomes could be improved by matching patients to either face-to-face or computerized low-intensity psychological interventions.
Methods: Archival clinical records were analysed for 85,664 patients who accessed face-to-face or computerized guided self-help (GSH). The primary outcome was early dropout (attending $\leq$3 sessions). Supervised machine learning analyses were applied in a training sample ($n = 55{,}529$). The trained algorithm was cross-validated in an independent test sample ($n = 30{,}135$). The clinical utility of the model was evaluated using logistic regression, chi-square tests, and sensitivity analyses in a balanced subsample.
Results: Patients who received their model-indicated treatment modality were 12% more likely to receive an adequate dose of treatment OR = 1.12 (95% CI = 1.02 to 1.24), $p = .02$, and the strength of this effect was larger in the balanced subsample (OR = 2.10, 95% CI = 1.65 to 2.68, $p < .001$). Patients had better treatment outcomes when matched to their model-indicated treatment modality.
Conclusions: Machine learning approaches may enable services to optimally match patients to the treatment modality that maximizes attendance.

Treatment dropout, defined as unilateral discontinuation of an agreed course of therapy, is a common problem in the field of psychotherapy, occurring in approximately 20% of cases (Swift & Greenberg, 2012). Early dropout -occurring in the earliest sessions-has been consistently associated with poor treatment outcomes and waste of limited healthcare resources (Barrett, Chua, Crits-Christoph, Gibbons, & Thompson, 2008; Hansen, Lambert, & Forman, 2002). In brief low-intensity psychological interventions, there is evidence that dropout is slightly higher compared to randomised controlled trials (Etzelmueller et al., 2020) or higher intensity treatments (Robinson, Delgadillo, & Kellett, 2020). Dose-response studies of low-intensity psychological interventions indicate that patients who receive a suboptimal dose (e.g., three or less sessions) have poor treatment outcomes compared to those accessing more sessions (Delgadillo et al., 2014; Robinson et al., 2020).

Treatment modality may also impact dropout rates. Historically, online interventions have been associated with a greater likelihood of dropout compared to face-to-face interventions, and this has been one of the main barriers to the uptake of effective computerized treatments (Waller & Gilbody, 2009). Although a recent meta-analysis found no significant differences in dropout between face-to-face and online cognitive behavioural therapy (CBT) interventions at an aggregate level (Andersson, Cuijpers, Carlbring, Riper, & Hedman, 2014; Etzelmueller et al., 2020; Van Ballegooijen et al., 2014), there is some initial evidence that certain subgroup of patients might present with a differential response across treatment modalities. For example, patients are more likely to dropout from computerized versus face-to-face CBT interventions if they are young, males, with lower educational level, and co-morbid anxiety (Karyotaki et al., 2015). These predictors overlap with those identified previously for dropout across other psychotherapies (Bower et al., 2013; McMurran, Huband, & Overton, 2010; Swift & Greenberg, 2012; Zimmermann, Rubel, Page, & Lutz, 2017). However, most individual predictors lack sufficient explanatory power on their own to be directly translated into clinical decisions (Van Ballegooijen et al., 2014). The ability to identify patients at high risk of dropout from psychological interventions is an important challenge for mental health services, to offer them a treatment modality that will enable patients to have an adequate dose of treatment.

Prior dropout prediction studies have been historically limited by

[*] Corresponding author.
E-mail address: p.gonzalez1@nhs.net (P. Gonzalez Salas Duhne).

https://doi.org/10.1016/j.brat.2022.104200
Received 2 March 2022; Received in revised form 9 September 2022; Accepted 12 September 2022
Available online 17 September 2022

small samples, applying heterogeneous methods, testing specific hypotheses with one or few predictors that often explain a very small proportion of variance in drop out. It is probable that dropout could be influenced by the combined effect of multiple variables, such as those described above. Further, prior dropout studies tend to have imbalanced datasets (i.e., dropout tends to have a lower base rate than treatment completion) which may undermine prediction accuracy (Japkowicz & Stephen, 2002). Given these challenges, machine learning methods could potentially be used to develop clinically-useful prediction models, since they are capable of discovering complex non-linear patterns in data and modelling the combined effect of multiple variables (Chekroud et al., 2021). However, not all machine learning methods are well-suited to predict binary events such as dropout (Bennemann, Schwartz, Giesemann, & Lutz, 2022) and many decisions made when developing machine learning algorithms need to be transparently reported to minimize risk of bias (Delgadillo, 2021). Recent studies and reviews of this literature indicate that machine learning methods could help to develop clinical prediction models to inform treatment decisions about diagnosis and treatment selection (see Chekroud et al., 2021; Dwyer, Falkai, & Koutsouleris, 2018; Yarkoni & Westfall, 2017).

The aim of the present study was to develop and evaluate a machine learning algorithm to identify patients who are more likely to drop out early from computerized versus face-to-face low-intensity cognitive behavioural therapy, and to examine if matching patients to treatments based on this algorithm could improve attendance and treatment outcomes.

1. Method

1.1. Setting and interventions

The study was conducted using anonymized archival records for 85,664 patients who accessed low-intensity psychological interventions for depression within the National Health Service (NHS) in England. Data were collected as part of routine mental healthcare over two years before the COVID-19 pandemic, and across eight NHS Trusts covering 16 Improving Access to Psychological Therapies (IAPT) services. These IAPT services covered diverse regions across England including London, Cambridge, Cheshire & Wirral, Bury, Heywood, Middleton, Rochdale, Oldham, Stockport, Tameside & Glossop, Trafford, Barnsley and East Riding.

IAPT services provide evidence-based psychological interventions through a stepped care model (Clark, 2018). Following identification of a suspected mental health disorder and referral (by GP or self-referral), an IAPT clinician conducts an initial assessment. Most patients with mild-to-moderate depression and anxiety symptoms are typically offered low-intensity psychological interventions as the initial treatment option in the stepped care model. Those who remain symptomatic after accessing a low-intensity intervention have the option to access lengthier high-intensity interventions (empirically supported psychotherapies). Patients with more severe symptoms or specific conditions for which only psychotherapy is recommended (e.g., post-traumatic stress disorder, social anxiety disorder etc.) can directly access high-intensity interventions such as cognitive behavioural therapy (CBT).

Low-intensity interventions include face-to-face guided self-help (GSH), computerized cognitive behavioural therapy (cCBT), and group-based psychoeducation. This study focused on two empirically-supported low-intensity interventions that are widely available in the IAPT programme: GSH and cCBT. Treatment selection in the participating services was informed by clinical guidelines (see Clark, 2018) and principles of shared decision-making after an initial assessment, which takes into consideration the patient's preferences about available treatment options (Richards & Whyte, 2011). In practice, the assessing therapists discuss the available treatment options and come to an agreement with the patient about which option to access.

GSH and cCBT are structured (e.g., following standard treatment protocols), didactic, low-intensity psychological interventions based on principles of CBT. These interventions aim to support people to learn about maintenance factors for common mental disorders (such as cognitive biases, avoidant and safety-seeking behaviours) and to learn and apply cognitive-behavioural coping strategies such as cognitive restructuring, problem solving, relaxation skills, graded exposure, behavioural activation, and behavioural experiments. In IAPT services, these interventions are delivered by psychological wellbeing practitioners qualified to a postgraduate level following a national curriculum (National IAPT Team, 2015). GSH is delivered face-to-face and cCBT is accessed via the internet.

GSH usually consists of 6–8 individual face-to-face sessions, each of which lasts between 20 and 30 min. The content and structure of the sessions are aligned to published practice guidelines for psychological wellbeing practitioners (Richards & Whyte, 2011; University College London, 2014). cCBT consists of highly structured interactive online modules which are similar in content to GSH, with each of the 7 to 10 modules lasting around 1 h to complete at the patients' own pace. The cCBT packages available in participating services were empirically supported; such as the 7-session modularised interventions available in the Silver Cloud platform (see Richards et al., 2020). Patients undergoing cCBT were supported by a psychological wellbeing practitioner who reviewed their individual progress and engaged with them via asynchronous personalized messaging and/or brief telephone contacts (of variable duration and frequency, dependent on the patient's needs and preferences), which prior research has suggested is clinically beneficial and conducive to a therapeutic relationship (Perera-Delcourt & Sharkey, 2019).

Ethical approval

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human subjects/patients were approved for a prior study by London City & East NHS Research Ethics Committee; January 06, 2016, Ref: 15/LO/2200, and approval was obtained for secondary analyses with the anonymized dataset. Verbal consent to use fully anonymous clinical data for research was obtained and documented in the clinical records for all patients included in the study sample.

1.2. Measures

All IAPT services routinely collect session-by-session validated patient-reported outcome measures for depression, anxiety, and functional impairment, alongside demographic and clinical information. The primary outcome measure for depression symptoms was measured by the Patient Health Questionnaire (Kroenke, Spitzer, & Williams, 2001; PHQ-9). Each of the nine items in the PHQ-9 are rated from 0 to 3, where the sum yields an overall score ranging between 0 and 27. Authors recommended a cut-off of $\geq$10 overall score to identify a major depressive disorder with an adequate trade-off between sensitivity (88%) and specificity (88%; (Kroenke et al., 2001). Reliable and clinically significant improvement (RCSI) was attained where significant change was observed in PHQ-9 scores (post-treatment was $\geq$6 points lower than pre-treatment scores) and post-treatment PHQ-9 scores were below the diagnostic cut-off ($\leq$10 points). This RCSI definition was based on the methodology proposed by Jacobson and Truax (1991), using the reliable change index ($\geq$6 points) proposed by Richards and Borglin (2011).

Secondary measures included anxiety symptoms (Spitzer, Kroenke, Williams, & Löwe, 2006; GAD-7) and functional impairment (Mundt, Marks, Shear, & Greist, 2002; WSAS). Self-reported clinical and demographic data was gathered at the initial assessment: age, gender, employment, ethnicity, Index of Multiple Deprivation (Smith et al., 2015; IMD), concurrent use of medication, long-term physical health condition, referral source, and number of sessions.

1.3. Early dropout definition

This study aimed to explore early dropout and adequate treatment dose, as defined below. Pre-treatment dropout data was not analysed as a predictor variable, but was also defined and reported in our sample, in line with recommendations advocating for transparency in adherence/dropout studies (Van Ballegooijen et al., 2014). Pre-treatment dropout was defined as patients having accessed only an initial assessment session, being referred to GSH or cCBT through shared decision making, and subsequently not attending any treatment sessions due to the patient's decision. This did not include those who were not deemed suitable during the initial assessment. Patients who accessed three or fewer sessions of treatment were classed as early dropouts in line with prior evidence from GSH interventions (Delgadillo et al., 2014). Patients who accessed four or more sessions of treatment were classed as having had an adequate dose of treatment, consistent with replicated evidence that at least four sessions are required to maximise treatment outcomes in brief and low-intensity psychological interventions (Robinson et al., 2020).

1.4. Overview of machine learning analyses

Machine learning refers to a data mining process that enables the discovery of patterns in a data-driven way, and it is commonly used to solve problems related to classification and prediction (Hastie, Friedman, & Tibshirani, 2009). Supervised machine learning refers to a process where a machine (e.g., a computerized algorithm) is trained to recognise *labels* (e.g., "this looks more like a dropout case") or to predict an outcome of interest (e.g., "this case has a 65% probability of dropping out"). Typically, the machine is provided with *labelled* examples (e.g., data labelled as belonging to completers vs. dropouts) and features (e.g., each patient's clinical and demographic characteristics) contained in a "training" dataset. Once an algorithm has been trained, its capacity to accurately make classifications or predictions is evaluated in a statistically independent "test" dataset, which is referred to as "external cross-validation".

The present study applied a supervised machine learning approach, informed by methodological recommendations for mental health studies (Delgadillo, 2021; Chekroud et al., 2021) including: a priori sample size calculation (for both training and validation sample); transparent reporting of pre-processing decisions (includes the reduction, transformation, imputation and balancing of the data); developing a targeted prescription model in a training sample; external cross-validation in a statistically independent test sample. Each of those steps is described in further detail below.

1.5. Pre-processing of data

Sample selection process and sample characteristics. The wider data set included 157,946 patients across 16 IAPT services. Of these, 95,088 patients accessed diverse low-intensity interventions: GSH (n = 84,053; 88.4%), psychoeducation groups (n = 8671; 9.1%), cCBT (1611; 1.7%), and other interventions (n = 753; 0.8%). To compare different delivery modes among equivalent low-intensity interventions, we only included patients who accessed GSH and cCBT. No other exclusion criteria were applied. The sample characteristics are summarized in Table 1. All baseline characteristics were significantly different across GSH and cCBT except for gender and use of medication. cCBT cases had lower average baseline severity on depression, anxiety, and global functioning, were younger, living in more favourable socioeconomic circumstances, more likely to be White British, employed, without a long-term medical condition, and self-referred. While systematic differences are expected in naturalist samples, to account for those differences, we used propensity score matching in secondary analyses (explained further in subsequent data analysis strategy).

Table 1
Sample characteristics across treatment modalities.

	All cases	GSH	cCBT
	N = 85,664	n = 84053	n = 1611
	Mean (SD) or %	Mean (SD) or %	Mean (SD) or %
PHQ-9	14.96 (6.22) *	15.01 (6.22)	12.32 (6.13)
GAD-7	13.53 (5.06) *	13.57 (5.05)	11.72 (5.16)
WSAS	19.63 (9.85) *	19.69 (9.86)	16.33 (8.81)
Age	39.81 (15.11) *	39.9 (15.1)	36.8 (13.0)
IMD decile	4.93 (2.79) *	4.91 (2.78)	5.92 (2.76)
Gender (% female)	65.10	65.16	62.24
Ethnicity (% White British)	84.60 *	84.48	90.77
Employment (% unemployed)	23.60 *	23.83	10.95
Referral (% self-referral)	51.68 *	51.51	60.55
Long term condition (% with LTC)	30.91 *	31.02	25.32
Medication (% prescribed and taking medication)	44.38	44.42	41.98

GSH = Guided Self-Help; cCBT = Computerized Cognitive Behavioural Therapy; PHQ-9 = Patient Health Questionnaire; GAD-7 = Generalized Anxiety Disorder Questionnaire; WSAS = Work and Social Adjustment Scale; IMD Decile = Index of multiple deprivation in deciles, where 1 represent the most deprived, and 10 the most affluent; * = significant differences at p < .001 between GSH and cCBT.

Sample size estimation. We performed an a priori calculation of the minimum sample size needed for the external test sample, to ensure the sample was large enough to reliably evaluate the clinical utility of the prediction model. This sample size calculation applied the four criteria proposed by Archer et al. (2021), which focus on estimating the proportion of the variance explained, the agreement between average predicted and observed values, the calibration slope and variance of observed outcomes. Each of the four criteria yield a sample size requirement, and the authors suggest utilizing the largest sample size requirement as a baseline. The largest sample size required to meet all criteria was n = 235.

Reduction of categorical variables. Categorical variables were collapsed into fewer categories before conducting any statistical analyses to minimize small-sample bias. This was done in line with recommendations for machine learning in psychotherapy research (Delgadillo, 2021), and following the methods outlined in a prior study (Delgadillo & Gonzalez Salas Duhne, 2020). Referral source originally included General Practitioner (GP), self, and other, where others accounted for less than 5% of the cases. Therefore, we grouped them into two categories: self and other. Similar procedures were done for: age (reduced into decades), gender (male/female, others were classified as missing data), ethnicity (White British/other), employment (employed/unemployed), disability (disabled/not disabled), and medication (prescribed and taking medication/not prescribed or taking medication).

Partitioning of data into training and test samples. We randomly assigned each of the 16 IAPT services into either the training or test sample aiming for a 70:30% split. This partition allowed for data from test and training samples to be completely independent (different clients, therapists, and geographical regions), as services were not partitioned. The partition method also ensured an equal balance of interventions provided (cCBT/GSH) and comparable base rates of pre-treatment dropout and early dropout. The training sample included 55,529 patients (64.8%) and the test sample included 30,135 patients (35.2%). The test sample included n = 540 patients who had cCBT (the intervention with the smallest sample), making this sample size almost twice as large as the minimum sample size required according to the calculations described above. In machine learning analyses larger data sets tend to have better prediction accuracy (Dwyer et al., 2018). While according to the sample size calculation a smaller sample would have sufficed to ensure adequate power, the authors decided to include more available cases for improvement on the prediction model.

Missing data. To address missing data in predictor variables (which carry the risks of reduced power, reduced external validity, and overfitting), we conducted multiple imputation separately in the training and test samples ensuring their independence. Demographic and clinical features (>10% and <30%) missing data were imputed with a Markov chain Monte Carlo method (Gilks, Richardson, & Spiegelhalter, 1995; MCMC) averaging 25 iterations for each dataset. Variables with more than 30% of missing data were excluded from the analyses (Stekhoven & Bühlmann, 2012).

Balancing of data. The naturalistic dataset (i.e., data came from routine practice) had an inherent class imbalance due to the low base rate of early dropout (i.e., more cases were classed as having dropped out rather than those classed as not having dropped out). Class imbalance a clinical prediction model may lead to an overestimation of the model's performance and may undermine the prediction accuracy (Hand & Vinciotti, 2003; Menardi & Torelli, 2014). Therefore, we divided the training dataset into subsamples to explore the potential effects of class imbalance via internal cross-validation, while maintaining the independence of data from the test sample. We compared the prediction accuracy between two internal (training) partitions: a random 50:50 split and a down-sampling approach, which has been recommended to explore the potential effects of class imbalance in dropout (Bennemann et al., 2022). Following the internal cross-validation of both approaches, sensitivity and specificity analyses comparing both approaches indicated there was no gain in prediction accuracy by correcting for class imbalance using random down-sampling. However, down-sampling considerably reduces the size of the available training sample. Therefore, we included the whole of the training sample in further analyses. It is noteworthy that despite the class imbalance, the number of early dropout cases in the training samples for each intervention was larger than the number of cases as recommended by Bennemann et al. (2022). Further details about predictive accuracy in the training sample are described in the Supplemental Materials.

1.6. Data analysis strategy

Development of prognostic indices and prediction equation. Prognostic indices were developed for cCBT and GSH cases separately within the training sample only. The prognostic index indicated the probability of early dropout. Clinical and demographic baseline patient characteristics (summarized in Table 1) were entered into a supervised machine learning approach, using Elastic Net regularization and optimal scaling, to determine the combined weight of patient characteristics that may predict early dropout in each of these interventions.

Elastic net combines LASSO (Least Absolute Shrinkage and Selection Operator; also called L1 regularization) and Ridge regression (also called L2 regularization). Ridge regression penalizes regression coefficients without excluding any predictors, while LASSO is a parsimonious way of selecting predictors by shrinking the coefficients to zero for variables that do not have reliable predictive value. Elastic net is adequate for use in contexts where multicollinearity between predictors is expected, and it achieves the goals of variable selection (via LASSO) and differential weight-setting (via Ridge). We pre-specified a penalty term increment of 0.01 and maximum of 0.10 (alpha hyperparameter $\alpha = 0.10$, defined a priori). Ten-fold internal cross-validation loops were applied to train and test each iterative model that was generated by increasing the penalty term by 0.01 each time (Efron & Tibshirani, 1997). A grid search procedure was applied to select the final prediction model among all available iterations, using the one-standard-error rule, which selected the most parsimonious (least complex) model within one standard error of the model that balanced the highest index of explained variance (pseudo r^2) and lowest expected prediction error. An additional feature of this machine learning approach was the application of optimal scaling, which models nonlinear relationships between dependent and independent variables by fitting splines (Gifi, 1990).

Development of a targeted prescription algorithm. Based on the prognostic equations yielded by the above Elastic net procedure, we constructed a personalized advantage index (PAI; DeRubeis et al., 2014) by expressing the difference between the two prognostic indices in a predicted probability scale (i.e., the predicted probability of early dropout, for each of the two interventions, cCBT vs GSH, for each patient). A positive % indicated participants were more likely to dropout early in cCBT (and therefore the model-indicated treatment was GSH), and a negative % indicated participants were more likely to dropout early in GSH (and therefore the model-indicated treatment was cCBT). For example, a PAI = 35% would indicate that a patient has a 35% higher probability of early dropout if they receive cCBT relative to accessing GSH.

External cross-validation procedure. We then applied the same targeted prescription algorithm to all cases in the test sample and determined with the same PAI rules, which was the model-indicated treatment. In this way, based on each patient's baseline characteristics, the algorithm could make a prediction about each patient's likelihood of early dropout for each type of treatment (the actual treatment received, and the counterfactual prediction for the treatment that was not accessed). We applied logistic regression to compare early dropout rates in cases that received the model-indicated (i.e., their "optimal" treatment) versus patients who did not receive their model-indicated treatment (i.e., received the "suboptimal" treatment). Subsequently, we compared the observed RCSI in cases between cases that did and those that did not receive their model-indicated using chi-square analyses and odds ratios adjusted after partialing baseline depression severity. To determine if baseline severity influenced this model, we used a two-way ANOVA regression after partialing baseline severity to compare post-treatment PHQ-9 scores between groups.

Propensity Score Matching. To further examine the robustness of the prediction model, we used propensity score matching to test the algorithm in a subsample of cases from the test sample, with equivalent numbers of cCBT and GSH cases. Propensity score matching (PSM) has been recommended for naturalistic data to guide treatment selection (Kessler, Bossarte, Luedtke, Zaslavsky, & Zubizarreta, 2019) by statistically balancing the baseline differences across covariates (Rosenbaum & Rubin, 1983). All cCBT cases from the test sample were selected and then matched to similar GSH cases using the PSM case-control matching procedure. All clinical and demographic variables were entered into a logistic regression PSM model predicting cCBT membership, using a one-to-one nearest neighbours approach, specifying a 0.2 calliper defined a priori, while allowing replacement. PSM resulted in 540 cCBT cases and 540 GSH cases with balanced characteristics. We then conducted the same statistical analyses described above (logistic regression to predict early dropout, and RCSI after partialing baseline severity).

2. Results

2.1. Base rate of pre-treatment dropout, early dropout, and adequate treatment dose

The base rate of pre-treatment dropout (never attending any therapy sessions) in the full sample was 29.08%, and the base rate of early dropout (accessing 3 or fewer treatment sessions) was 54.09%. Independent samples t-tests and chi-square tests were calculated to compare baseline differences between cCBT and GSH (Table 2). This naturalistic sample differed systematically across treatments in recovery, pre-treatment drop out, early dropout, and treatment length. It is noteworthy that most of the cases of cCBT that dropped out early did so at the

Table 2
Comparison of recovery and dropout base rate across treatment modality.

	All cases	GSH	cCBT	test statistic	p
	85664	N = 84053	N = 1611		
RCSI (%)	14922 (17.42%)	14652 (17.43%)	270 (16.76%)	x^2 (1) = 20.31	<.001
Mean last PHQ-9 (SD)	10.75 (7.31)	10.81 (7.32)	7.67 (6.04)	t (79492) = −96.31	<.001
Mean number of sessions (SD)	3.69 (2.86)	3.59 (2.60)	4.83 (3.02)	t (85662) = 18.94	<.001
Pre-treatment dropout (% <1 treatment session)	24915 (29.08%)	24328 (28.94%)	587 (36.44%)	x^2 (1) = 35.84	<.001
Early dropout (% ≤3 treatment sessions)	47031 (54.90%)	46400 (55.20%)	631 (39.17%)	x^2 (1) = 151.52	<.001
Adequate dose (% ≥4 treatment sessions)	38209 (44.60%)	37258 (44.33%)	951 (59.03%)	x^2 (1) = 151.52	<.001

Note. GSH = Guided self-help; cCBT = computerized self-help based on cognitive behavioural therapy. RCSI = Reliable and Clinically significant improvement. PHQ-9 = Patient Health Questionnaire.

point of referral (i.e., they never activated their cCBT account), while the opposite pattern was observed for GSH (indicated by the difference between pre-treatment dropout and early dropout across treatments).

For those who accessed at least one treatment session/module, the mean number of sessions across treatments was 3.69, with cCBT cases showing a marginally higher mean number of sessions (4.83, $SD = 3.02$) compared to GSH (3.59, SD = 2.60; t (85662) = 74.41, $p < .001$).

Exploratory analyses of the relationship between the number of sessions and RCSI (Fig. 1), clearly indicate a relationship between treatment duration and outcomes. Those who received an adequate dose of low-intensity treatment had between a 38% and 55% probability of meeting RCSI criteria compared to between 0 and 28% probability of meeting RCSI criteria in those who dropped out early. Furthermore, the number of sessions attended explained 7.6% of the variance in RCSI (Nagelkerke r^2), and participants were 21% more likely to meet RCSI per each additional session attended (x^2 (1) = 4221.05; $p < .001$).

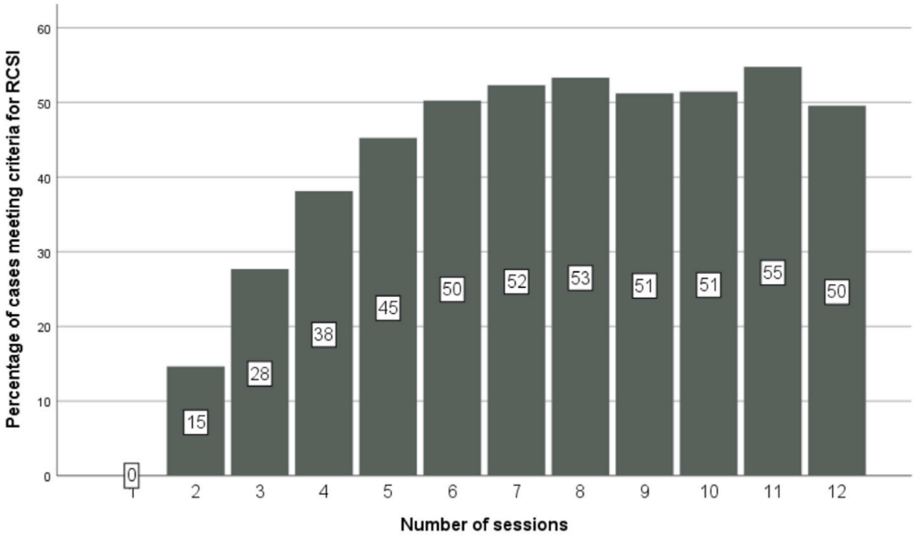
2.2. Predictors of early dropout

Table 3 presents the regularized coefficients for the variables selected as the most reliable predictors of early dropout in the training sample. The regularization procedure shrunk the beta coefficients for some variables to zero, where they had no reliable prognostic value. This variable selection procedure selected all eleven variables for GSH and only five variables for cCBT. In GSH, 1.6% of the variance in early dropout was explained by the model ($r^2 = .016$, $F(21,54144) = 41.85$, $p < .001$), while in cCBT 2.3% of the variance was explained by the model ($r^2 = .023$, F [12,1034] = 1.39, $p = .162$). The five common variables that predicted higher probably of early dropout across treatment were: younger age, ethnicity (patients from an ethnic minority were more likely to dropout of cCBT and less likely in GSH), socioeconomic deprivation (patients living in more deprived circumstances were more likely to dropout of cCBT but less likely to drop out of GSH), medication

Table 3
Regularized regression coefficients for the prediction of early dropout in guided self-help and computerized CBT.

	GSH Training sample r-square = .017		cCBT Training sample r-square = .023	
Variables	B	SE	B	SE
PHQ-9	.079	.005	.030	.028
GAD-7	−.034	.007	.000	.014
WSAS	−.025	.005	.000	.019
Gender (% female)	.007	.004	.000	.015
Age	.047	.004	.050	.024
Ethnicity (% White British)	.021	.004	.026	.026
IMD decile	.005	.004	−.007	.022
Employment (% unemployed)	.055	.005	.000	.007
Referral (% self-referral)	.007	0004	.000	.007
Long term condition (% with LTC)	.014	.004	.000	.012
Medication (% prescribed and taking medication)	.016	.004	.001	.019

B = regularized regression coefficient; SE = standard error; GSH = Guided self-help; cCBT = computerized self-help based on cognitive behavioural therapy. PHQ-9 = Patient Health Questionnaire; GAD-7 = Generalized Anxiety Disorder Questionnaire; WSAS = Work and Social Adjustment Scale; IMD Decile = Index of multiple deprivation in deciles, where 1 represent the most deprived, and 10 the most affluent.



RCSI = Reliable and Clinically Significant Improvement.

Fig. 1. Dose-response curve of recovery per number of treatment sessions. RCSI = Reliable and Clinically Significant Improvement.

(those taking medication more likely to drop out of cCBT and less likely in GSH), and higher baseline severity of depression symptoms (more likely to drop out in general). The other six variables that predicted a higher probability of early dropout specifically in GSH were: being male, unemployed, having a long-term medical condition, being referred by someone else (not self-referral), and having a higher baseline severity of anxiety symptoms and functional impairment. The main predictor demonstrated by the magnitude of the regularized coefficients in cCBT was age (younger, below 30's, more likely to drop out), and in GSH was employment (unemployed, more likely to drop out).

2.3. External cross-validation

The prediction models and PAI equation developed in the training sample ($n = 55529$) were applied in the test sample ($n = 30,135$). According to the PAI equation, few cases were classified as having accessed their model-indicated treatment across therapies in the test sample. From the GSH cases (n = 29595), 95.88% ($n = 28376$) did not receive their model-indicated treatment (which would have been cCBT) and only 4.12% ($n = 1219$) received their model-indicated treatment. From the cCBT cases ($n = 540$), 97.78% ($n = 528$) received their model-indicated treatment and 2.2% ($n = 12$) received the suboptimal treatment (GSH). According to the PAI equation, cCBT was the model indicated for 95.92% ($n = 28904$) of the cases, while only 1.79% ($n = 540$) actually received cCBT in the whole of the test sample.

Cases that received the model-indicated treatment were 12% more likely to receive an adequate dose of treatment (rather than to drop out early) compared with those who received their suboptimal treatment (x^2 [DF = 1] = 5.37, $p = .020$; OR = 1.12 [95% CI = 1.02, 1.24]).

Exploring the impact on clinical outcomes, cases that received their model-indicated treatment were more likely to meet RCSI criteria (30%) compared to those who received the suboptimal treatment (26%), (x^2 [DF = 1] = 9.85, p < .01; OR = 1.26 [95% CI = 1.09, 1.46]), and this was statistically significant after partialing baseline severity (adjusted OR = 1.17 [95% CI = 1.02, 1.36]; p = .029). Similarly, a two-way ANOVA showed that receiving the model-indicated treatment significantly increased the probability of having lower post-treatment PHQ-9 scores after partialing baseline severity (F [2,25454] = 13170.79, $p <$ .001; model-indicated treatment Beta = −.375, [95% CI = −.65, −0.11]).

2.4. Sensitivity analysis using propensity score matching

cCBT cases in the test sample (n = 540), were matched to similar test-sample GSH cases on all clinical and demographic variables using propensity score matching. The treatment selection algorithm developed in the testing sample was then applied only to these case-control matched cases. Matched cases that received their model-indicated treatment were twice as likely to receive an adequate dose (rather than to drop out early) compared with those who received the suboptimal treatment (x^2 [DF = 1] = 36.07, $p < .001$; OR = 2.10 [95% CI = 1.65, 2.68]).

Cases that received their model-indicated treatment were more likely to meet RCSI criteria (36%) compared to those who received the suboptimal treatment (22%), (x^2 [DF = 1] = 15.36, $p < .001$; OR = 1.97 [95% CI = 1.40, 2.77]), and this was statistically significant after partialing baseline severity (adjusted OR = 1.69 [95% CI = 1.19, 2.40]; $p =$ .003). Similarly, a two-way ANOVA analysis showed that receiving the model-indicated treatment significantly increased the probability of having lower post-treatment PHQ-9 scores after partialing baseline severity (F [2929] = 520.99, $p < .001$; model-indicated treatment Beta = −1.51, [95% CI = −2.21, −0.81]). Fig. 2 summarizes the effect sizes of the prediction models in the full sample and in the subsample using propensity score matching.

3. Discussion

Early dropout is a major barrier to the adoption of brief, low-intensity, evidence-based psychological interventions. Our results show that, overall, nearly one third (29%) of patients offered brief interventions in routine care never access a single session of treatment, and more than half (54%) of those who do initiate treatment drop out early and do not receive an adequate dose. Consistent with prior evidence (e.g., Delgadillo et al., 2014), the present results show that patients who drop out early have poorer treatment outcomes. Given the detrimental impact of dropout, this study developed a targeted prescription algorithm using machine learning methods, aiming to support treatment selection in a way that minimises early dropout. Overall, the results indicate that early dropout could be significantly reduced and treatment outcomes could be improved by using targeted prescription of low intensity treatment options instead of the usual treatment selection procedure that is applied in routine care. These results were verified in a statistically independent test sample and they remained robust in sensitivity analyses that addressed potential confounders.

A particularly striking finding was that the vast majority of patients

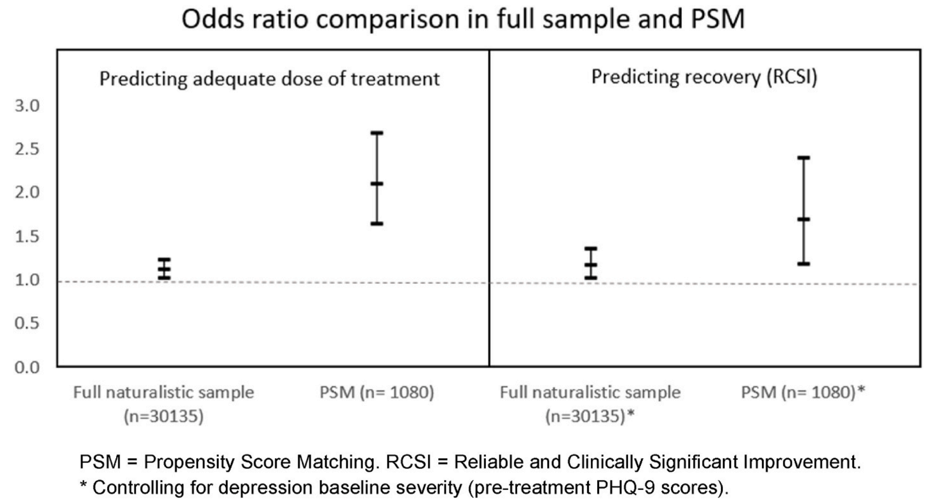


Fig. 2. Effect size of prediction models for adequate treatment dose and recovery.
PSM = Propensity Score Matching. RCSI = Reliable and Clinically Significant Improvement.
* Controlling for depression baseline severity (pre-treatment PHQ-9 scores).

in this naturalistic sample were referred for GSH (98.1% - see Table 1), even though cCBT was available in all participating services and it had minimal waiting times due to its computerized format. This observation is likely to reflect patients' preferences for face-to-face treatment, since both choices are typically offered to patients with mild-to-moderate symptoms following principles of shared decision-making. An alternative explanation may be that some clinicians may not consistently offer the choice of cCBT, thus biasing treatment selection towards GSH, although we had insufficient data to examine this. Nevertheless, it is clear that cCBT is rarely accessed in this treatment setting, despite its proven efficacy (see Richards et al., 2020). Yet, the present findings indicate that the machine learning algorithm would recommend cCBT for around 96% of patients with mild-to-moderate symptoms. This suggests that many patients currently accessing GSH could indeed benefit from cCBT, even though it may not immediately seem like a preferable treatment option. Hence, the shared decision-making approach used in routine care seems to be error prone, since it results in patients receiving a suboptimal treatment option in many cases. Offering further information to patients using a machine learning algorithm, such as the increased probability of treatment completion and symptomatic recovery with treatment A vs. treatment B, could be a viable way to optimise treatment selection while retaining principles of shared decision-making. Offering online treatments to those who are most likely to access an adequate dose could carry a substantial benefit in reducing the costs of routine healthcare while achieving similar clinical outcomes (Barak, Hen, Boniel-Nissim, & Shapira, 2008). The present results suggest that targeted prescription could improve the cost-effectiveness of services offering these types of low-intensity interventions, although this inference is drawn from observational data, and this should be empirically tested using a prospective randomised controlled trial design.

The overall base rate of early dropout was higher in face-to-face GSH and lower in computerized cCBT, contrary to prior literature where no differences have been found at an aggregate level across treatment modalities (Andersson et al., 2014; Etzelmueller et al., 2020; Van Ballegooijen et al., 2014). This discrepancy with prior evidence is likely to be explained by the treatment selection approach applied in this clinical context. As shown in Table 1, characteristics associated with increased probability of early dropout from GSH were particularly high in the GSH sample relative to the cCBT sample (e.g., higher depression severity, anxiety severity and functional impairment, and higher percentage of unemployed patients). This distribution of patient characteristics is systematically different between samples that accessed GSH and cCBT (contrary to the balanced samples found in randomised controlled trials).

The present study elucidated several features that are associated with dropout from both of these treatment modalities. Younger people with higher baseline depression symptoms were generally more likely to drop out, although the strength of these associations varied across treatments. A range of other predictors summarized in the results section predicted early drop out differentially in GSH vs. cCBT. A prior meta-analysis with data from 10 RCTs of self-help web-based interventions for depression (equivalent to cCBT in the present study), found that being younger, male, having lower education and comorbid anxiety increased the risk of dropping out (Karyotaki et al., 2015). Age was the only common predictor with our study. This highlights a current challenge in the field of precision mental healthcare: the diversity of variables collected across studies precludes us from drawing firm conclusions that generalise across samples and settings. Nevertheless, the present results show that leveraging information from multiple variables (without drawing cause-effect inferences about each predictor in the model) can help to develop clinically useful targeted prescription methods for the treatment setting where the data is collected.

3.1. *Limitations and methodological considerations*

The presence of confounding by indication (i.e., systematic differences in which patients are referred to GSH vs. cCBT through shared decision-making) is an important confounder. It is likely that patients have a strong preference for GSH and that clinicians are less likely to recommend cCBT for systematic reasons (e.g., concerns about risk/safety, likelihood of engagement, severity of symptoms, computer literacy, etc.). To address this methodological issue, we carried out rigorous sensitivity analyses using the PSM method. Class imbalance (different base rates of early drop out across treatments) was also a methodological challenge, which can bias prediction models (Hand & Vinciotti, 2003). However, the results remained robust after using a down-sampling technique through the application of PSM. While we cannot eliminate confounding by indication or class imbalance, the PSM case-control matching process adjusts for these issues statistically, which potentially yields the most robust results that we can derive from observational data. Future clinical trials of targeted prescription could benefit from including suitable decision rules to account for those systematic differences in situations where they may be clinically appropriate (e.g., do not recommend an online intervention where individuals present with suicide risk).

The large sample and independent cross-validation procedure used in the present study are essential to avoid overfitting and in line with recommendations (Perlis, 2016). However, the study and data pre-processing strategy were not pre-registered in the public domain and were conducted fully in retrospective data. Furthermore, in the cCBT condition, the time spent via asynchronous messaging with therapists was not known or adjusted for, potentially introducing differences in treatment delivery not accounted for.

4. Conclusions

In summary, it is possible to predict which patients are at risk of early drop out from brief and low-intensity psychological interventions such as GSH and cCBT, using patient-reported data collected at initial pre-treatment assessments. This information can be used to decide which type of intervention to offer to each patient, in a targeted way, to maximise engagement with treatment. Implementing this targeted prescription model has potential to reduce early dropout and to improve the cost-effectiveness of services offering low-intensity interventions. The present evidence warrants the investigation of the above hypothesis prospectively, using a randomised controlled trial design.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Paulina Gonzalez Salas Duhne: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Jaime Delgadillo:** Conceptualization, Methodology, Formal analysis, Writing – review & editing, Supervision. **Wolfgang Lutz:** Methodology, Formal analysis, Writing – review & editing.

Declaration of competing interest

None.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2022.104200.

References

Andersson, G., Cuijpers, P., Carlbring, P., Riper, H., & Hedman, E. (2014). Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: A systematic review and meta-analysis. *World Psychiatry, 13*(3), 288–295. https://doi.org/10.1002/wps.20151

Archer, L., Snell, K. I., Ensor, J., Hudda, M. T., Collins, G. S., & Riley, R. D. (2021). Minimum sample size for external validation of a clinical prediction model with a continuous outcome. *Statistics in Medicine, 40*(1), 133–146. https://doi.org/10.1002/sim.8766

Barak, A., Hen, L., Boniel-Nissim, M., & Shapira, N. A. (2008). A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. *Journal of Technology in Human Services, 26*(2–4), 109–160. https://doi.org/10.1080/15228830802094429

Barrett, M. S., Chua, W.-J., Crits-Christoph, P., Gibbons, M. B., & Thompson, D. (2008). Early withdrawal from mental health treatment: Implications for psychotherapy practice. *Psychotherapy: Theory, Research, Practice, Training, 45*(2), 247–267. https://doi.org/10.1037/0033-3204.45.2.247

Bennemann, B., Schwartz, B., Giesemann, J., & Lutz, W. (2022). Predicting patients who will drop out of out-patient psychotherapy using machine learning algorithms. *The British Journal of Psychiatry, 220*(4), 192–201. https://doi.org/10.1192/bjp.2022.17

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D. A., Gilbody, S., et al. (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

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., et al. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry, 20*(2), 154–170.

Clark, D. M. (2018). Realizing the mass public benefit of evidence-based psychological therapies: The IAPT program. *Annual Review of Clinical Psychology, 14*, 159–183. https://doi.org/10.1146/annurev-clinpsy-050817-084833

Delgadillo, J. (2021). Machine learning: A primer for psychotherapy researchers. *Psychotherapy Research, 31*(1), 1–4. https://doi.org/10.1080/10503307.2020.1859638

Delgadillo, J., & Gonzalez Salas Duhne, P. (2020). Targeted prescription of cognitive–behavioral therapy versus person-centered counseling for depression using a machine learning approach. *Journal of Consulting and Clinical Psychology, 88*(1), 14–24. https://doi.org/10.1037/ccp0000476.

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

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The personalized advantage index: Translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS One, 9*(1), Article e83875. https://doi.org/10.1371/journal.pone.0083875

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology, 14*, 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Efron, B., & Tibshirani, R. (1997). Improvements on cross-validation: The 632+ bootstrap method. *Journal of the American Statistical Association, 92*(438), 548–560. https://doi.org/10.1080/01621459.1997.10474007

Etzelmueller, A., Vis, C., Karyotaki, E., Baumeister, H., Titov, N., Berking, M., … Ebert, D. 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), Article e18100. https://doi.org/10.2196/18100

Gifi, A. (1990). *Nonlinear multivariate analysis*. Chichester, England: Wiley.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (Eds.). (1995). *Markov chain Monte Carlo in practice*. CRC press.

Hand, D. J., & Vinciotti, V. (2003). Local versus global models for classification problems: Fitting models where it matters. *The American Statistician, 57*(2), 124–131. https://doi.org/10.1198/0003130031423

Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9*(3), 329–343. https://doi.org/10.1093/clipsy.9.3.329

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science.

Jacobson, N., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis, 6*(5), 429–449. https://doi.org/10.3233/IDA-2002-6504

Karyotaki, E., Kleiboer, A., Smit, F., Turner, D., Pastor, A., Andersson, G., … Cuijpers, P. (2015). Predictors of treatment dropout in self-guided web-based interventions for depression: An 'individual patient data' meta-analysis. *Psychological Medicine, 45*(13), 2717–2726. https://doi.org/10.1017/S0033291715000665

Kessler, R. C., Bossarte, R. M., Luedtke, A., Zaslavsky, A. M., & Zubizarreta, J. R. (2019). Machine learning methods for developing precision treatment rules with observational data. *Behaviour Research and Therapy, 120*, Article 103412. https://doi.org/10.1016/j.brat.2019.103412

Kroenke, K., Spitzer, R. L., & Williams, J. B. (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

McMurran, M., Huband, N., & Overton, E. (2010). Non-completion of personality disorder treatments: A systematic review of correlates, consequences, and interventions. *Clinical Psychology Review, 30*(3), 277–287. https://doi.org/10.1016/j.cpr.2009.12.002

Menardi, G., & Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery, 28*(1), 92–122. https://doi.org/10.1007/s10618-012-0295-5

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The work and social adjustment scale: A simple measure of impairment in function00ing. *The British Journal of Psychiatry, 180*(5), 461–464. https://doi.org/10.1192/bjp.180.5.461

National IAPT Team. (2015). *National curriculum for the education of psychological wellbeing practitioners* (3rd ed.). London: NHS England/Department of Health.

Perera-Delcourt, R., & Sharkey, G. (2019). Patient experience of supported computerized CBT in an inner-city IAPT service: A qualitative study. *The Cognitive Behaviour Therapist, 12*, E13. https://doi.org/10.1017/S1754470X18000284

Perlis, R. H. (2016). Abandoning personalization to get to precision in the pharmacotherapy of depression. *World Psychiatry, 15*(3), 228–235. https://doi.org/10.1002/wps.20345

Richards, D. A., & Borglin, G. (2011). Implementation of psychological therapies for anxiety and depression in routine practice: Two year prospective cohort study. *Journal of Affective Disorders, 133*(1–2), 51–60. https://doi.org/10.1016/j.jad.2011.03.024

Richards, D., Enrique, A., Eilert, N., Franklin, M., Palacios, J., Duffy, D., et al. (2020). A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. *NPJ Digital Medicine, 3*, 85. https://doi.org/10.1038/s41746-020-0293-8

Richards, D. A., & Whyte, M. (2011). *Reach out. National programme student materials to support the delivery of training for psychological wellbeing practitioners delivering low intensity interventions*. London, UK: Rethink.

Robinson, L., Delgadillo, J., & Kellett, S. (2020). The dose-response effect in routinely delivered psychological therapies: A systematic review. *Psychotherapy Research, 30*(1), 79–96. https://doi.org/10.1080/10503307.2019.1566676

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Smith, T., Noble, M., Noble, S., Wright, G., McLennan, D., & Plunkett, E. (2015). *The English indices of deprivation 2015*. London: Department for Communities and Local Government.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092–1097, 1092–7. pmid:16717171.

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

Swift, J. K., & 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

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

Van Ballegooijen, W., Cuijpers, P., Van Straten, A., Karyotaki, E., Andersson, G., Smit, J. H., et al. (2014). Adherence to internet-based and face-to-face cognitive behavioural therapy for depression: A meta-analysis. *PLoS One, 9*(7), Article e100674. https://doi.org/10.1371/journal.pone.0100674

Waller, R., & Gilbody, S. (2009). Barriers to the uptake of computerized cognitive behavioural therapy: A systematic review of the quantitative and qualitative evidence. *Psychological Medicine, 39*(5), 705–712. https://doi.org/10.1017/S0033291708004224

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Zimmermann, D., Rubel, J., Page, A. C., & Lutz, W. (2017). Therapist effects on and predictors of non-consensual dropout in psychotherapy. *Clinical Psychology & Psychotherapy, 24*(2), 312–321. https://doi.org/10.1002/cpp.2022