
Article:
Nye, A., Delgadillo, J. orcid.org/0000-0001-5349-230X and Barkham, M. (2023) Efficacy of personalized psychological interventions: a systematic review and meta-analysis. Journal of Consulting and Clinical Psychology, 91 (7). pp. 389-397. ISSN 0022-006X

https://doi.org/10.1037/ccp0000820

Citation: Nye, A., Delgadillo, J., & Barkham, M. (2023). Efficacy of personalized psychological interventions: A systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology*. doi: 10.1037/ccp0000820

Efficacy of Personalized Psychological Interventions:

A Systematic Review and Meta-Analysis

Arthur Nye*, Jaime Delgadillo, & Michael Barkham

Clinical and Applied Psychology Unit, Department of Psychology,

University of Sheffield, UK

Conflicts of interest: None.

Data sharing policy: Data access requests should be addressed to the corresponding author.

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

*** Corresponding author:** Arthur.nye@shsc.nhs.uk

Abstract

Objective: To evaluate the efficacy of different approaches to personalization in psychological therapy.

Method: This was a systematic review and meta-analysis of randomized controlled trials that compared the mental health outcomes of personalized treatment with standardized treatment and other control groups. Eligible studies were identified through three databases (Scopus, PsychINFO and Web of Science). We conducted a narrative synthesis and random effects meta-analysis of available outcomes data, including subgroup analyses to examine sources of effect size heterogeneity. The review protocol was pre-registered in the Open Science Framework.

Results: Seventeen studies (N = 7617) met inclusion criteria for the review, nine of which (N = 5134) provided sufficient data for inclusion in meta-analysis. Eight studies were classed as having high risk of bias, eight had moderate risk, and one had low risk. There was no significant evidence of publication bias. A statistically significant effect size was found in favor of personalized treatments relative to standardized treatments (d = 0.22 [95% CI = 0.05, 0.39], p = 0.011). When studies with a high risk of bias were removed, this effect size was smaller but remained statistically significant (d = 0.14 [95% CI = 0.08, 0.20], p <0.001).

Conclusions: Current evidence indicates that personalization is an effective strategy to improve outcomes from psychological therapy, and the seemingly small effect size advantage of personalization could have an important impact at a clinical population level.

Public health significance:

- Personalized psychological interventions are associated with superior outcomes compared to standardized psychological interventions.
- If applied across a clinical population of patients accessing therapy, the small effect size in favor of personalized treatment has the potential to improve outcomes for a large number of patients.
- In particular, personalized treatment has the potential to improve outcomes in the context of depression.

Keywords: personalized medicine; stratified medicine; precision mental health care; treatment matching; psychotherapy

Evidence-based psychological therapies are effective for the treatment of various mental health problems (Barkham & Lambert, 2021). For example, a meta-analysis of 40 years of studies examining psychological therapies for depression found effect sizes of 0.71 for cognitive behavioral therapy (CBT), 0.74 for behavioral activation, 0.60 for interpersonal psychotherapy, and 0.61 for short-term psychodynamic therapy (Cuijpers, 2017). In another example, a meta-analysis of 41 studies comparing CBT to placebos for various anxiety disorders, in addition to obsessive compulsive disorder (OCD) and post-traumatic stress disorder (PTSD), found between-group effect sizes of 1.01 for generalized anxiety disorder (GAD), 0.41 for social anxiety disorder (SAD), 0.39 for panic disorder (PD), 1.13 for OCD, and 0.48 for PTSD on disorder-specific measures (Carpenter et al., 2018). Furthermore, According to a network meta-analysis of 101 studies including 11,910 patients, psychological therapy demonstrates similar efficacy to pharmacotherapy (Cuijpers et al., 2020). Overall, various forms of therapy are effective for some of the most prevalent mental health problems.

However, estimates of effect size in meta-analyses of psychological therapy studies vary according to methodological factors. For example, in a review of 115 randomized controlled trials (RCTs) investigating psychological therapies for depression, an overall mean effect size of 0.68 reduced to an effect size of 0.22 when only high-quality studies were included (Cuijpers et al., 2010). More recently, a meta-analysis found that effect sizes for psychological therapy varied according to the type of control condition: effect sizes in favor of psychological therapy were 0.89 when compared to waitlist, 0.61 relative to care-as-usual, and 0.51 versus other controls (Cuijpers et al., 2019). Cuijpers et al. (2019) also found that the effect size of psychological therapies reduced to 0.31 when publication bias was considered. Hence, the results of meta-analyses should be interpreted in light of such

methodological features. In particular, conventional meta-analyses only enable us to draw conclusions for "the average" member of a clinical population (e.g., patients meeting criteria for depression), but not for individuals with specific features.

Meta-analyses that measure treatment effects at the group-level indicate approximately equal efficacy when comparing evidence-based psychological therapies (e.g., in the case of depression [Barth et al., 2016; Cuijpers et al., 2008]), which has led some to argue that all psychological therapies work through common factors. However, there is also evidence that some psychological therapies may be more effective for some patients than others (e.g., Mulder et al., 2017). More recently, a meta-analysis of clinical trials of psychotherapies for depression found a 9% higher variance in the intervention groups compared with the control groups, which provides evidence of heterogeneity in individual treatment response across different patients (Kaiser et al., 2022). One hypothesis is that heterogeneity in treatment effects may be explained by aptitude-by-treatment interactions (Cronbach & Snow, 1977), where different psychological therapy models, components or techniques have differential effects for patients depending upon their specific characteristics. The hypothesis that some patients may respond differentially to alternative treatment options has motivated an interest in the development of various forms of personalized treatments. Multiple approaches to personalization have been documented in the psychotherapy literature, of which Treatment Matching (TM) and Individually Tailored (IT) designs are most common. TM studies are those which prospectively matched subgroups of patients to treatments based on hypothesised aptitude-by-treatment interactions (ATIs). IT studies are those which tailored treatments to individual patients (e.g., based on co-morbidities or idiosyncratic case conceptualisations).

In a narrative review of this literature, Cohen et al. (2021) have described several examples of methods used to personalize psychological treatments, ranging from methods where patients are matched to alternative interventions to methods where the treatment components (e.g., techniques) or delivery style (e.g., more or less directive) are tailored to the individual case. Cohen et al. (2021) propose a conceptual framework to describe the different forms of personalization of psychological therapy. According to this framework, there are Three Dimensions of Personalization (3DP). In the 3DP framework, the first dimension is the *timing* at which personalization decisions are made in a patient's treatment pathway, such as before, during or after treatment. The second dimension is the *level of intervention*. This refers to the level of specificity of personalization, such as the intensity of treatment, choice of modality, choice of techniques, or style of delivery. The third dimension is *structure*. This is the formality of the method of personalization, on a continuum from informal idiosyncratic personalization to using a formal statistical model.

There has been growing interest in personalized treatment in recent years. For example, Fisher et al. (2019) designed tailored treatments for individual CBT patients, based on their unique symptom profiles (*structure*) developed using pre-treatment measures (*timing*). Treatment was tailored through the selection of specific treatment modules (*level*) targeting indicated symptom domains. This resulted in a large pre-post treatment effect (g = 1.86) for reducing symptoms of anxiety and depression. In another example, Delgadillo and Gonzalez Salas Duhne (2020) used a data-driven machine learning approach (*structure*) to identify subgroups of patients with differential responses to CBT and Counselling for Depression (CfD) based on pre-treatment (*timing*) demographic and clinical information. This resulted in the development of a targeted prescription algorithm to optimally match patients to CBT or CfD (*level*). In a retrospective analysis, cases that received their model-

indicated treatment had a significantly higher rate of reliable and clinically significant improvement (62.5%) relative to cases who did not (41.7%). Numerous other examples of the development of personalized treatment selection and adaptation methods have been documented in recent years (see reviews by Checkroud et al., 2021; Cohen et al., 2021).

However, it is yet to be established whether such personalized interventions lead to improved outcomes compared to standardized interventions, particularly since most studies in this emerging area draw conclusions from secondary analyses of data from clinical trials that did not specifically test a form of personalized treatment, or from retrospective cohort datasets and uncontrolled study designs. Some prospective clinical trials have compared the efficacy of personalized treatments relative to passive or no-treatment control groups (e.g., Silfvernagel et al., 2012). However, even if supportive, such evidence does not enable clinical services to determine if personalization may be preferable to available evidence-based interventions. In order to establish the efficacy of treatment personalization, evidence form randomized controlled trials of personalized interventions compared to standard evidence-based interventions is necessary. Therefore, the present systematic literature review and meta-analysis aimed to examine whether personalized treatment is associated with improved mental health outcomes relative to passive control groups and to standardized treatments, additionally investigating the efficacy of different methods of personalization.

Methods

Transparency and Openness

Transparency and Openness Promotion (TOP) guidelines (Nosek et al., 2015) were followed. The protocol for this systematic review (including plans for the search strategy, data extraction and analysis) was registered in the Open Science Framework (OSF) database prior to conducting the literature search (MD5: e730c768b5b99d9e911984befc9aea5e).

Studies excluded at the full text screening stage are outlined in supplemental material B, with reasons for exclusion. Effect sizes reported by studies included in the meta-analysis are available in supplemental material E. Data analysis was conducted using the statistical package Meta-Analysis via Shiny (MAVIS) (Hamilton et al., 2016).

Search strategy

Table 1 displays the inclusion and exclusion criteria that guided the review. The criteria were developed following a PICOS framework, which has demonstrated greater sensitivity than the SPIDER and greater specificity than the PICO search tools (Methley et al., 2014). Key search terms (related to personalization, psychological therapy and randomized controlled trials) were combined using Boolean operators (see supplemental material A). The search was conducted in April 2022 using three databases: SCOPUS, Web of Science and PsychINFO. No restrictions were applied relating to the date of publication.

The first author screened titles, abstracts and full texts against the inclusion and exclusion criteria. Queries about the inclusion of studies were discussed and resolved by the research team. Forward and backward citation searches were conducted for each included study, and authors of included studies were contacted via email to identify further studies meeting the inclusion and exclusion criteria.

Table 1

The inclusion and exclusion criteria presented in a PICOS framework

	Inclusion Criteria	**Exclusion Criteria**
Population	Adult patients (18 years or older) accessing psychological treatment for a mental health problem	Studies where more than 50% of patients were under 18 years old

Intervention	Studies in which patients were prospectively matched to psychological treatments, or where tailored psychological treatment was examined, and where the matching or tailoring method was the primary experimental intervention	Studies that did not prospectively match patients to treatments Treatment matching only to pharmaceutical treatments Treatment matching outside of a mental health context
Comparator		
Outcome	Outcome is recorded using a validated patient-reported measure, therapist-reported measure or diagnostic interview A quantitative analysis of outcome is included	Outcome not recorded using any validated measure, therapist-reported measure or diagnostic interview No quantitative analysis of outcome is included
Study design	The study design is a randomized controlled trial	Studies which are not randomized controlled trials Articles written in languages other than English Articles which have not been peer reviewed (i.e., grey literature)

Data Extraction

Data extraction was performed by a single reviewer. The primary outcome of interest was whether personalized treatment (i.e., via treatment matching or individual tailoring) led to improved mental health outcomes versus standardized treatment. Data were extracted

relating to the effect of personalized treatment vs. standardized treatment and personalized treatment vs. control groups (a heterogenous category including online discussion group, weekly check-in, waitlist). Quantitative data derived from all primary outcome measures or diagnostic interviews were extracted at all available timepoints. In addition to statistical outcomes, data were extracted pertaining to: study design; type of personalization; country; setting; number of participants; participant age, gender and mental health conditions; interventions provided; total *N*; analysed *N*; narrative outcome.

Risk of Bias Assessment

The Cochrane risk-of-bias tool for randomized trials (RoB 2) (Sterne et al., 2019) was used. All included articles were rated by the first author, with 50% of articles ($k = 8$) selected at random to be independently assessed by a second reviewer. The first and second reviewer subsequently compared their ratings and resolved any discrepancies. Interrater reliability before discrepancies were resolved was calculated using Cohen's kappa coefficient, $k = 0.40$, indicating fair agreement (Landis & Koch, 1977).

Data Synthesis

A narrative synthesis was carried out including all eligible studies. In addition, all studies which provided sufficient statistical data (i.e., some measure of between-group effect size) were included in a random effects meta-analysis using MAVIS (Hamilton et al., 2016). Between-group (personalized vs. non-personalized) effect sizes were converted to a common metric (Cohen's *d*) to enable a meta-analysis. Where studies reported more than one primary outcome measure, a pooled within-study effect size was calculated by combining effect sizes across all measures. In addition, Q and I^2 statistics were calculated to test for heterogeneity (Huedo-Medina et al., 2006). Furthermore, three tests were used to investigate publication bias: a weighted regression model with multiplicative dispersion; a rank correlation test for

funnel plot asymmetry; and a fail-safe N calculation using Rosenthal's approach (Oswald & Plonsky, 2010). As there were fewer than 20 studies included in the meta-analysis, conducting moderator analyses was deemed inappropriate. However, subgroup analyses based on the outcome measures (depression, general distress), risk of bias (high, low), type of personalization (treatment matching, tailoring, level of personalization), and structure of personalization (informal, semi-formal, formal) were conducted to investigate potential sources of effect size heterogeneity.

Results

Study Characteristics

The results of the search and selection process are displayed in a PRISMA diagram (Figure 1). The reasons for excluding each study at the full text screening stage (k = 72) are outlined in the supplemental material B. In total, seventeen studies met the inclusion criteria. Study characteristics and study outcomes are presented in the supplemental materials D and E.

Fourteen studies compared personalized treatment to standardized treatment and six studies compared personalized treatment to control groups. Seven studies examined a TM approach to personalization and ten studies examined an IT approach to personalization. Studies using a TM approach to personalization assigned patients to treatments based upon a decision-support tool (k = 2), a statistical model (k = 1), levels of sociopathy and psychopathology (k = 1), presenting problem and patient characteristics (k = 1), responses to the Addiction Severity Index (ASI) (k = 1) and diagnosis and patient goals (k = 1). Studies which used an IT approach tailored treatment based on comorbidities (k = 3), case conceptualisations (k = 2), psychometric questionnaire responses (k = 1), individual symptoms (k = 1), pre-treatment assessment scores (k = 1), pre-treatment OCD processes

endorsed (k = 1) and a combination of pre-treatment interview and clinical impression (k = 1). Personalization was achieved by prescribing specific treatment modules (k = 7), selecting treatment intensity (k = 4), selecting treatment modality (k = 3), selecting specific cognitive-behavioral techniques (k = 2) and tailoring specific online text (k = 1).

Using the 3DP framework, in terms of the *timing* of personalization, 16 studies examined prospective treatment personalization, while one study (Lutz et al., 2021) employed a combination of prospective and live adaptive treatment personalization. With regards to *level* of personalization, 10 studies personalized treatment at the level of treatment components (techniques), four studies personalized at the level of treatment intensity (e.g., brief intervention vs. psychotherapy), and three studies personalized by treatment package (e.g., therapy modality A vs. therapy modality B). The *structure* of personalization was more difficult to determine, as this exists across a continuum from an informal (e.g., clinical intuition) method of personalization to a formal statistical model of personalization. However, broadly, six studies were deemed to use an informal method of personalization, one study was deemed to use a formal statistical model of personalization and the remaining 10 studies were deemed to use a semi-formal method of personalization (i.e., following guidelines, a decision guide or decision rule).

The total number of participants across all included studies was N = 7,617, with a sample size range of 54 – 1868 participants. The gender of participants across all studies ranged from 76% male to 73% female, and mean age ranged from 25 to 44 years. In total, 5 studies were conducted in Sweden, 3 in Australia, 3 in Germany, 2 in the United States, 1 in the Netherlands, 1 in Russia, 1 in the United Kingdom and 1 was split between Switzerland, Germany and Austria. A total of 7 studies examined psychological therapy in an online setting with the remaining 10 studies examining predominantly face-to-face psychological

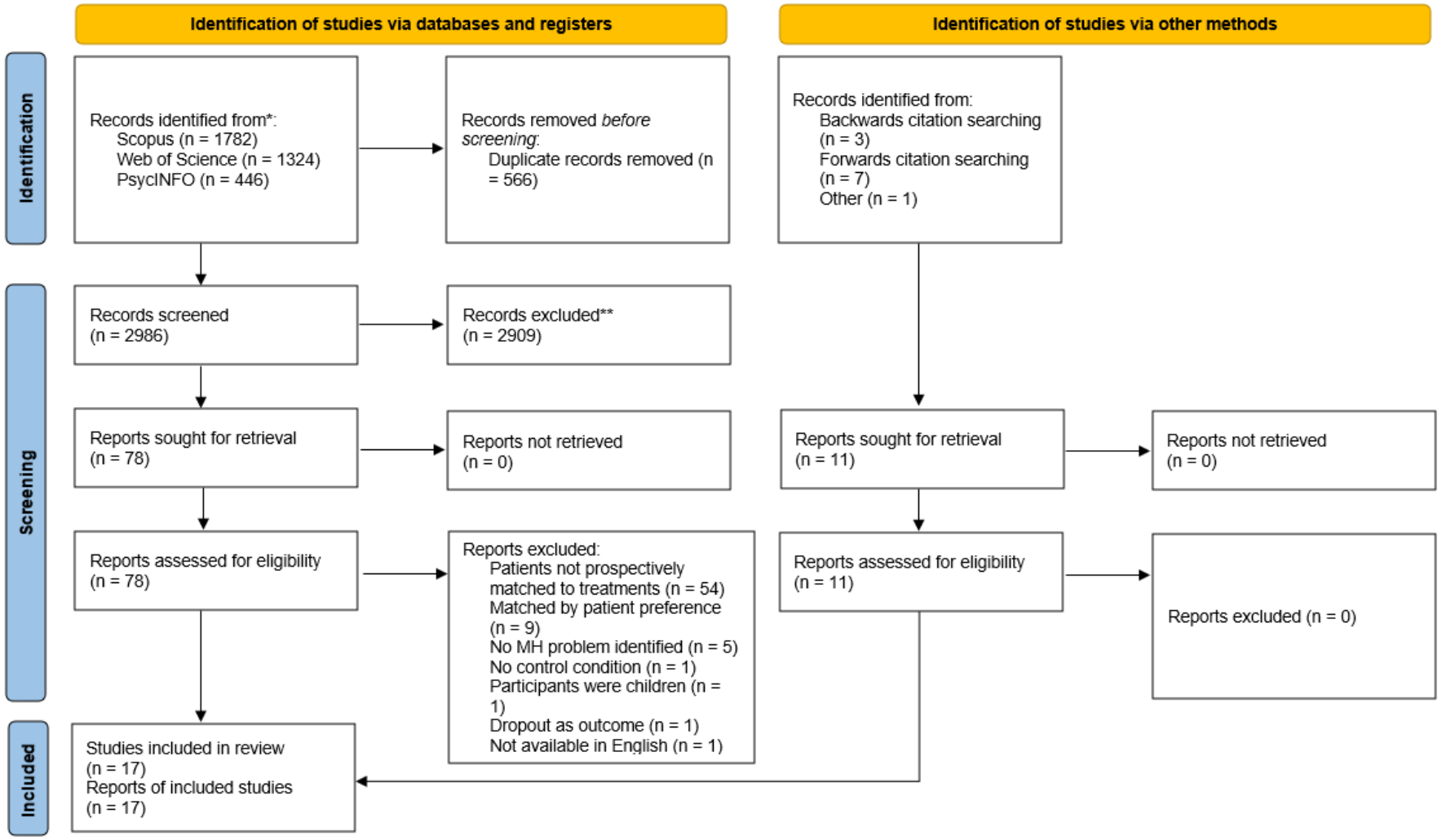
therapy, which took place in primary care settings (k = 3), university research clinics (k = 3), substance use services (k = 2), outpatient mental health services (k = 1) and psychotherapeutic inpatient settings (k = 1). The number of studies by each primary mental health condition examined were: depression (k = 6), anxiety disorders (k = 5), drug or alcohol dependence (k = 3) obsessive compulsive disorder (k = 1), any mental health problem (k = 1), psychological distress (k = 1).

The most commonly used primary outcome measures among the studies were the Beck Depression Inventory (BDI-I & BDI-II) (k = 4), the Beck Anxiety Inventory (BAI) (k = 2), the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM) (k = 2), and the Patient Health Questionnaire (PHQ-9) (k = 2). Other primary outcome measures are outlined in supplemental material F.

Risk of Bias Assessment

All studies except one were rated as having "some concerns" or a "high risk" of bias. Eight studies were rated as "high risk", eight were rated as having "some concerns" and one was rated as "low risk" overall. The most common sources of bias related to: (1) the measurement of the outcome; or (2) the selection of the reported result. Respectively, these two sources of bias were typically due to: (1) the possibility that participants may have experienced increased expectancy effects because of being informed they were offered a personalized rather than standardized intervention; and (2) many studies not pre-registering analysis plans. Supplemental material C displays the overall risk of bias ratings for all reviewed studies, including ratings for each domain of the RoB 2.

Figure 1. PRISMA Flow Diagram



Narrative Synthesis

All six studies which included a comparison of personalized treatment versus control groups (including waitlist, discussion group, weekly-check-in) found superior outcomes for personalized treatment. Eight out of 14 studies which included a comparison of personalized treatment versus standardized treatment found a superior outcome for personalized treatment, five studies found no significant differences and one study reported a superior outcome for standardized treatment (Schulte et al., 1992). Of the eight studies reporting a superior outcome for personalized treatment versus standardized treatment, three studies identified superiority of personalized treatment only for a subsample of participants and not in the overall sample (Johansson et al., 2012 [patients with higher baseline depression severity or comorbidity]; Lutz et al., 2021 [patients with a clear treatment recommendation identified by the treatment algorithm]; Watzke et al., 2010 [patients systematically assigned to psychodynamic therapy]) and one study only identified superiority on one of three primary outcome measures (Kadden et al., 2001 [fewer negative consequences of drinking, but not fewer days of abstinence or heavy drinking]). Two of the eight studies reporting a superior outcome for personalized treatment versus standardized treatment reported on follow-up measurements: one study found that the differences favoring personalization were maintained at 12-month post-treatment (Fletcher, Chondros et al., 2021) and one study found that the differences were maintained at 6-month post-treatment but became non-significant at 12-month post-treatment (Fletcher, Spittal et al., 2021).

Meta-Analysis

Nine studies (N = 5,134) provided sufficient data to be included in the primary meta-analysis comparing outcomes for personalized treatment versus standardized treatment. The mean effect size for personalized treatment versus standardized treatment was d = 0.22 (95% CI = 0.05, 0.39), p = 0.011, indicating that personalized treatment brought about significantly improved outcomes relative to standardized treatment (Figure 2). Cochran's Q test (Q[8] = 46.43, p < .001) indicated significant evidence of heterogeneity, I^2 = 87.32%. The test for funnel plot asymmetry (t[7] = 0.56, p = 0.592) and Kendall's tau (0.22, p = 0.477) indicated no significant evidence of publication bias, fail-safe N = 144 (see supplemental material G for funnel plot).

Sub-group analyses for the meta-analysis examining personalized versus standardized treatment were carried out to investigate potential sources of heterogeneity. The subgroup analysis of studies using a depression measure (k = 4; n = 2395) yielded a mean effect size of d = 0.16 (95% CI = 0.08, 0.25), p <.001, indicating personalized treatment was associated with significantly improved depression outcomes relative to standardized treatment. No significant heterogeneity was found, Q(3) = 0.13, p = 0.99, I^2 = 0%.

The subgroup analysis of studies using a distress measure (k = 3, n = 1439) yielded a mean effect size of d = 0.09 (95% CI = -0.01, 0.20), p = 0.075, indicating no significant difference in distress outcomes between personalized and standardized treatment. There was no significant heterogeneity, Q(2) = 0.58, p = 0.748, I^2 = 0%.

The subgroup analysis of studies rated as having a high risk of bias (k = 2, n = 452) yielded an effect size of d = 0.57 (95% CI = -0.03, 1.18), p = 0.063. This is a higher effect size than the primary meta-analysis, but not statistically significant. There was significant heterogeneity, Q(1) = 5.78, p = 0.016, I^2 = 82.7%.

The subgroup analysis of studies rated as having low risk of bias or some concerns (k = 7, n = 4682) yielded an effect size of d = 0.14 (95% CI = 0.08, 0.20), p <0.001. This effect size is smaller than that found in the primary meta-analysis, but remains statistically significant. There was no significant heterogeneity, $Q(6) = 6.09$, $p = 0.413$, $I^2 = 0\%$.

The subgroup analysis of studies which used a TM approach to personalization (k = 5, n = 4,034) yielded an effect size of d = 0.13 (95% CI = 0.06 to 0.19), $p < 0.001$, indicating that TM was associated with statistically significant superior outcomes to standardized (unmatched) treatment. There was no significant heterogeneity, $Q(4) = 2.94$, $p = 0.568$, $I^2 = 0\%$.

The subgroup analysis of studies which used an IT approach to personalization (k = 4, n = 1100) yielded an effect size of d = 0.37 (95% CI = 0.01, 0.74), $p = 0.045$, indicating that IT was associated with statistically significant superior outcomes to standardized (untailored) treatment. Significant heterogeneity was found, $Q(3) = 22.65$, $p < 0.001$, $I^2 = 85.55\%$.

Subgroup analyses were conducted for the three different levels of personalization of the 3DP framework: component-level matching (where patients are matched to particular treatment components or modules), intensity-level matching (where patients are matched to different intensities of treatment) and package-level matching (where patients are matched to a particular treatment modality).

The subgroup analysis for component-level matching yielded the same result as that for IT studies reported directly above, as this evaluated the same results from the same studies (d = 0.37 [95% CI = 0.01, 0.74], $p = 0.045$).

The subgroup analysis of studies which examined intensity-level matching (k = 3, n = 3,359) yielded an effect size of d = 0.14 (95% CI = 0.07, 0.21), $p < 0.001$, indicating that intensity-level matching was associated with significantly superior outcomes relative to

standardized (unmatched) treatment. There was no significant heterogeneity, $Q(2) = 1.03$, $p = 0.598$, $I^2 = 0\%$.

The subgroup analysis of studies using package-level matching ($k = 2$, $n = 675$) yielded an effect size of $d = 0.06$ (95% CI = -0.10, 0.21), $p = 0.489$, indicating no significant difference between package-level matching and standardized (unmatched) treatment. There was no significant heterogeneity, $Q(1) = 0.99$, $p = 0.320$, $I^2 = 0\%$.

Subgroup analyses were conducted for different structures of personalization based on the 3DP framework. Studies which examined a semi-formal structure were grouped with the study which examined a statistical (formal) structure, as only one study evaluated the latter structure of personalization (Delgadillo et al., 2022).

The results of the subgroup analysis for studies examining an artisanal (informal) structure were identical to that of the package-level subgroup analysis above, as it included the same studies ($d = 0.06$ [95% CI = -0.10, 0.21], $p = 0.489$).

The subgroup analysis of studies examining a formal or semi-formal structure ($k = 7$, $n = 4459$) yielded an effect size of $d = 0.26$ (95% CI = 0.06, 0.47), $p = 0.013$, indicating a significant advantage in favour of formal or semi-formal personalization vs standardized (unmatched) treatment. There was significant heterogeneity for this result, $Q(6) = 42.07$, $p < 0.001$, $I^2 = 90.07\%$.

Six studies ($N = 426$) provided sufficient data for inclusion in the secondary meta-analysis comparing outcomes for personalized treatment versus control groups (either waitlist or [k=3], support group [k=1], discussion group [k=1] or weekly check-ins [k=1]). The mean effect size for personalized treatment versus control groups was $d = 0.89$ (95% CIs = 0.69, 1.09), $p < 0.001$, suggesting that personalized treatment was associated with significantly superior outcomes relative to control groups (Figure 3). Cochran's Q test ($Q[5] = 3.84$, $p =$

0.573) suggested no significant heterogeneity, $I^2 = 0\%$. The test for funnel plot asymmetry ($t[4] = 1.15$, $p = 0.315$) and Kendall's tau (0.20, $p = 0.719$) indicated no significant evidence of publication bias, fail-safe $N = 166$ (see supplemental material H for funnel plot).

A sensitivity analysis for the secondary meta-analysis was conducted, in which effect sizes derived from personalized treatment versus active control groups with minimally intensive interventions (support group, discussion group, weekly check-in) were removed, leaving only the effect sizes generated from personalized treatment versus waitlist control groups ($k = 3$, $n = 203$). This resulted in a slightly larger effect size for personalized treatment versus control groups (d = 1.01 [95% CI = 0.67, 1.34], p <0.001), with no significant heterogeneity ($Q[2] = 2.58$, $p = 0.277$, $I^2 = 20.96\%$).

Figure 2. Random-effects meta-analysis forest plot: Outcomes for personalized treatment versus standardized treatment

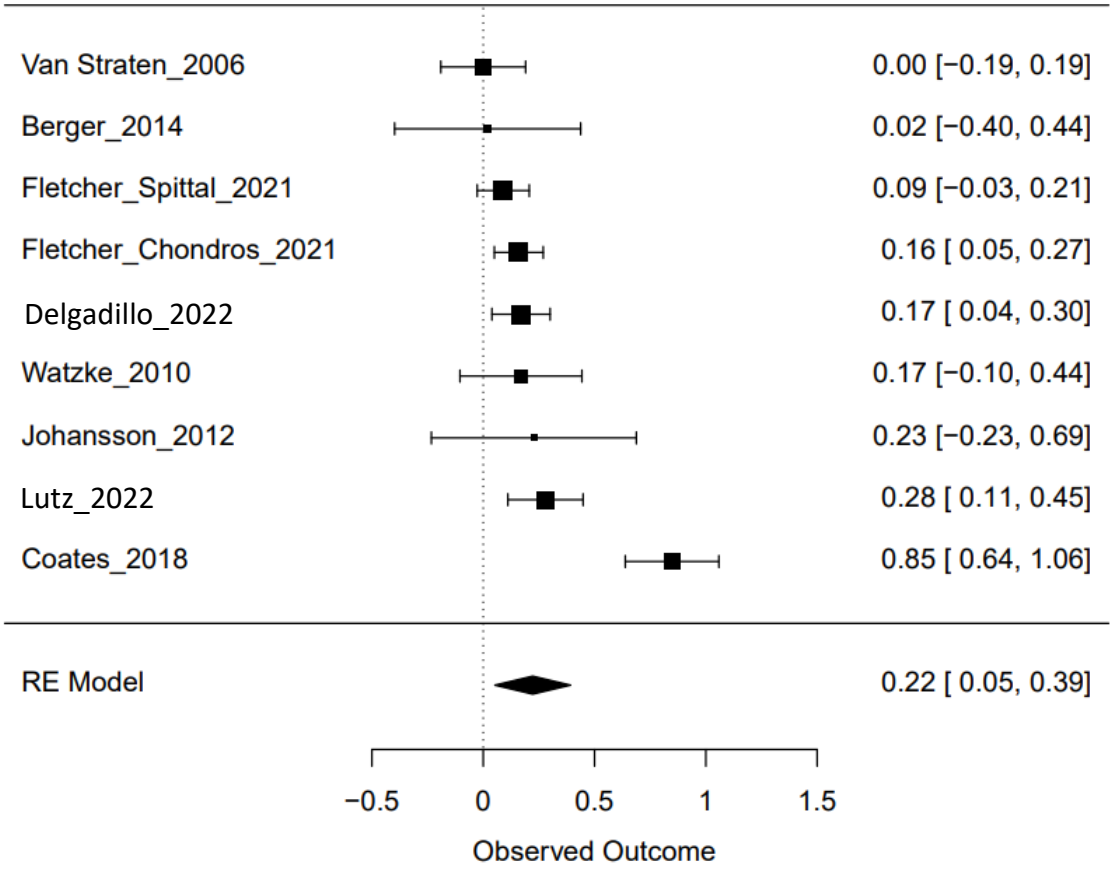
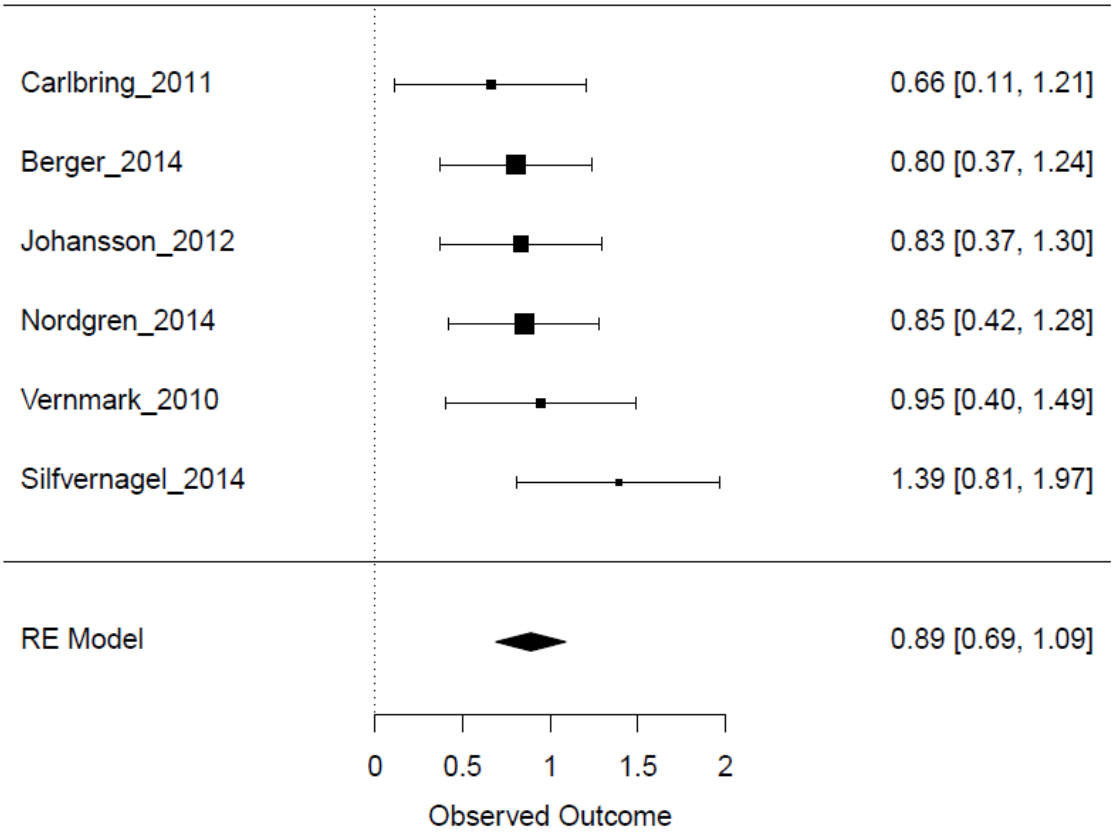


Figure 3. Random-effects meta-analysis forest plot: Outcomes for personalized treatment versus control groups



Discussion

Summary of Results

This systematic review and meta-analysis found that personalized psychological treatment was associated with improved outcomes relative to standard evidence-based psychological treatment. There was no evidence of publication bias for this result, with the failsafe *N* calculation indicating that 144 studies with null results would be required to overturn this finding. The effect size for personalized treatment versus standardized treatment was small ($d = 0.22$) but statistically significant, and this effect remained statistically significant in studies with low risk of bias or with some concerns, albeit with a reduced effect

size (d = 0.14) relative to studies with high risk of bias (d = 0.57). Personalized treatment had superior outcomes relative to standardized treatment on measures of depression (d = 0.16), but not on measures of distress (d = 0.09).

In terms of the type of personalization, both TM (d = 0.13) and IT approaches (d = 0.37) to personalization were associated with superior outcomes relative to standardized (unmatched or untailored) treatment. In terms of the level of personalization, intensity-level matching was associated with improved outcomes relative to standardized treatment (d = 0.14), as was component-level matching (d = 0.37). The magnitude of the latter comparison indicates that component-level matching is a particularly effective method of personalization. However, package-level-matching (d = 0.06) was not associated with improved outcomes relative to standardized treatment. With regards to the structure of personalization, formal or semi-formal structures were associated with better outcomes than standardized treatment (d = 0.26), whereas artisanal (informal) structures were not (d = 0.06). It is important to point out that the latter two subgroup analyses included the same studies, so it is not clear if package-level matching is no more effective than standardized treatment, or if these studies were undermined by an informal method of personalization. Future research is needed to determine if a data-driven (formal) method of package-level matching may be more effective than standardized treatment.

In the meta-analysis comparing personalized treatment to control groups, personalized treatment had a large advantage relative to control groups (d = 0.89). We note that the magnitude of this effect size compared to control groups is larger than the effect size of approximately 0.70 that is typical for evidence-based (standardized) psychotherapies (e.g., see Cuijpers, 2017). There was no evidence of publication bias for this result, with the

failsafe *N* calculation indicating that 166 studies with null results would be required to overturn it.

Strengths and Limitations

This is the first systematic review and meta-analysis investigating the efficacy of personalized psychological interventions tested in prospective and randomized controlled trials. Strengths of the review include the pre-registration of the study protocol, a systematic search conducted across multiple databases, the use of forward and backwards citation searches, a risk of bias assessment with reliability checks, and a meta-analysis of quantitative outcomes.

A limitation of the review was the relatively small number of studies available for the secondary meta-analysis assessing outcomes for personalized treatment versus different types of control groups (waitlist, support group, discussion group, weekly check-in). This meant that "minimally intensive" active control groups (support group, discussion group, weekly-check-in) were combined with passive control groups (waitlist) into one heterogenous category to ensure sufficient studies were available to conduct the secondary meta-analysis. A sensitivity analysis indicated that removing the minimally intensive active control groups from this calculation resulted in a slightly larger effect size (d = 0.89 vs. heterogenous control groups, d = 1.01 vs waitlist). This is consistent with research indicating that waitlist control conditions may be associated with a nocebo effect, leading to inflated effect sizes for intervention conditions (Mohr et al., 2014). Therefore, the primary meta-analysis which examined outcomes for personalized treatment versus standardized treatment provides the most relevant and important results.

Additionally, the small number of studies available meant that it was not possible to conduct subgroup analyses for measures of anxiety or different follow-up durations. As a

consequence of the relatively few number of studies, the subgroup analysis for studies with a low risk of bias were grouped with studies rated as having some concerns, and studies which examined formal and semi-formal structures were also grouped. Although other subgroup analyses were conducted to investigate possible sources of heterogeneity, the relatively small number of studies identified by the review meant that it was not possible to conduct more robust moderator analyses.

Furthermore, while the use of reliability checks for the risk of bias assessment was a strength, the inter-rater reliability was only fair ($k = 0.40$). As the primary risk of bias ratings may be somewhat unreliable, the results of subgroup analyses investigating high and low risk of bias studies should be interpreted with caution. Additionally, as only one author conducted the initial selection of the studies, this process could have been prone to bias. However, the selection was performed according to pre-determined inclusion and exclusion criteria, ad any queries about the inclusion of studies were discussed with the wider research team. Other limitations of the review were the exclusion of grey literature and studies not written in the English language, and study selection and data extraction performed by a single reviewer.

Clinical Implications

The results of the meta-analysis indicate that personalized treatment is associated with significantly improved outcomes relative to standardized treatment. While the effect size for this difference was small by conventional standards ($d = 0.22$), multiplying this difference over a large population of patients who engage in psychological therapy would result in a substantial number of patients experiencing improved mental health outcomes. This effect size equates to an approximate number-needed-to-treat of NNT = 8.5. By this logic, if personalized psychological care were implemented, approximately 1 out of 8 (12.5%) patients would have a better outcome by comparison to standardized interventions. Even

taking a highly conservative and modest effect size of $d = 0.14$ (for intensity-level matching), this equates to a NNT = 12.7 (7.9%), which for a population of 1000 patients would mean that 127 patients would have better outcomes by implementing personalized interventions.

Providing a formal or semi-formal structure of personalized treatment to improve depression outcomes may be particularly appropriate given: (1) the results of the subgroup analysis indicating significantly improved depression outcomes for personalized versus standardized treatment; (2) the results of the subgroup analyses indicating that formal or semi-formal structures of personalization are associated with improved outcomes, whereas informal structures are not; and (3) evidence of approximately equivalent efficacy across evidence-based psychological therapies for depression (Barth et al., 2016).

Further Research

A common source of bias in most studies included in the review was that patients self-reporting their outcome were aware of which intervention they had received, leading to the potential for increased expectancy effects upon being informed they were receiving a "personalized", "tailored" or "matched" treatment rather than a "standardized" treatment or "usual care". Some studies using a TM approach to personalization addressed this by randomly assigning therapists to matched or unmatched treatment, with patients blinded to the allocation of their therapist (and therefore unaware of whether they were engaged in a matched or unmatched treatment). Another common source of bias originated from studies not pre-registering their analysis plans. Further research could therefore reduce the risk of bias by blinding patients to their allocated treatment and pre-registering analysis plans, such as in the double-blind design applied by Delgadillo et al. (2022).

In relation to the 3DP framework, while studies investigated various levels of personalization, the review did not identify any studies which examined personalization at the

level of the format of delivery (e.g., group vs. face-to-face) or interactional style (e.g., directive vs. non-directive). In terms of the structure of personalization, only one study investigated a statistical model of personalization (Delgadillo et al., 2022). Additionally, only one study investigated personalization during treatment (Lutz et al., 2022). Therefore, further research could investigate these relative gaps in the literature to assess whether these other forms of personalization could lead to improved outcomes relative to standardized treatment.

Conclusions

Personalized psychological treatments are associated with improved outcomes relative to standardized treatments, and this seems to be particularly relevant for the treatment of depression. This evidence indicates that adopting personalized and precision mental health methods is a promising avenue to enhance the efficacy of psychological care.

References*

Barkham, M., & Lambert, M. J. (2021). The efficacy and effectiveness of psychological therapies. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.). *Bergin and Garfield's handbook of psychotherapy and behavior change,* 7th ed. (pp. 135–189). Wiley.

Barth J, Munder T, Gerger H, Nüesch, E., Trelle, S., Znoj, H., Jüni, P., & Cuijpers, P. (2016). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: A network meta-analysis. *PLoS Medicine, 10*(5) e1001454. https://doi.org/10.1176/appi.focus.140201

*Berger, T., Boettcher, J., & Caspar, F. (2014). Internet-based guided self-help for several anxiety disorders: A randomized controlled trial comparing a tailored with a standardized disorder-specific approach. *Psychotherapy*, *51*(2), 207. https://psycnet.apa.org/doi/10.1037/a0032527

Bloom, D. E., Chisholm, D., Jané-Llopis, E., Prettner, K., Stein, A., & Feigl, A. (2011). From burden to "best buys": Reducing the economic impact of non-communicable diseases. *World Health Organisation, Geneva, Switzerland*.

Calvert, S. J., Beutler, L. E., & Crago, M. (1988). Psychotherapy outcome as a function of therapist–patient matching on selected variables. *Journal of Social and Clinical Psychology*, *6*(1), 104–117. https://psycnet.apa.org/doi/10.1521/jscp.1988.6.1.104

*Carlbring, P., Maurin, L., Törngren, C., Linna, E., Eriksson, T., Sparthan, E., & Andersson, G. (2011). Individually-tailored, Internet-based treatment for anxiety disorders: A randomized controlled trial. *Behaviour Research and Therapy*, *49*(1), 18–24. https://doi.org/10.1016/j.brat.2010.10.002

* Note. Asterisks (*) denote papers included in the review.

Carpenter, J. K., Andrews, L. A., Witcraft, S. M., Powers, M. B., Smits, J., & Hofmann, S. G. (2018). Cognitive behavioral therapy for anxiety and related disorders: A meta-analysis of randomized placebo-controlled trials. *Depression and Anxiety, 35*(6), 502–514. https://doi.org/10.1002/da.22728

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

*Coates, J. M., Gullo, M. J., Feeney, G. F., Young, R. M., & Connor, J. P. (2018). A randomized trial of personalized cognitive-behavior therapy for alcohol use disorder in a public health clinic. *Frontiers in Psychiatry*, *9*, 297. https://dx.doi.org/10.3389%2Ffpsyt.2018.00297

Cohen, Z.D., Delgadillo, J., DeRubeis, R.J. (2021). Personalized treatment approaches. In M. Barkham, W. Lutz, L.G. Castonguay (Eds.), *Bergin and Garfield's handbook of psychotherapy and behavior change*, 7th ed. (pp. 667–700). Wiley.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.

Cuijpers, P. (2017). Four decades of outcome research on psychotherapies for adult depression: An overview of a series of meta-analyses. *Canadian Psychology/Psychologie Canadienne, 58*(1), 7–19. https://psycnet.apa.org/doi/10.1037/cap0000096

Cuijpers, P., Noma, H., Karyotaki, E., Vinkers, C. H., Cipriani, A., & Furukawa, T. A. (2020). A network meta-analysis of the effects of psychotherapies, pharmacotherapies

and their combination in the treatment of adult depression. *World Psychiatry*, *19*(1), 92–107. https://doi.org/10.1002/wps.20701

Cuijpers, P., Quero, S., Papola, D., Cristea, I. A., & Karyotaki, E. (2019). Care-as-usual control groups across different settings in randomized trials on psychotherapy for adult depression: a meta-analysis. *Psychological Medicine*, 1–11. https://doi.org/10.1017/s0033291719003581

Cuijpers, P., Van Straten, A., Andersson, G., & Van Oppen, P. (2008). Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology*, *76*(6), 909–922. https://doi.org/10.1037/a0013075

Cuijpers, P., Van Straten, A. A., Bohlmeijer, E. E., Hollon, S. D., & Andersson, G. G. (2010). The effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychological Medicine*, *40*(2), 211–223. https://doi.org/10.1017/s0033291709006114

*Delgadillo, J., Ali, S., Fleck, K., Agnew, C., Southgate, A., Parkhouse, L., Cohen, Z., DeRubeis, R., & Barkham, M. (2022). Stratified care vs stepped care for depression: a cluster randomized clinical trial. *JAMA Psychiatry*, *79*(2), 101–108. https://doi.org/10.1001/jamapsychiatry.2021.3539

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. https://psycnet.apa.org/doi/10.1037/ccp0000476

Fisher, A. J., Bosley, H. G., Fernandez, K. C., Reeves, J. W., Soyster, P. D., Diamond, A. E., & Barkin, J. (2019). Open trial of a personalized modular treatment for mood and

anxiety. *Behaviour Research and Therapy*, *116*, 69–79. https://doi.org/10.1016/j.brat.2019.01.010

*Fletcher, S., Chondros, P., Densley, K., Murray, E., Dowrick, C., Coe, A., Hegarty, K., Davidson, S., Wachtler, C., Mihalopoulos, C., Lee, Y., Chatterton, M., Palmer, V., & Gunn, J. (2021). Matching depression management to severity prognosis in primary care: results of the Target-D randomised controlled trial. *British Journal of General Practice*, *71*(703), e85-e94. https://doi.org/10.3399/BJGP.2020.0783

*Fletcher, S., Spittal, M. J., Chondros, P., Palmer, V. J., Chatterton, M. L., Densley, K., Potiriadis, M., Harris, H., Bassilios, B., Burgess, P., Mihalopoulos, C., Pirkis, J., & Gunn, J. (2021). Clinical efficacy of a Decision Support Tool (Link-me) to guide intensity of mental health care in primary practice: a pragmatic stratified randomised controlled trial. *The Lancet Psychiatry*, *8*(3), 202–214. https://doi.org/10.1016/s2215-0366(20)30517-4

Hamilton, W.K., Aydin, B., & Mizumoto, A. (2016). *MAVIS: Meta analysis via Shiny* [Online]. Retrieved January 10, 2022, from http://kylehamilton.net/shiny/MAVIS/

Huedo-Medina, T. B., Sánchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I² index?. *Psychological methods*, *11*(2), 193. https://doi.org/10.1037/1082-989x.11.2.193

James, S., Abate, D., Abate, K., Abay, S., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R., Abebe, Z., Abera, S., Abil, O., Abraha, H., Abu-Raddad, L., Abu-Rmeileh, N., Accrombessi, M., ... & Murray, C. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study

2017. *The Lancet*, *392*(10159), 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7

*Johansson, R., Sjöberg, E., Sjögren, M., Johnsson, E., Carlbring, P., Andersson, T., Rousseau, A., & Andersson, G. (2012). Tailored vs. standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: a randomized controlled trial. *PloS One*, *7*(5), e36905. https://doi.org/10.1371/journal.pone.0036905

*Kadden, R. M., Litt, M. D., Cooney, N. L., Kabela, E., & Getter, H. (2001). Prospective matching of alcoholic clients to cognitive-behavioral or interactional group therapy. *Journal of Studies on Alcohol*, *62*(3), 359–369. https://doi.org/10.15288/jsa.2001.62.359

Kaiser, T., Volkmann, C., Volkmann, A., Karyotaki, E., Cuijpers, P., & Brakemeier, E.-L. (2022). Heterogeneity of treatment effects in trials on psychotherapy of depression. *Clinical Psychology: Science and Practice, 29*(3), 294–303. https://doi.org/10.1037/cps0000079

Kazdin, A. E. (1986). Comparative outcome studies of psychotherapy: methodological issues and strategies. *Journal of Consulting and Clinical Psychology*, *54*(1), 95. https://psycnet.apa.org/doi/10.1037/0022-006X.54.1.95

Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual Review of Public Health*, *34*, 119–138. https://doi.org/10.1146/annurev-publhealth-031912-114409

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363–374. https://doi.org/10.2307/2529310

*Lutz, W., Deisenhofer, A.-K., Rubel, J., Bennemann, B., Giesemann, J., Poster, K., & Schwartz, B. (2022). Prospective evaluation of a clinical decision support system in psychological therapy. *Journal of Consulting and Clinical Psychology, 90*(1), 90–106. https://doi.org/10.1037/ccp0000642

*McLellan, A. T., Grissom, G. R., Zanis, D., Randall, M., Brill, P., & O'Brien, C. P. (1997). Problem-service 'matching' in addiction treatment: A prospective study in 4 programs. *Archives of General Psychiatry*, *54*(8), 730–735. https://doi.org/10.1001/archpsyc.1997.01830200062008

Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Services Research*, *14*(1), 1–10. https://doi.org/10.1186/s12913-014-0579-0

Mohr, D. C., Ho, J., Hart, T. L., Baron, K. G., Berendsen, M., Beckner, V., Cai, X., Cuijpers, P., Spring, B., Kinsinger, S. W., Schroder, K. E., & Duffecy, J. (2014). Control condition design and implementation features in controlled trials: a meta-analysis of trials evaluating psychotherapy for depression. *Translational Behavioral Medicine*, *4*(4), 407-423. https://doi.org/10.1007/s13142-014-0262-3

*Moritz, S., Stepulovs, O., Schröder, J., Hottenrott, B., Meyer, B., & Hauschildt, M. (2016). Is the whole less than the sum of its parts? Full versus individually adapted metacognitive self-help for obsessive-compulsive disorder: A randomized controlled trial. *Journal of Obsessive-Compulsive and Related Disorders*, *9*, 107–115. https://doi.org/10.1016/j.jocrd.2016.04.001

Mulder, R., Murray, G., & Rucklidge, J. (2017). Common versus specific factors in psychotherapy: Opening the black box. *The Lancet Psychiatry*, *4*(12), 953–962. https://doi.org/10.1016/S2215-0366(17)30100-1

NHS Digital (2021). *Psychological Therapies, Annual report on the use of IAPT services, 2020-21.* https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-annual-reports-on-the use-of-iapt-services/annual-report-2020-21

*Nordgren, L. B., Hedman, E., Etienne, J., Bodin, J., Kadowaki, Å., Eriksson, S., Lindkvist, E., Andersson, G., & Carlbring, P. (2014). Effectiveness and cost-effectiveness of individually tailored Internet-delivered cognitive behavior therapy for anxiety disorders in a primary care population: a randomized controlled trial. *Behaviour Research and Therapy*, *59*, 1–11. https://doi.org/10.1016/j.brat.2014.05.007

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., ... & Yarkoni. (2015). Promoting an open research culture. *Science*, *348* (6242), 1422-1425. https://doi.org/10.1126/science.aab2374

Oswald, F. L., & Plonsky, L. (2010). Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics*, *30*, 85–110. https://doi.org/10.1017/S0267190510000115

Project MATCH Research Group. (1998). Matching patients with alcohol disorders to treatments: Clinical implications from Project MATCH. *Journal of Mental Health, 7*(6), 589–602. https://doi.org/10.1080/09638239817743

*Schulte, D., Künzel, R., Pepping, G., & Schulte-Bahrenberg, T. (1992). Tailor-made versus standardized therapy of phobic patients. *Advances in Behaviour Research and Therapy, 14*(2), 67–92. https://doi.org/10.1016/0146-6402(92)90001-5

*Silfvernagel, K., Carlbring, P., Kabo, J., Edström, S., Eriksson, J., Månson, L., & Andersson, G. (2012). Individually tailored internet-based treatment for young adults and adults with panic attacks: randomized controlled trial. *Journal of Medical Internet Research*, *14*(3), e65. https://dx.doi.org/10.2196%2Fjmir.1853

Sterne, J. A., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Jungqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., ... & Higgins, J. P. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ,* 366. https://doi.org/10.1136/bmj.l4898

*Van Straten, A., Tiemens, B., Hakkaart, L., Nolen, W. A., & Donker, M. C. H. (2006). Stepped care vs. matched care for mood and anxiety disorders: a randomized trial in routine practice. *Acta Psychiatrica Scandinavica*, *113*(6), 468–476. https://doi.org/10.1111/j.1600-0447.2005.00731.x

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

*Watzke, B., Rüddel, H., Jürgensen, R., Koch, U., Kriston, L., Grothgar, B., & Schulz, H. (2010). Effectiveness of systematic treatment selection for psychodynamic and cognitive–behavioural therapy: Randomised controlled trial in routine mental

healthcare. *The British Journal of Psychiatry*, *197*(2), 96–105.

https://psycnet.apa.org/doi/10.1192/bjp.bp.109.072835

Supplemental Materials

Supplemental Material A: Search Terms

1. (match* OR selection OR targeted OR personalised OR personalized)
2. AND (psychological therap* OR psychotherap*)
3. AND (randomised control*" OR "randomized control*" OR RCT)

Supplemental Material B: Studies Excluded at Full Text Screening Stage

Author	Year	Title	DOI	Reason for Exclusion
Aardoom et al.	2017	Moderators of change in an internet-based intervention for eating disorders with different levels of therapist support: What works for whom?	10.1016/j.brat.2016.11.012	Did not prospectively match patients to treatments
Allen et al.	1997	Project MATCH secondary a priori hypotheses	10.1111/j.1360-0443.1997.tb02889.x	Did not prospectively match patients to treatments
Anderson et al.	2020	Predictors and moderators of treatment outcome in a randomized clinical trial for binge-eating disorder	10.1037/ccp0000503	Did not prospectively match patients to treatments
Arch & Ayers.	2013	Which treatment worked better for whom? Moderators of group cognitive behavioral therapy versus adapted mindfulness based stress reduction for anxiety disorders	10.1016/j.brat.2013.04.004	Did not prospectively match patients to treatments
Avants et al.	1998	When is less treatment better? The role of social anxiety in matching methadone patients to psychosocial treatments	10.1037/0022-006X.66.6.924	Did not prospectively match patients to treatments
Bagby et al.	2008	Personality and differential treatment response in major depression: A randomized controlled trial comparing cognitive-behavioural therapy and pharmacotherapy	10.1177/070674370805300605	Did not prospectively match patients to treatments
Beaucham et al..	2013	Do personality traits matter when choosing a group therapy for early psychosis?	10.1111/j.2044-8341.2011.02052.x	Did not prospectively match patients to treatments
Beutler.	2003	A comparison of the Dodo, EST, and ATI factors among comorbid stimulant-dependent, depressed patients	10.1002/cpp.354	Did not prospectively match patients to treatments
Brown et al.	2002	Matching substance abuse aftercare treatments to client characteristics	10.1016/S0306-4603(01)00195-2	Did not prospectively match patients to treatments

Bulmash et al.	2009	Personality, stressful life events, and treatment response in major depression	10.1037/a0017149	Did not prospectively match patients to treatments
Carter et al.	2018	Patient predictors of response to cognitive behaviour therapy and schema therapy for depression	10.1177/0004867417750756	Did not prospectively match patients to treatments
Chen et al.	2014	5-HTTLPR moderates naltrexone and psychosocial treatment responses in heavy drinking men who have sex with men	10.1111/acer.12492	Did not prospectively match patients to treatments
Cohen et al..	2020	A demonstration of a multi-method variable selection approach for treatment selection: Recommending cognitive–behavioral versus psychodynamic therapy for mild to moderate adult depression	10.1080/10503307.2018.1563312	Did not prospectively match patients to treatments
D'Antonio et al.	2013	Depression and traumatic brain injury: symptom profiles of patients treated with cognitive-behavioral therapy or supportive psychotherapy	10.2217/npy.13.75	Did not prospectively match patients to treatments
Delgadillo et al.	2017	Case complexity as a guide for psychological treatment selection	10.1037/ccp0000231	Did not prospectively match patients to treatments
Dew et al.	2001	Initial recovery patterns may predict which maintenance therapies for depression will keep older adults well	10.1016/S0165-0327(00)00280-9	Did not prospectively match patients to treatments
Donker et al.	2013	Predictors and moderators of response to internet-delivered interpersonal psychotherapy and cognitive behavior therapy for depression	10.1016/j.jad.2013.06.020	Did not prospectively match patients to treatments
Driessen et al.	2016	Differential efficacy of cognitive behavioral therapy and psychodynamic therapy for major depression: A study of prescriptive factors	10.1017/S0033291715001853	Did not prospectively match patients to treatments
Dunlop et al.	2015	Preliminary findings supporting insula metabolic activity as a predictor of outcome to psychotherapy and medication treatments for depression	10.1176/appi.neuropsych.14030048	Did not prospectively match patients to treatments
Eskildsen et al.	2020	Personalized psychotherapy for outpatients with major depression and anxiety disorders: transdiagnostic versus diagnosis-specific group cognitive behavioural therapy	10.1007/s10608-020-10116-1	Did not prospectively match patients to treatments

Friedl et al.	2020	Using the personalized advantage index for individual treatment allocation to cognitive behavioral therapy (CBT) or a CBT with integrated exposure and emotion-focused elements (CBT-EE)	10.1080/10503307.2019.1664782	Did not prospectively match patients to treatments
Gomez Penedo et al.	2017	Markers for context-responsiveness: Client baseline interpersonal problems moderate the efficacy of two psychotherapies for generalized anxiety disorder	10.1037/ccp0000233	Did not prospectively match patients to treatments
Heather et al.	2008	UK Alcohol Treatment Trial: Client-treatment matching effects	10.1111/j.1360-0443.2007.02060.x	Did not prospectively match patients to treatments
Huibers et al.	2015	Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the personalized advantage index approach	10.1371/journal.pone.0140771	Did not prospectively match patients to treatments
Joutsenniemi et al.	2012	Prediction of the outcome of short- and long-term psychotherapy based on socio-demographic factors	10.1016/j.jad.2012.03.027	Did not prospectively match patients to treatments
Kikkert et al.	2016	The role of avoidant and obsessive-compulsive personality disorder traits in matching patients with major depression to cognitive behavioral and psychodynamic therapy: A replication study	10.1016/j.jad.2016.08.017	Did not prospectively match patients to treatments
Kim et al.	2019	Initial severity-dependent longitudinal model with application to a randomized controlled trial of women with depression	10.1002/sim.8072	Did not prospectively match patients to treatments
Kuerbis et al.	2018	Exploration of treatment matching of problem drinker characteristics to motivational interviewing and non-directive client-centered psychotherapy	10.1016/j.jsat.2017.12.002	Did not prospectively match patients to treatments
Le Grange et al.	2014	Predictors and moderators of outcome for severe and enduring anorexia nervosa	10.1016/j.brat.2014.03.006	Did not prospectively match patients to treatments
Le Grange et al.	2012	Moderators and mediators of remission in family-based treatment and adolescent focused therapy for anorexia nervosa	10.1016/j.brat.2011.11.003	Did not prospectively match patients to treatments

Lloyd et al.	2014	Comorbidity in the prediction of cognitive processing therapy treatment outcomes for combat-related posttraumatic stress disorder	10.1016/j.janxdis.2013.12.002	Did not prospectively match patients to treatments
Lorenzo-Luaces et al.	2017	A prognostic index (PI) as a moderator of outcomes in the treatment of depression: A proof of concept combining multiple variables to inform risk-stratified stepped care models	10.1016/j.jad.2017.02.010	Did not prospectively match patients to treatments
Maude-Griffin et al.	1998	Superior efficacy of cognitive-behavioral therapy for urban crack cocaine abusers: Main and matching effects	10.1037/0022-006X.66.5.832	Did not prospectively match patients to treatments
McGrath et al.	2013	Toward a neuroimaging treatment selection biomarker for major depressive disorder	10.1001/jamapsychiatry.2013.143	Did not prospectively match patients to treatments
Moggia et al.	2020	Patterns of change and their relationship to outcome and follow-up in group and individual psychotherapy for depression	10.1037/ccp0000562	Did not prospectively match patients to treatments
Najafzadeh et al.	2017	Economic evaluation of implementing a novel pharmacogenomic test (IDgenetix®) to guide treatment of patients with depression and/or anxiety	10.1007/s40273-017-0587-0	Did not prospectively match patients to treatments
Newman et al.	2017	Interpersonal problems predict differential response to cognitive versus behavioral treatment in a randomized controlled trial	10.1016/j.beth.2016.05.005	Did not prospectively match patients to treatments
Newman et al.	2019	Time-varying moderation of treatment outcomes by illness duration and comorbid depression in generalized anxiety disorder	10.1037/ccp0000385	Did not prospectively match patients to treatments
Norr et al.	2018	Virtual reality exposure versus prolonged exposure for PTSD: Which treatment for whom?	10.1002/da.22751	Did not prospectively match patients to treatments
O'Keeffe et al.	2018	Predicting dropout in adolescents receiving therapy for depression	10.1080/10503307.2017.1393576	Did not prospectively match patients to treatments
Peters et al.	2016	Medical burden, body mass index and the outcome of psychosocial interventions for bipolar depression	10.1080/10503307.2017.1393576	Did not prospectively match patients to treatments
Piper et al.	1999	Prediction of dropping out in time-limited, interpretive individual psychotherapy	10.1037/h0087787	Did not prospectively match patients to treatments

Presnell et al.	2012	Therapist and client race/ethnicity match: An examination of treatment outcome and process with rural older adults in the deep south	10.1080/10503307.2012.673022	Did not prospectively match patients to treatments
Rychtarik et al.	2000	Treatment settings for persons with alcoholism: Evidence for matching clients to inpatient versus outpatient care	10.1037/0022-006X.68.2.277	Did not prospectively match patients to treatments
Sahin et al.	2018	Clinical severity as a moderator of outcome in psychodynamic and dialectical behavior therapies for borderline personality disorder	10.1037/per0000276	Did not prospectively match patients to treatments
Serbanescu et al.	2020	Combining baseline characteristics to disentangle response differences to disorder-specific versus supportive psychotherapy in patients with persistent depressive disorder	10.1016/j.brat.2019.103512	Did not prospectively match patients to treatments
Sundquist et al.	2020	Macrophage migration inhibitory factor as a predictor for long-term improvements after mindfulness-based group therapy or treatment as usual for depression, anxiety or stress and adjustment disorders	10.1007/s12671-020-01352-3	Did not prospectively match patients to treatments
Van Bronswijk et al.	2021	Cross-trial prediction in psychotherapy: External validation of the personalized advantage index using machine learning in two Dutch randomized trials comparing CBT versus IPT for depression	10.1080/10503307.2020.1823029	Did not prospectively match patients to treatments
Van Bronswijk et al.	2021	Selecting the optimal treatment for a depressed individual: Clinical judgment or statistical prediction?	10.1016/j.jad.2020.09.135	Did not prospectively match patients to treatments
Van Bronswijk et al.	2018	The impact of personality disorder pathology on the effectiveness of cognitive therapy and interpersonal psychotherapy for major depressive disorder	10.1016/j.jad.2017.08.043	Did not prospectively match patients to treatments
Vitinius et al.	2019	Somatic and sociodemographic predictors of depression outcome among depressed patients with coronary artery disease - a secondary analysis of the SPIRR-CAD study	10.1186/s12888-019-2026-6	Did not prospectively match patients to treatments
Wallace et al.	2013	A novel approach for developing and interpreting treatment moderator profiles in randomized clinical trials	10.1001/jamapsychiatry.2013.1960	Did not prospectively match patients to treatments

Young et al.	2018	Drinking to cope moderates the efficacy of changing veteran drinking norms as a strategy for reducing drinking and alcohol-related problems among U.S. veterans	10.1037/adb0000347	Did not prospectively match patients to treatments
Zilcha-Mano et al.	2016	Reducing dropout in treatment for depression: Translating dropout predictors into individualized treatment recommendations	10.4088/JCP.15m10081	Did not prospectively match patients to treatments
Chilvers et al.	2001	Antidepressant drugs and generic counselling for treatment of major depression in primary care: Randomised trial with patient preference arms	10.1136/bmj.322.7289.772	Matched by patient preference
Dunlop et al.	2017	Effects of patient preferences on outcomes in the predictors of remission in depression to individual and combined treatments (PReDICT) Study	10.1176/appi.ajp.2016.16050517	Matched by patient preference
Handelzalt & Keinan.	2010	The effect of choice between test anxiety treatment options on treatment outcomes	10.1080/10503300903121106	Matched by patient preference
Hegerl et al.	2010	Effects of pharmacotherapy and psychotherapy in depressed primary-care patients: A randomized, controlled trial including a patients' choice arm	10.1017/S1461145709000224	Matched by patient preference
Hell et al.	2021	The impact of free choice in alcohol treatment. Primary outcomes of the self-match study	10.1016/j.drugalcdep.2021.108587	Matched by patient preference
Kwan et al.	2010	Treatment preference, engagement, and clinical improvement in pharmacotherapy versus psychotherapy for depression	10.1016/j.brat.2010.04.003	Matched by patient preference
Leykin et al.	2007	The relation of patients' treatment preferences to outcome in a randomized clinical trial	10.1016/j.beth.2006.08.002	Matched by patient preference
Lin et al.	2005	The influence of patient preference on depression treatment in primary care	10.1207/s15324796abm3002_9	Matched by patient preference
Lindegaard et al.	2020	Internet-based psychodynamic therapy vs cognitive behavioural therapy for social anxiety disorder: A preference study	10.1016/j.invent.2020.100316	Matched by patient preference

Bergbom et al.	2014	Early psychologically informed interventions for workers at risk for pain-related disability: Does matching treatment to profile improve outcome?	10.1007/s10926-013-9478-1	No mental health problem identified
Conrad et al.	2015	The changeability and predictive value of dysfunctional cognitions in cognitive behavior therapy for chronic tinnitus	10.1007/s12529-014-9425-3	No mental health problem identified
Jansen et al.	2019	Stepped care targeting psychological distress in head and neck cancer and lung cancer patients: which groups specifically benefit? Secondary analyses of a randomized controlled trial	10.1007/s00520-019-04714-3	No mental health problem identified
Krebber et al.	2016	Stepped care targeting psychological distress in head and neck cancer and lung cancer patients: a randomized, controlled trial	10.1093/annonc/mdw230	No mental health problem identified
Lackner et al.	2019	Factors associated with efficacy of cognitive behavior therapy vs education for patients with irritable bowel syndrome	10.1016/j.cgh.2018.10.033	No mental health problem identified
Fisher et al.	2019	Open trial of a personalized modular treatment for mood and anxiety	10.1016/j.brat.2019.01.010	Not a randomised controlled trial
Gunlicks-Stoessel et al.	2019	Latent profiles of cognitive and interpersonal risk factors for Adolescent depression and implications for personalized treatment	10.1007/s10802-019-00552-3	>50% of participants under 18 years old
Arndt et al.	2020	Identifying change-dropout patterns during an internet-based intervention for depression by applying the muthen-roy model	10.1080/16506073.2018.1556331	Outcome not recorded using a validated patient-reported measure, therapist-reported measure or diagnostic interview
Vaiva et al.	2018	Combining postcards, crisis cards, and telephone contact into a decision-making algorithm to reduce suicide reattempt: A randomized clinical trial of a personalized brief contact intervention	10.4088/JCP.17m11631	Unable to access in English

Supplemental Material C: Risk of Bias Assessments

Author and Date	Risk of bias arising from randomisation process	Risk of bias due to deviations from the intended interventions	Risk of bias due to missing outcome data	Risk of bias in measurement of the outcome	Risk of bias in selection of the reported result	Overall rating
Berger_2014	Low	Low	Low	Some Concerns	Some Concerns	Some Concerns
Carlbring_2011	Low	Some Concerns	Low	High	Some Concerns	High
Coates_2018	Low	Low	High	High	Low	High
Delgadillo_2022	Low	Low	Low	Low	Low	Low
Fletcher_Chondros_2021	Low	Low	Some Concerns	Some Concerns	Low	Some Concerns
Fletcher_Spittal_2021	Low	Low	Low	Low	Some Concerns	Some Concerns
Johansson_2012	Low	Some Concerns	Some Concerns	Some Concerns	Some Concerns	High
Kadden_2001	High	Low	Low	High	Some Concerns	High
Lutz_2022	Some Concerns	Some Concerns	Low	Low	Some Concerns	Some Concerns
McLellan_1997	Some Concerns	Some Concerns	High	High	Some Concerns	High
Moritz_2016	Low	High	Low	Some Concerns	Some Concerns	High
Nordgren_2014	Low	Low	Low	Some Concerns	Low	Some Concerns
Schulte_1992	Some Concerns	Some Concerns	High	High	Some Concerns	High
Silfvernagel_2012	Low	Low	Low	High	Low	High
Van Straten_2006	Low	Low	Some Concerns	Low	Some Concerns	Some Concerns
Vernmark_2010	Low	Some Concerns	Low	Some Concerns	Some Concerns	Some Concerns
Watzke_2010	Low	Low	Low	Low	Some Concerns	Some Concerns

Supplemental Material D: Characteristics of Included Studies

Study	Design and Type of Personalization	Country	Setting	Participants	Mental Health Conditions	Interventions	Primary Outcome Measures
Berger et al. (2014)	RCT IT Component-Level, Semi-formal Structure	Switzerland, Germany, and Austria	Online	*N* = 132, 44% male, 56% female). Mean age = 35.1 years (*SD* = 11.4, range = 18 – 65)	86% of participants met diagnostic criteria for SAD, 33% for PD and 25% for GAD	Tailored CBT-based internet treatment, Disorder-specific CBT-based internet treatment, Waitlist (control)	BAI BDI GSI
Carlbring et al .(2011)	RCT IT Component-Level, Semi-formal Structure	Sweden	Online	*N* = 54, 76% male, 24% female. Mean age = 38.8 (*SD* = 10.7)	Participants recruited from a waiting list of people interested in internet-based treatment for GAD, SAD or PD. For inclusion, participants had to meet DSM-IV criteria for an anxiety disorder	Tailored CBT-based internet treatment (modules prescribed based on comorbidities and SCID-I), Confidential online support group targeting anxiety problems (control)	BAI CORE-OM MADRS-S QOLI
Coates et al. (2018)	RCT IT Component-Level, Semi-formal Structure	Australia	Hospital outpatient drug and alcohol service	*N* = 379, 65% male, 35% female; mean age = 44.3 years, *SD* = 10.8	All participants met DSM-IV criteria for alcohol dependence	Targeted face-to-face CBT (modules chosen based upon pre-treatment assessment scores) Standardized face-to-face CBT	Percentage of drinking days Quantity of alcohol consumed

Delgadillo et al. (2022)	Double-blind, cluster RCT TM Intensity-Level, Statistical Model	United Kingdom	IAPT services – Primary Care	N = 951, 35% male, 65% female, mean age = 38.3 years, SD = 14.5 (stratified care = 583, stepped care = 368)	Patients presenting with depression and/or anxiety disorders.	Low intensity CBT and high intensity CBT. Stratified treatment selection based on a statistical model using machine learning, drawing upon a range of clinical and demographic factors.	Reliable and clinically significant improvement on the PHQ-9
Fletcher, Chondros et al. (2021)	RCT TM Intensity-Level, Semi-formal Structure	Australia	GP Practices – Primary Care	*N* = 1868 adults aged 18-65, 32% male, 68% female, mean age = 35.5 years, SD = 12.1	Patients with a score of 2 or more on the two-item version of the PHQ-9 or GAD-7, or patients using medication for their mental health.	Matched care, or usual care plus attentional control. Matched care involved an e-health platform (Target-D), consisting of symptom feedback, priority-setting and treatment matched to prognosis. Treatments included online self-help, online psychological therapy and nurse-ed collaborative care.	PHQ-9 scores at 3 months post-randomisation
Fletcher, Spittal et al. (2021)	Stratified RCT TM Intensity-Level, Semi-Formal Structure	Australia	GP Practices – Primary Care	*N* = 1671 adults aged 18-75 years Control *N* = 837, 28% male, 73% female, mean age = 39.5 (*SD* = 14.8) Intervention *N* = 834, 26% male, 73% female, mean age = 39.7, *SD* = 15.1	Patients with a score of 2 or more on the two-item version of the PHQ-9 or GAD-7, or patients using medication for their mental health.	Prognosis-matched care, or usual care plus attention control. Prognosis was determined by patient responses to a 23-item decision support tool which assessed psychosocial factors. Interventions included low intensity care (an online programme) and high intensity care (nurse led collaborative care).	Change in score on the K10 at 6 months post-randomisation.

Johansson et al. (2012)	RCT IT Component-Level, Semi-Formal Structure	Sweden	Online	*N* = 121, 29% male, 71% female, mean age = 44.7, (*SD* = 12.1)	Participants were recruited from a waiting list of people who had expressed an interest in internet-based treatment for depression. All participants had a diagnosis of MDD, 55% of whom had a comorbid anxiety disorder	Tailored (specific chapters for comorbid symptoms) online CBT-based guided self-help Standardized (non-tailored), online CBT-based guided self-help A monitored online discussion group (control)	BDI-II
Kadden et al. (2001)	RCT TM Package-Level, Artisanal Structure	United States	University Research Clinic	*N* = 250, 66% male, 34% female, mean age = 45 (*SD* = 10.7)	All participants met DSM-IV criteria for alcohol dependence (98%) or alcohol abuse (2%)	Patients were matched to face-to-face CBT, or Interactional Therapy based on levels of sociopathy and psychopathology, or randomly assigned to these same two treatments	Proportion of days abstinent Proportion of days heavy drinking
Lutz et al. (2022)	RCT IT Component-Level, Semi-Formal Structure	Germany	University Outpatient Clinic	*N* = 538 therapist-patient dyads, 64% of patients were female, 36% male, mean age = 36.3 (*SD* = 13.7)	According to SCID-I, primary diagnoses were affective disorder (50.7%), anxiety disorder (16.2%), adjustment disorder (12.6%), PTSD (8.6%), obsessive compulsive disorder (3.5%), somatoform disorder (3.4%), eating disorder (1.3%)	Patients were treated with CBT. A data-driven clinical decision support system based upon psychometric questionnaires gave therapists recommendations about treatment components and delivery before and during treatment.	Composite measure combining the: PHQ-9 GAD-7 HSCL-11 OQ-30 QEP-2
McLellan et al. (1997)	RCT TM Intensity-Level, Artisanal Structure	Philadelphia, United States	Inpatient and Outpatient private treatment programmes for substance use	*N* = 130 adults, 70% male, 30% female, average age = 38 years,	All participants were dependent on alcohol and/or drugs according to DSM-III)	Patients were randomised to standard or matched services. In matched services, patients received treatment according to responses to the ASI. Treatments included medication and face-to-face therapy.	ASI

Moritz et al. (2016)	RCT IT Component-Level, Semi-formal Structure	Russia	Online	*N* = 89, 52% male, 48% female. Mean ages (standardized condition = 24.72 years, *SD* =6.65; tailored condition = 25.50 , *SD* = 8.05; waitlist = 25.41, *SD* = 5.97)	All participants reported OCD symptoms. Participants did not necessarily have an OCD diagnosis, but severity of symptoms were in line with inpatient populations according to validated measures (OCI-R, Y-BOCS, BDI)	Standard MCT self-help book, Tailored MCT self-help book (participants only received chapters relevant to OCD processes they had endorsed in a questionnaire), Waitlist (control)	Y-BOCS OCI-R BDI
Nordgren et al. (2014)	RCT IT Component-Level, Semi-formal Structure	Sweden	Online	*N* = 100, 37% male, 63% female, mean age = 35.4 years	All participants had an anxiety disorder as primary diagnosis, with 58% also having at least one co-morbid axis-I disorder. Participants were predominantly recruited by GPs or nurses in primary care settings	Tailored CBT-based internet treatment (modules selected for comorbidities), Control group (participants were asked weekly about their wellbeing by a therapist)	CORE-OM
Schulte et al. (1992)	RCT IT Component-Level, Artisanal Structure	Germany	University Research Clinic	*N* =120 patients, 36% male, 64% female. Average age = 39.4 years (range = 19-65).	Patients with a diagnosis of a phobia according to DSM-III, without any other diagnosis, following the CIDI	Tailored face-to-face behaviour therapy (choice of techniques personalized according to a case-by-case problem analysis. Standardized behaviour therapy (not tailored), Yoked "variable standard therapy" (not tailored), in which treatment was provided that was originally tailored to a patient in the experimental group	Outcomes on various measures were combined to result in a discrete outcome for each patient

Silfvernagel et al. (2012)	RCT IT Component-Level, Semi-Formal Structure	Sweden	Online	N = 57 participants, 35% male, 65% female. Mean age = 32.4 (*SD* = 6.9)	Participants were recruited from on an online list of people interested in internet CBT research for panic disorder and generalised anxiety disorder. Participants had to report recurrent panic attacks for inclusion	Tailored CBT-based online treatment (modules selected according to responses to the SCID-I and clinical impression), Waitlist (control)	PDSS
Van Straten et al. (2006)	RCT in routine practice TM Package-Level, Artisanal Structure	Netherlands	Outpatient mental health services	*N* = 702 patients, 61% female, 39% male, mean age = 36.4, *SD* = 10.2	88% participants met DSM-IV criteria for a mood disorder, 53% for an anxiety disorder	Patients were randomised to matched care (not protocolised) or stepped care (protocolised). In matched care a multidisciplinary team matched clients to face-to-face therapy based on presenting problem and patient characteristics. Treatments included interpersonal, supportive, psychoanalytic and eclectic. In stepped care, brief therapy or CBT were the first steps, with patients allowed to switch.	Recovery after 12 months and at study completion, defined as having no mood or anxiety disorder according to DSM-IV by telephone interview following the CIDI
Vernmark et al. (2010)	RCT IT Component-Level, Artisanal Structure	Sweden	Online	*N* = 88 participants, 68% female, 32% male, mean age = 36.8 years (*SD* = 12.9)	Participants were required to score at less than 31 and more than 14 on the MADRS-S, or have a diagnosis of MDD according to DSM-IV	Tailored CBT-based email therapy (text created by therapists based on individual case conceptualisation), CBT-based guided self-help (not tailored), Waitlist (control)	BDI

Watzke et al. (2010)	Two-level RCT TM Package-Level, Artisanal Structure	Germany	Psychotherapeutic in-patient	*N* = 291, 27% male, 73% female, mean age = 43 years (*SD* = 10.7)	All participants were psychotherapeutic inpatients diagnosed with a mental health problem according to the ICD–10	Participants were randomly assigned to Systematic Treatment Selection (STS) or random treatment selection. STS recommendations were based on diagnosis and patient goals. Interventions were brief group CBT and brief group Psychodynamic therapy.	GSI

Note: ASI = Addiction Severity Index; BAI = Beck Anxiety Inventory; BDI = Beck Depression Inventory; CIDI = Composite International Diagnostic Interview; CORE-OM = Clinical Outcomes in Routine Evaluation Outcome Measure; DSM = Diagnostic and Statistical Manual for Mental Disorders; GAD = Generalised Anxiety Disorder; GAD-7 = Generalised Anxiety Disorder-7 questionnaire; GSI = Global Severity Index; ICD = International statistical Classification of Diseases; IT = Individually Tailored treatment study; K10 = Kessler Psychological Distress Scale-10; MADRS = Montgomery-Åsberg Depression Rating Scale self-report; OCD = Obsessive Compulsive Disorder; OCI-R = Obsessive Compulsive Inventory-Revised; PHQ-9 = Patient Health Questionnaire-9; PD = Panic Disorder; PDSS = Panic Disorder Severity Scale; QOLI = Quality of Life Inventory; RCSI = Reliable and Clinically Significant Improvement; RCT = Randomised Controlled Trial; SAD = Social Anxiety Disorder; SCID = Structured Clinical Interview for the DSM-IV; SD = Standard Deviation; TM = Treatment Matching study; YBOCS = Yale Brown Obsessive Compulsive Scale.

Supplemental Material E: Primary Outcomes of Included Studies

Study	Total *N*	Analysed *N*	Narrative Outcome	Statistical Outcome at Post Treatment and Follow-up (FU)
Berger et al. (2014)	132	132 (Tailored = 44, Standardized = 44, Waitlist = 44)	Tailored interventions superior to waitlist at posttreatment and 6-month follow-up. However, no differences found between tailored- and disorder-specific interventions at posttreatment or 6-month follow-up.	The two intervention groups did not differ significantly on any measure at posttreatment, all *p*s > 0.26. Individual *P* values for differences on each measure were not reported. Effect sizes for Tailored vs Standardized were -0.05 (BAI), 0.12 (BDI-II) and -.02 (GSI). Effect sizes for Tailored vs Waitlist were 0.87 (BAI), 0.83 (BDI-II) and 0.75 (GSI). 6-month FU: No group by time interaction was found, F(1, 58.6 – 80.9) = 0.03 to 2.06, all ps > .20), indicating no significant difference between standard and tailored treatments at 6-month follow-up. Effect sizes not provided.
Carlbring et al. (2011)	54	54 (Tailored = 27, support = 27)	A moderate effect was found in favour of the tailored intervention group relative to the control group across all measures at posttreatment.	BAI: tailored intervention significantly more effective, F (1,51) = 9.53, $p < .01$, effect size = 0.38. CORE-OM: tailored intervention significantly more effective, F (1, 51) = 22.04, $p < .001$, effect size = 1. MADRS-S: tailored intervention significantly more effective, F (1, 51) = 10.64, $p < .01$, effect size = 0.69. QOLI: tailored intervention significantly more effective, F (1, 51) = 8.56, $p < .01$, effect size = 0.69.
Coates et al. (2018)	379	379 (TAU = 193, Targeted = 186)	No significant difference between targeted and standardized interventions in drinking days or consumption at posttreatment.	No significant difference in drinking days, b = 0.90 (SE = 1.07), $p = 0.096$. No significant difference in consumption, b = 0.94 (SE = 1.08), $p = 0.422$.
Delgadillo et al. (2022)	951	951 (stratified care = 583, stepped care = 368)	Remission significantly higher in the stratified care group relative to standard care at posttreatment.	PHQ-9: At posttreatment, participants in the stratified (personalized) care group were significantly more likely to improve than in the stepped care group (RCSI: 52.3% vs. 45.1%; OR = 1.40, $p = 0.025$). Stratified care cases were also significantly more likely to meet IAPT criteria for recovery (48.2% vs 43.7%; OR = 1.33, $p = .043$).

Fletcher, Chondros et al. (2021)	1868	1262 (analysed at 3-month follow-up, 610 = matched, 673 = usual care plus attentional control)	Matched care saw greater improvements in depressive symptoms relative to usual care plus attentional control at 3-month post-treatment. This was maintained at 12-month follow-up.	PHQ-9: The difference in mean depression scores at 3-month post-intervention was -.088 (95% CI = -1.45 to -.031) in favour of the matched care group, p = .003. The standardized mean difference was -0.16 (95% CI = -0.26 to -0.05). 12-month FU: PHQ-9: The difference in mean depression scores was -.059 (95% CI = -1.18 to 0.01) in favour of the matched care group, p = .05. Standardized mean difference was calculated as -0.10 (95% CI = -0.21 to 0.002).
Fletcher, Spittal et al. (2021)	1671	1125 (analysed at 6-month follow-up, 547 = matched, 578 = usual care plus attentional control)	Greater reductions in psychological distress were found in the matched care group than in usual care plus attention control group at 6-month post-treatment. This difference became non-significant at 12-month post-treatment.	K10: The mean difference between groups was -0.88 (95% CI -1.66 to -0.11) in favour of the matched care group, p = .03. The standardized mean difference was calculated at -.09 (-0.17 to -0.01). 12-month FU: K10: The mean difference between groups was -0.55 (95% CI -1.39 to 0.30), p = .21. The standardized mean difference was calculated at -0.06 (-0.14 to 0.03) (non-significant).
Johansson et al. (2012)	121	115 (36 = tailored treatment, 37 = standardized treatment, 42 = control)	Both tailored treatment and standardized treatment resulted in greater symptom reduction than the discussion group. Tailored treatment led to greater symptom reduction and higher recovery rates than the standardized treatment in the subgroup of participants with higher levels of depression at baseline and more comorbidity. However, these differences were not significant in the subgroup of participants with lower levels of depression, or in the overall sample.	BDI-II: In the higher depression severity subgroup there was a significant interaction effect of intervention group and time (F(1, 102.4) = 6.19, p<.05) in favour of the tailored treatment relative to standardized treatment. The effect size in the higher depression severity subgroup was d = 0.51 in favour of tailored treatment, and for tailored treatment vs control this was d = 1.29. In the total sample, and in the lower depression severity subgroup, no interaction effects of group and time were found (F and p values not reported). In the total sample, effect size for tailored vs standardized treatment was d = 0.23, and for tailored vs control this was d = 0.84. 12-month FU: BDI-II: Effect size at 12-month follow-up in the high-severity depression group was d = 0.69 (in favour of tailored treatment), in the low-severity depression group this was d = -0.11, and in the overall sample this was d = 0.27.
Kadden et al. (2001)	250	250 (122 = matched, 128 = randomly assigned)	Participants who were prospectively matched did not have superior drinking outcomes to those who were assigned randomly, but did have fewer negative consequences of drinking. Participants who were randomly assigned were	PDA and PDH: Type of treatment assignment (prospectively matched vs random) did not have a significant effect on percentage of days abstinent (PDA) or percentage of days heavy drinking (PDH) (statistical results not reported).

			more likely to be abstinent at end of treatment, but not at follow-up.	3,6,9,12-month FU: No differences were found in PDA or PDH between prospectively matched, randomly-assigned matched or randomly-assigned mismatched participants across follow-up times. Statistical results not reported.
Lutz et al. (2022)	817	614 (328 = decision-support, 232 = standardized)	When therapists followed the recommended treatment strategy in the first ten sessions (i.e. motivational vs problem-solving focus), patients had a better outcome. However, feedback during treatment had no effect on patient outcome.	Composite measure: There was a significant difference in the percentage of change until session ten between patients treated with their optimal and patients treated with a nonoptimal strategy (d = 0.28, p = .033).
McLellan et al. (1997)	130	94 (matched = 45, standardized = 49)	Matched patients did not see a significant improvement on the ASI at 6-months post-treatment relative to standard care patients.	ASI: F = 2.15 (df = 1, 89), p <0.09.
Moritz et al. (2016)	89	50	Both metacognitive conditions demonstrated significant improvements in OCD symptoms relative to waitlist. However, no significant differences were found between the tailored and standardized interventions.	Y-BOCS, OCI-R and BDI: The authors did not report the statistics for the differences on these measures between the standardized and tailored groups, other than to state none of the analyses produced any differences (p > .3 or greater for all).
Nordgren et al. (2014)	100	91 (46 = tailored, 45 = control), 75 at one-year follow-up	Rates of clinically significant improvement were significantly higher in the tailored group than the control group (weekly check-ins)	CORE-OM: Clinically significant improvement was 46% for the tailored group at post-treatment, compared to 12% for the control condition. Scores on the CORE-OM were significantly improved in the tailored group relative to the control group, F (1,89-97) = 5.097, p < .05, between group d = 0.86. 12-month FU: Between-group effect sizes not reported.
Schulte et al. (1992)	120	97 after drop-out (33 = individualized, 30 = standardized, 34 = yoke control)	The authors concluded that standardized group was superior to the tailored and yoked groups	When the scores for the measures were combined to result in a discrete outcome for each patient, patients receiving standardized treatment had better outcomes than those receiving the individualized or yoked treatment, Chi^2 (4) = 13.6, p < .01.

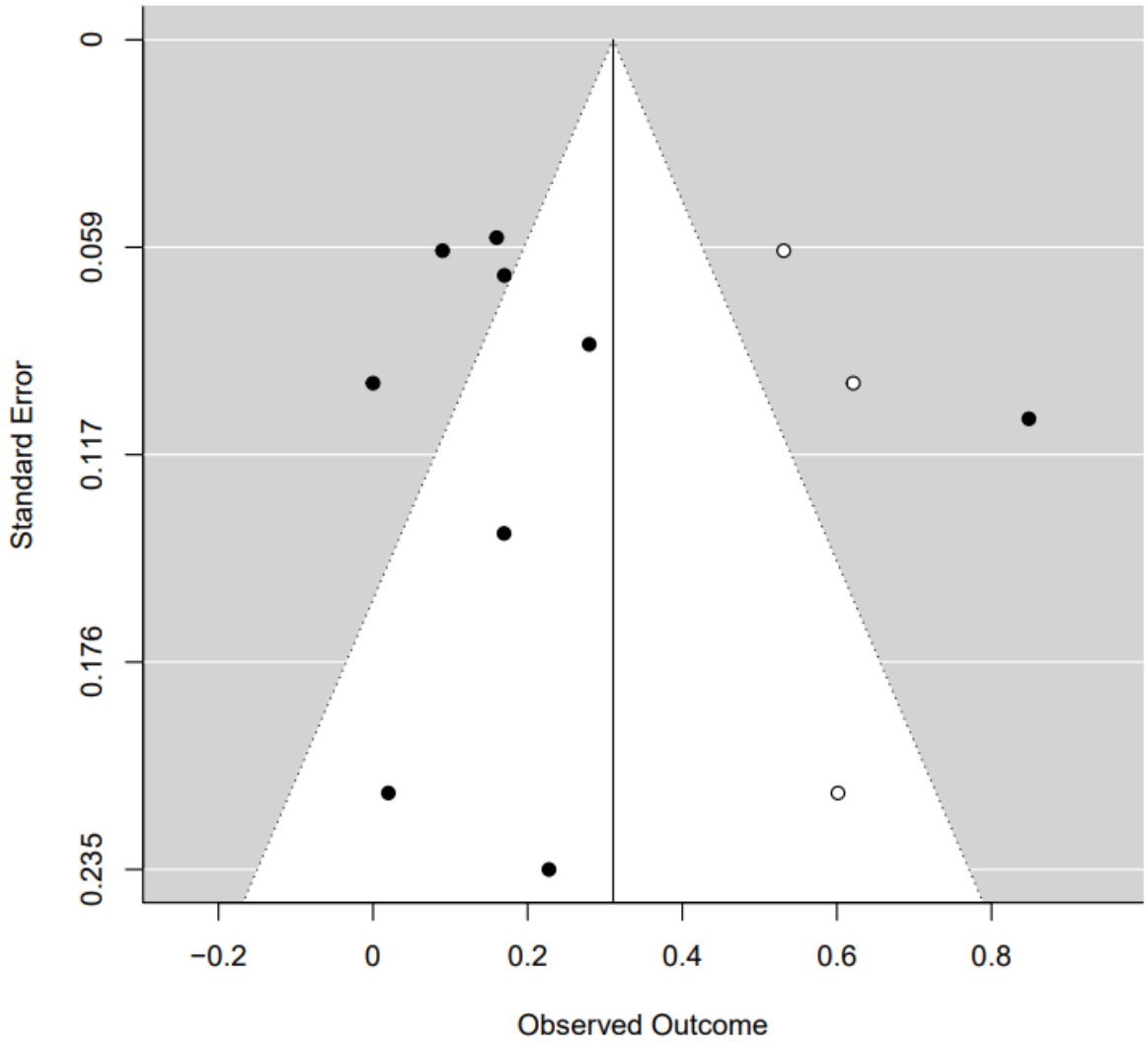
Silfvernagel et al. (2012)	57	57 at post-treatment (29= tailored, 28 = waitlist), 29 at 12-month follow-up	The tailored condition was significantly more effective in reducing panic symptoms than waitlist at post-treatment and 12 month follow up.	PDSS: At post treatment, the tailored condition was significantly more effective than the control condition, F (1,47.3) = 29.6, p < .001, between group d = 1.41. 12-month FU: PDSS: A mixed-models within-group analysis for the treatment condition showed a significant effect of time F (1, 18.3) = 19.5, p < .001, estimated mean = 4.80 (SE = 1.18), (SD 4.66), within-group d = 1.66.
Van Straten et al. (2006)	702	451 at 12-month (171 = matched, 139 = CBT, 139 = BT), 479 at 18-24 month	No statistically significant difference was found between matched care and stepped care.	CIDI: Odds ratios of recovery using ITT at 12-months were: 1.0 for matched care (reference), 1.36 (95% CI: 0.87-2.12) for CBT and 1.48 (0.94-2.32) for BT in stepped care. The difference between matched and stepped care was not statistically significant (p = .09) 18–24-month FU: CIDI: Odds ratios of recovery using ITT were: 1.0 for matched care (reference), 1.26 (95% CI: 0.81-1.98) for CBT and 1.41 (0.89-2.25) for BT in stepped care (not statistically significant).
Vernmark et al. (2010)	88	85 at post-treatment (29 = tailored email, 27 = self-help, 29 = waitlist), 75 at 6-month follow-up	Both tailored and standardized groups demonstrated significant reductions in depressive symptoms relative to waitlist. The authors argued that differences between the tailored and standardized groups were small, but favoured the tailored group.	BDI: At posttreatment the tailored group demonstrated significant improvement compared to waiting list, p = .002, between groups effect size d = 0.96, within group effect size for tailored group, d = 2.27 BDI: However, there was no significant difference between the tailored and standardized groups, p = .41. 6-month FU: BDI: Mean BDI score at pre-treatment was 22.2 (SD = 5.3) in the tailored group, changing to 9 (SD = 5.6) at 6-month follow-up. The tailored group within-group effect size was d = 2.42. Between-group statistics not available at follow-up.
Watzke et al. (2010)	291	226 (RTS = 147, STS = 79)	No general effect was observed for STS relative to RTS. However, STS resulted in better long-term outcome for PDT specifically, but not for CBT.	GSI: At 6-months, there was no significant main effect of type of treatment assignment (STS: marginal mean = 0.98, SE = 0.06) (RTS: marginal mean = 1.00, SE = .07), F(1, 226) = 0.13, p = 0.721, partial n^2 = 0.001. However, in terms of differential effectiveness, STS patients in the psychodynamic group benefited more from treatment than patients randomly assigned to psychodynamic (STS: marginal mean = 0.98, SE = 0.11) (RTS: marginal mean = 1.15, SE = 0.09), F(1, 226) = 4.72, p = 0.031, partial n^2= 0.021

Note. ASI = Addiction Severity Index; BAI = Beck Anxiety Inventory; BDI = Beck Depression Inventory; BT= Brief Therapy; CIDI = Composite International Diagnostic Interview; CORE-OM = Clinical Outcomes in Routine Evaluation Outcome Measure; GAD-7 = Generalised Anxiety Disorder-7 questionnaire; GSI = Global Severity Index; IAPT = Improving Access to Psychological Therapies; ITT = Intention To Treat; K10 = Kessler Psychological Distress Scale-10; MADRS = Montgomery-Åsberg Depression Rating Scale self-report; OCI-R = Obsessive Compulsive Inventory-Revised; PDA = Proportion of Days Abstinent; PDH = Proportion of Days Heavy drinking; PHQ-9 = Patient Health Questionnaire-9; PDSS = Panic Disorder Severity Scale; QOLI = Quality of Life Index; RCSI = Reliable and Clinically Significant Improvement; RTS = Random Treatment Selection; SCID = Structured Clinical Interview for the DSM-IV; SD = Standard Deviation; STS = Systematic Treatment Selection; TAU= Treatment As Usual; YBOCS = Yale Brown Obsessive Compulsive Scale

Supplemental Material F: Other Primary Outcome Measures used by Included Studies

Other primary outcome measures used included the Global Severity Index (GSI), the Montgomery-Åsberg Depression Rating Scale Self-report (MADRS-S), the Generalised Anxiety Disorder questionnaire (GAD-7), the Kessler Psychological Distress Scale (K10), the Drinker Inventory of Consequences (DrinC), the Proportion of Days Abstinent (PDA), the Proportion of Days Heavy drinking (PDH), the Addiction Severity Index (ASI), the Obsessive Compulsive Inventory-Revised (OCI-R), the Yale-Brown Obsessive Compulsive Scale (Y-BOCS) and the Panic Disorder Severity Scale (PDSS). Two studies (Schulte et al., 1992; Lutz et al., 2021) used their own composite measures of outcome based upon responses to a range of questionnaires, while another study (Coates et al., 2018) used the percentage of drinking days and quantity of alcohol consumed as their primary measures of outcome. While these three studies did not use a validated measure to assess their primary outcomes, all studies used validated measures within the procedure and therefore met the inclusion criteria. One study (Van Straten et al., 2006) measured outcome based upon recovery according to Diagnostic and Statistical Manual of mental disorders (4^{th} version) (DSM-IV) criteria established by the Composite International Diagnostic Interview (CIDI).

Supplemental Material G: Funnel Plot for Meta-Analysis assessing Personalized Treatment versus Standardized Treatment



Supplemental Material H: Funnel Plot for Meta-Analysis assessing Personalized Treatment versus Control Groups