Therapist facilitative interpersonal skills and the relation to psychotherapy outcome

Charlotte Bentham

Submitted in partial fulfilment for the award of Doctor of Clinical Psychology at the University of

Sheffield

The University of Sheffield

Faculty of Science

Department of Psychology

May 2021

Declaration

I declare that this work has not been submitted for any other degree at the University of Sheffield or any other institution. This thesis is my own original work and all other sources have been referenced accordingly.

Word Count

Literature Review

Excluding references, tables, and appendices	7805
Including references, tables, and appendices	16115

Research Report

Excluding references, tables, and appendices	7995
Including references, tables, and appendices	19003

Total

Excluding references, tables, and appendices	15800
Including references, tables, and appendices	35118

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Gillian Hardy, for her support, positivity, and advice throughout the course of this project. My thanks also go to Professor Michael Barkham and Dr. Dave Saxon for their theoretical and statistical knowledge which helped to bring this project together. Thank you also to Associate Professor Kim de Jong and Kane Steggles for the time and expertise they shared during the FIS training. Thank you to Merinda Haake and Melanie Simmonds-Buckley for their ongoing support with the project. It has been a pleasure working with you all.

A massive thank you to my fellow trainees. I feel lucky to have made life-long friendships with such a lovely bunch of thoughtful, kind, and passionate people.

A special thanks goes to my fellow FIS researcher, and thesis buddy, Annie Faulds. Thank you for the constant messages of encouragement and reassurance, this journey would have been much less fun, and more stressful, without you.

To my grandparents, Alwyn and Julie, I would like to express my thanks for all of the support, advice and food care parcels they have provided over the last three years.

I could not have made it through the last three years without the love and support of my partner, Matt. Thank you for your patience, for holding things together when I was busy doing work, and for renovating a beautiful home for us to live in.

And finally, to my lovely mum, Tracy, thank you for your love, selflessness, and for your unwavering faith that I would one day become a Clinical Psychologist. I dedicate this thesis to you.

Lay Summary

The interpersonal characteristics of a therapist, such as their ability to show warmth, empathy, and hope, have been found to influence how beneficial therapy is for a client. A measure called the facilitative interpersonal skills task was developed to rate how much therapists use interpersonal skills in simulated therapy sessions with clients. One important interpersonal skill is called the therapeutic alliance, which measures agreement between a client and therapist on the "what" and "how" of therapy, as well as a feeling of closeness between them.

Part One is a meta-analytic review aiming to look at whether clients with a weak therapeutic alliance with their therapist are more likely to drop out of therapy early. The review brought together the findings of 25 studies and confirmed that there is a relationship of medium strength between the therapeutic alliance and a client dropping out of psychotherapy. Other factors such as the client's diagnosis or educational history, the type and length of therapy, and the way studies measured alliance and dropout, did not make an important difference to the strength of this relationship. However, the strength of the relationship was weaker in studies that had not published their findings in journals, compared to those that had.

Part Two is a research study aiming to look at whether a therapist's facilitative interpersonal skills are linked to therapy being more successful for clients. To do this, the measure had to be adapted to rate therapists' skills in actual client therapy sessions that were not going well. Overall, the interpersonal skills a therapist used in the session did not link to more successful treatment for clients. Furthermore, the relationship between a therapist's skills and the success of treatment did not differ based on the type of therapy clients received. The study did find that there were different patterns in how therapist's skills impact on treatment success based on how severe the client's depression was when they started therapy. For clients with moderate depression, treatment was more successful when the therapist used more interpersonal skills. For clients with severe depression, treatment was more successful when the therapist used less interpersonal skills.

Overall, the findings suggest that the interpersonal skills of a therapist can make a difference to how successful psychotherapy is for a client and whether they might drop out early. The findings suggest the need for therapists to pay close attention to their use of interpersonal skills and the therapeutic alliance with a client to maximise the benefit of psychotherapy.

Table of Contents

Part One: Literature Review

Therapeutic alliance and premature termination: An updated meta-analysis of adult individual

psychotherapy

Part Two: Research Report

The role of therapists' interpersonal skills during "not on track" therapy sessions in predicting

response to counselling person-centred experiential and cognitive behavioural therapy interventions

for depression

Part One: Literature Review

Therapeutic alliance and premature termination: An updated meta-analysis of adult individual psychotherapy

Abstract

Objectives

Premature termination of psychotherapy is a common and problematic issue. Therapeutic alliance, the agreement between a client and therapist on the tasks and goals of therapy, in the presence of an affective bond, has been associated with psychotherapy outcomes. The current meta-analytic review aimed to quantify the associations between therapeutic alliance and psychotherapy dropout and examine potential modulatory factors in this relationship.

Method

A systematic database search was conducted to identify twenty-five papers reporting an association between therapeutic alliance and dropout from individual adult psychotherapy, which were synthesised in a meta-analytic review. Additional variables were examined as potential moderators. Quality appraisal assessed the risk of methodological bias of the research field.

Results

Aggregating data across studies revealed a small-moderate effect size ($d = 0.443$; 95% CIs: $[0.295, 0.591]$) between therapeutic alliance and premature termination. Moderator analyses indicated the alliance-dropout relationship was moderated by publication status, with published studies reporting larger effect sizes than unpublished research. The alliance-dropout association was not moderated by psychotherapy orientation, therapy length, client diagnosis, client education, alliance rater or timepoint, or the operationalisation of dropout used.

Conclusions

The current meta-analysis provides further evidence that establishing robust therapeutic alliance in psychotherapy is an important factor in reducing client unilateral termination. The association between therapeutic alliance and dropout is largely consistent across psychotherapy orientation and treatment length, client diagnosis and education, and definitions of alliance and dropout. The effect sizes vary as a function of publication status, suggestive of publication bias across the literature.

Practitioner Points

- Psychotherapy dropout is common and problematic for clients and therapists.

- Therapy dyads with higher levels of therapeutic alliance are less likely to end with client premature termination.

- Monitoring and discussing the therapeutic alliance across psychotherapy may be beneficial for identifying ruptures/disagreement in the working relationship that may increase likelihood of premature termination.

Limitations

- The findings of the current review are not generalisable beyond face-to-face individual adult psychotherapy.

- The review focuses on the association between alliance strength and premature termination, potentially masking dynamic temporal and alliance rupture patterns that may be underlying the association.

<center>**Introduction**</center>

Premature Termination

Premature termination encompasses a range of situations where a client begins an intervention but discontinues prior to a resolution or recovery from the problems for which they were seeking treatment. Dropout from psychotherapy is a relatively common problem, with an estimated rate of 17% dropout in efficacy trials, and 26% in effectiveness studies (Swift & Greenberg, 2012). Premature termination may occur for many reasons: it may indicate a perceived lack of progress in treatment or discontentment with the goals and tasks of therapy, dropout may also be the result of extra-therapeutic reasons, or personal factors interfering with engagement. On occasion, psychotherapy dropout may indicate the client's perceptions that their problem has been resolved, and therapy is no longer needed. Alternatively, clients unilaterally terminating therapy have often been found to have less improvement over the course of therapy compared to continuers (Holmes et al., 2019). Premature termination is problematic for several reasons. Clients prematurely terminating therapy have been shown to have poor treatment outcomes (Cahill et al., 2003) and reported more dissatisfaction with treatment (Björk et al., 2009). Furthermore, attrition and missed appointments are a strain on mental health services, are an ineffective use of staff time, increase waiting lists, and prevent others accessing services (Barrett et al., 2008). Given the detrimental impact, increasing understanding of factors contributing to psychotherapy dropout is essential.

Across the psychotherapy literature, there are inconsistencies in how premature termination is defined. The first operational definition identifies dropouts as clients who attend less than the specified number of sessions agreed upon by the therapist and client, based upon the dose-response literature that an adequate number of sessions are required for symptomatic improvement (Lambert, 2007). The second definition is dropout prior to the completion of an entire treatment, or treatment protocol (Swift & Greenberg, 2012). An issue with the first two methods of defining premature termination is the potential for misclassification of clients who recover early in treatment (Swift &

Greenberg, 2012). The third operationalisation of dropout identifies clients who have missed scheduled appointments without attempting to reschedule or attend future appointments (Swift & Greenberg, 2012). The fourth characterises those who do not return after their initial therapy intake session (Huang et al., 2013; Longo et al., 1992). The fifth definition employs the therapist's judgement, asking therapists to use their experience of the client and therapy to discern between terminators who have demonstrated sufficient treatment gains (completers) from terminators who have not demonstrated clinically significant improvement (dropouts). A limitation of this approach is the potential discrepancy in perceptions of reasons for discontinuance, and disagreement in the goals and expectations of therapy, between therapists and clients (Todd et al., 2003). A sixth conceptualisation was introduced by Hatchett and Park (2003), defining dropout as termination before reaching clinically significant change and/or scoring within the non-clinical range on symptom outcome measures. Rates of dropout from psychotherapy have significant variation based on the definition used (Wierzbicki & Perarik, 1993); with dropout rates higher when determined by therapist judgement (37.6%) compared with other methods of operationalisation, including completing a specified number of sessions (18.3%) or treatment protocol (18.4%; Swift & Greenberg, 2012).

Barriers to treatment and strategies to encourage engagement and retention are poorly understood (Barrett et al., 2008). Among potential predictors of dropout, pre-treatment client demographic and clinical characteristics have been the most extensively examined to date (Roos & Werbart, 2013). Client age, gender, socioeconomic status, educational levels and clinical diagnosis have been associated with differential dropout rates (Swift & Greenberg, 2012; Williams et al., 2005). Therapist influences are less well researched; however, therapist experience level has been found to influence dropout, but other therapist demographic variables such as age, gender and ethnicity do not (Swift & Greenberg, 2012). In clinical practice, the therapist's capability to engage the client may be key to treatment retention, therefore understanding relationship and process variables predicting and preventing premature termination is required (Roos & Werbart, 2013).

Relational qualities, such as therapists' emotional intelligence, empathy, warmth and regard, and process factors such as the quality of the therapeutic alliance, disagreement, and negative process, have been found to influence attrition (Roos & Werbart, 2013). Research is now increasingly focused on transdiagnostic therapy-related common factors that influence psychotherapy dropout (Sly, 2009). One such common factor is the therapeutic alliance.

Therapeutic Alliance

Theorists, researchers, and clinicians use various terms to describe aspects of the therapeutic relationship, including the therapeutic alliance, working alliance, and helping alliance (Martin et al., 2000). Despite variability in the conceptualisation of therapeutic alliance, most theoretical operationalisations define the alliance as comprising of three components in the relationship between a client and therapist: 1) agreement on the goals of therapy; 2) collaborative agreement on the tasks of therapy; and 3) the quality of the affective interpersonal bond (Bordin, 1979). Within research and clinical practice, a variety of measures have been developed and used to characterise alliance in psychotherapy relationships. The primary measures used across the literature include, but are not limited to, the Working Alliance Inventory (WAI; Hatcher & Gillaspy, 2006; Horvath & Greenberg, 1989; Tracey & Kokotovic, 1989), Helping Alliance Questionnaire (HAQ; Alexander & Luborsky, 1986; Luborsky et al., 1983), Vanderbilt Therapeutic Alliance Scale (VTAS; Suh et al., 1986) and California Psychotherapy Alliance Scales (CALPAS; Marmar et al., 1986). Together these measures have been used in over two-thirds of the research on the alliance (e.g. Flückiger et al., 2018), and have been found to be highly correlated with a shared underlying factor structure (Hatcher & Barends, 1996). Evidence from meta-analytic reviews consistently support a moderate association between therapeutic alliance and a range of psychotherapy outcomes, across treatment orientations, even when controlling for baseline clinical severity (Flückiger et al., 2018; Horvath et al., 2011; Martin et al., 2000). One outcome related to therapeutic alliance is premature termination from psychotherapy.

Therapeutic Alliance and Premature Termination

The relationship between premature termination and alliance is important to understand in isolation from other measures of psychotherapy outcome. The strength and patterns of alliance in clients who ultimately terminate treatment may be distinct from those who complete therapy, due to the reduced length of treatment and potential for problematic interactions or interpersonal ruptures between the client and therapist. In a meta-analytic review, Sharf et al. (2010) examined the association between psychotherapy dropout and the strength of therapeutic alliance, reporting a moderately-strong association ($d = .55$). As hypothesised by the authors, therapy dyads who exhibited a weaker therapeutic alliance were more likely to end in unilateral termination of psychotherapy. However, the small number of studies meeting review criteria ($k = 11$) increased the chance studies were confounded by idiosyncratic sample characteristics and limited the ability of the review to reliably examine important moderator variables.

Potential Moderators of the Therapeutic Alliance–Dropout Association

Several client characteristics have been assessed as moderating the association between therapeutic alliance and psychotherapy outcome. Client demographic factors, such as age, gender and ethnicity, have been inconsistently linked to differential rates of dropout (Swift & Greenberg, 2012), but demonstrate no evidence of moderating the alliance-dropout association (Sharf et al., 2010). Client education has been found to be important, with higher rates of dropout in those with lower levels of education (Swift & Greenberg, 2012), and a stronger association between alliance and premature termination in this demographic group (Sharf et al., 2010). Client diagnosis/presenting problem also influences dropout rates, with higher degrees of termination in clients diagnosed with a personality or eating disorder, and lower dropout rates in those with an anxiety disorder or psychosis (Swift & Greenberg, 2012). Diagnosis has also been reported as a significant moderator of the alliance-outcome association, with weaker associations reported in those with an eating disorder, and a high degree of variability in effect sizes for those diagnosed

7

with a personality disorder (Flückiger et al., 2018). However, Sharf et al. (2010) did not support client diagnosis as a moderator of the alliance-dropout association, perhaps due to the limited numbers of papers included.

A range of treatment variables have previously been assessed as potential moderators of dropout or important factors in the alliance-outcome association. As therapeutic alliance is defined as a common factor of psychotherapy and a transtheoretical construct (Bordin, 1979), it has been proposed and previously supported that the alliance-outcome association does not vary between treatment orientation (Flückiger et al., 2018), although this has not been specifically assessed with regard to premature termination. The length of treatment may also be a potential moderator, as higher dropout rates have been found in therapies that are not time-limited (Swift & Greenberg, 2012), and longer treatments have been found to have stronger alliance-dropout associations (Sharf et al., 2010).

The operationalisation of therapeutic alliance and premature termination also require consideration as potential moderators of the alliance-dropout association. Previous meta-analytic reviews have found that the alliance-outcome relation differs in strength across raters of alliance (Flückiger et al., 2018; Horvath et al., 2011). Flückiger et al. (2018) reported a trend towards observer-rated alliance being associated with a marginally smaller alliance-outcome effect in comparison to client-rated alliance. Conversely, Horvarth et al. (2011) indicated a trend towards a lower alliance-outcome association based on therapists' evaluations (Horvath et al., 2011). However, alliance rater was not found to be a signficant moderator in the alliance-dropout association in the Sharf et al. (2010) review. The measure of alliance used does not seem to significantly moderate the alliance-outcome relationship (Flückiger et al., 2018). However, there may be a significant modulatory influence of the timepoint when alliance is rated, with the alliance-outcome association reportedly stronger when alliance is measured late in therapy, rather than early in psychotherapy (Flückiger et al., 2018; Horvath et al., 2011). Finally, the definition of dropout

used between studies has been found to signficantly moderate dropout rates, with higher rates of premature termination reported in studies where dropout is characterised by therapist judgement (Swift & Greenberg, 2012). Again, this was not found to be a significant moderator of the alliance-dropout association (Sharf et al., 2010).

Study variables have previously been examined as potential moderators of the alliance-outcome association. Year of publication was not a significant moderator of the alliance-dropout association in the Sharf et al. (2010) review. Publication status is, however, proposed as a study variable of interest in the current review. Stronger, more significant findings are published more often than weaker, less significant findings (Borenstein et al., 2009). As the current review intends to include unpublished work, publication status may be considered an important moderator of the association between premature alliance and dropout from individual psychotherapy.

The Current Review

The primary aim of the current meta-analytic review is to complete an updated examination of the strength of the relationship between therapeutic alliance and premature termination from adult individual psychotherapy. Many variables of interest have been shown to modulate dropout rates and the association between alliance and outcome, but were inconsistently replicated in the previous alliance-dropout meta-analytic review (Sharf et al., 2010), which was potentially underpowered to reveal subgroup differences in effect size. Given the number of potential moderator variables of interest, categorical and continuous variables were selected *a-priori* based on the links to alliance and dropout in the extant literature, and availability of data in the current review papers. The final selection of moderator variables included: psychotherapy orientation, length of therapy, client diagnosis, client education, alliance rater, timepoint of alliance rating, definition of dropout and the publication status of the research.

Hypotheses

1) Therapeutic alliance will be negatively associated with premature termination of psychotherapy; in that lower levels of therapeutic alliance will be associated with a higher degree of dropout (Hypothesis 1 – H1).

2) Associations between therapeutic alliance and premature termination will not significantly vary across different psychotherapy orientations (Hypothesis 2 – H2).

3) Associations between therapeutic alliance and premature termination will vary dependent on the length of therapy; with psychotherapies without session limitations and longer therapies having the strongest alliance-dropout association, relative to short-duration therapies (Hypothesis 3 – H3).

4) There will be diagnosis-related differences in the association between therapeutic alliance and premature termination, with stronger associations reported in clients with a personality disorder than other diagnostic categories (Hypothesis 4 – H4).

5) Studies with a greater representation of clients with higher-level education will report a weaker alliance-dropout association than those with a lower proportion of clients with high levels of education (Hypothesis 5 – H5).

6) Differences in the strength of association between therapeutic alliance and dropout will be influenced by the rater of alliance, with client-rated alliance being the strongest predictor of dropout, relative to therapist- and observer-rated alliance (Hypothesis 6 – H6).

7) The alliance-dropout association will vary dependent on the timepoint in therapy in which the alliance is rated; there will be a stronger association between therapeutic alliance and dropout in studies that rate alliance later in psychotherapy, compared to studies rating alliance early or as an average across the course of psychotherapy (Hypothesis 7 – H7).

8) The proposed association between alliance and premature termination will vary as a function of the operational definition of dropout used, with studies using therapist judgement to assign

dropout status reporting stronger effect sizes than other operational definitions (Hypothesis 8 – H8).

9) The strength of the association will vary dependent on publication status; unpublished studies will report a weaker association between alliance and premature termination compared to published work (Hypothesis 9 – H9).

<center>**Method**</center>

Search Strategy

Prior to commencement of the review, a protocol was published on the Prospero database (https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=235532). The systematic search utilised four electronic databases: PsycINFO, PubMed, Scopus and Web of Science. Search criteria were translated into two categories of key words that were used to identify papers from the title, abstract, and subject headings of primary research papers. The categories of key words selected studies that a) examined the therapeutic alliance (key words: "therapeutic alliance" OR "working alliance" OR "helping alliance"; as used in the review by Flückiger et al., 2018) and b) included clients who had prematurely terminated individual psychotherapy (key words: "drop out" OR dropout OR drop-out OR attrition OR withdrawal OR discontin* OR terminat* OR non-complet* OR noncomplet*). The previous meta-analytic review of papers examining the association between therapeutic alliance and dropout included papers published up to December 2006 (Sharf, 2008; Sharf et al., 2010). Therefore, date restrictions identified papers published between January 2007 and December 2020 (inclusive).

Existing grey literature was also included in the current review to maximise the number of studies for the meta-analysis. Excluding unpublished work can introduce a systematic publication bias, as studies with significant results, and stronger associations, are more likely to be published in scientific journals (Borenstein et al., 2009). The inclusion of grey literature in meta-analyses has, however, been argued to lower the quality of studies, as unpublished work has not been through the journal peer-review process (Borenstein et al., 2009). However, the current study aims to highlight study quality as part of the quality/bias assessment process. Grey literature was identified as part of the main database searches, as well as handsearching of OpenGrey and ProQuest databases.

In accordance with best-practice guidelines, the reference and citation networks of papers meeting the review eligibility criteria were searched for relevant literature that may have been

missed in the initial searches (Aguinis et al., 2011). Additionally, the reference and citation networks of relevant literature reviews were also searched for relevant studies (Flückiger et al., 2018, 2020; Roos & Werbart, 2013; Sharf et al., 2010; Swift & Greenberg, 2012).

Study Selection

In total, 792 references were identified from the systematic searches. The titles and abstracts were screened by the author to establish if papers were relevant to the review question. The full text of all relevant articles ($k = 190$) was then reviewed by the author to determine if the study met the review eligibility criteria. Study eligibility was established using the inclusion and exclusion criteria presented in Table 1. Points of clarification regarding review eligibility were discussed with the research team (GH, DS). A second reviewer (AF) independently assessed the eligibility of a randomly selected 10% of papers assessed at the full-text stage ($k = 19$). Initial interrater agreement on study eligibility was consistent for 94.7% of papers, and discrepancies or queries arising from rating were discussed until consensus was reached. Figure 1 displays the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram illustrating the search procedure (Moher et al., 2009).

Data Extraction

Data was extracted from each of the 25 papers by the author, including therapeutic alliance and premature termination variables, as well as 21 other treatment, client, and study variables that were to be considered as potential moderators. Study protocol papers and other primary or secondary papers were reviewed for relevant data, when needed. To establish the reliability of the data extraction and coding procedures, an independent rater completed secondary rating of a randomly selected 20% ($k = 5$) of the identified papers. An initial 96.4% agreement rate was observed between raters. Discrepancies in data extraction were resolved through discussion and reference to the original papers.

Figure 1

PRISMA Diagram
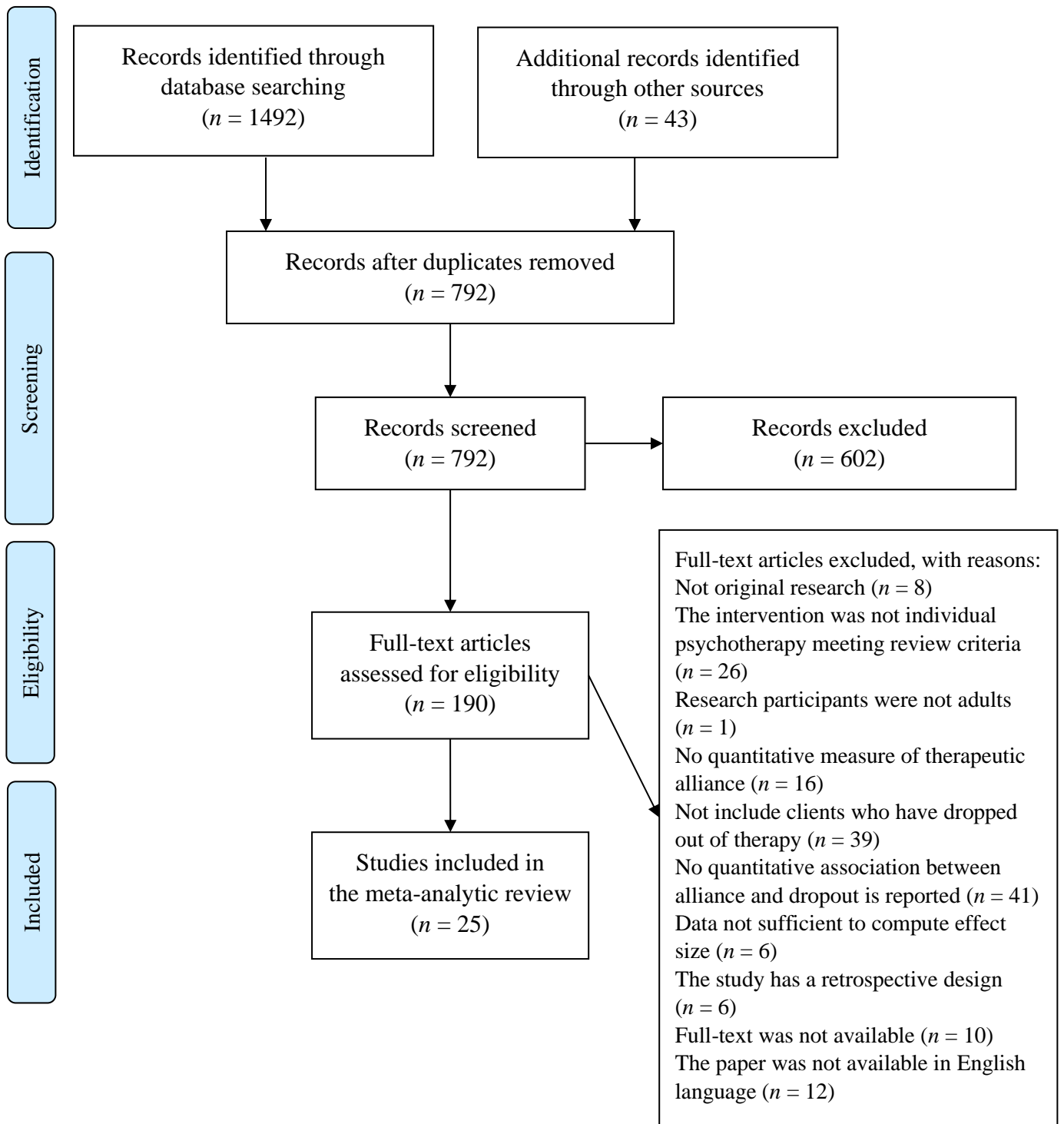
Table 1

Inclusion and Exclusion Criteria for Review Eligibility

Criterion	Inclusion	Exclusion
Original Research	The research must present original data.	Meta-analyses, systematic reviews, study protocols and book chapters were excluded. Case studies and case series were also excluded.
Adult Clients	The research participants were adults (aged 18 years and above).	Studies of child and adolescent clients were excluded from the current review.
Individual Psychotherapy Intervention	The research must assess the outcome of an individual psychotherapy intervention. All psychological and counselling interventions were considered in the current review regardless of theoretical orientation (e.g., cognitive behavioural therapy, psychodynamic psychotherapy, person-centred therapy, etc.).	To minimise the heterogeneity of study characteristics, the following studies were excluded from the current meta-analysis: a) Studies limited exclusively to drug and alcohol treatment. b) Studies limited exclusively to couple, family and/or group psychotherapy. c) Studies limited exclusively to telephone or internet psychotherapy. d) Studies limited exclusively to psychotherapy treatment for offending or violent behaviour. e) Studies limited exclusively to inpatient treatment. f) Studies which use analogue subjects (e.g., simulated psychotherapy patients) rather than genuine psychotherapy clients. g) In studies were there are multiple treatment conditions, and one or more of the treatments is not individual psychotherapy (e.g., pharmacotherapy treatment alone), studies will only be included if the alliance dropout association is reported for the individual psychotherapy condition separately.

Criterion	Inclusion	Exclusion
Measure of Therapeutic Alliance	The research includes a prospective measure of therapeutic alliance. All quantitative measures of therapeutic alliance were considered eligible for the current review.	A quantitative measure of therapeutic alliance is not available. Studies with survey designs which rely on clients and/or therapists retrospectively rating therapeutic alliance and dropout/attrition were excluded due to the influence of recall bias.
Assess Premature Termination	The study included participants who have prematurely terminated psychotherapy. All operational definitions of dropout were considered.	The study did not examine or report on clients who had dropped out of psychotherapy.
Reported an Association between Therapeutic Alliance and Dropout	Studies included in the review must present a quantitative measure of association between the therapeutic alliance and premature termination. The data reported must be sufficient to extract or estimate a value indicating the relation between the two primary review variables.	There was no quantitative association reported between alliance and dropout, or otherwise the data was not sufficient to calculate an effect size.
English Language	The paper was available in English language.	The research paper was not available in English language.
Full-Text Availability	The full-text research paper must be available.	The full-text research paper was not available.

Study definition of dropout was assigned a code based on the operationalisation of premature termination described by Swift and Greenberg (2012), specifically: 1) attending less than a specified number of sessions; 2) failure to complete a treatment protocol; 3) missing a scheduled appointment without rescheduling or attending future appointments; 4) termination after the initial intake session; 5) therapist judgement of dropout status; 6) discontinuance of treatment prior to reliable and/or clinically significant improvement in symptoms and 7) other. Dropout rate was defined as the percentage of clients who started an intervention who prematurely terminated. This variable was only collated for studies in which the dropout rate was representative of the full sample starting treatment, and not when clients had been selected based on dropout status *post hoc*.

Therapeutic alliance variables were also extracted, including the measure of therapeutic alliance used, the rater of therapeutic alliance (coded as: client, therapist, and observer) and the timepoint (session number) when therapeutic alliance was rated. The early phase of treatment is defined in the alliance literature as the first five or six sessions (Flückiger et al., 2018; Muran et al., 2009). Therefore, timepoint in which the therapeutic alliance was measured was coded as early (rating in the first six sessions), late (therapeutic alliance rated after session six), and average (an average of therapeutic alliance ratings taken across treatment).

Data was extracted to compute an effect size. For each paper, descriptive (e.g., means, standard deviations and sample sizes) or inferential statistics (e.g., *d*, z, *t, F, r*, Odds ratio) describing the association between therapeutic alliance and dropout were identified. In instances when the data reported was not sufficient to compute a comparable effect size, the corresponding author of the paper was contacted and asked to provide further data. Papers where an effect size could not be calculated, and further data was not provided, were excluded from the review. A summary of the reasons for exclusion of such papers is included in Appendix A.

Treatment variables extracted included treatment length (number of sessions/weeks of treatment), time limitation of treatment, orientation of psychotherapy, and treatment setting. Time

limitation of treatment was coded as none (treatment duration was not specified or was not restricted), low (treatment duration of less than 20 sessions) and high (treatment duration of 20 sessions and above). Treatment orientation was coded as cognitive-behavioural (inclusive of cognitive, behavioural, and cognitive-behavioural therapies), counselling, integrative, psychodynamic, solution-focused, supportive/client-centred, and other. A "mixed" code was also added for studies where multiple psychotherapy orientations were included, but the analyses were completed on the full sample.

Five client variables were extracted and coded. Client presenting problem was categorised as depressive disorder, anxiety disorder, personality disorder, eating disorder, post-traumatic stress disorder, other or mixed (when clients with a range of presenting problems received psychotherapy within the same study). Client age, gender, and ethnicity was also extracted when reported. Client education status was coded as the percentage of clients with a higher-level qualification (e.g. university, college or equivalent qualification).

Finally, data was extracted from identified papers regarding the study characteristics. Study author(s), publication status, year of publication/completion, country in which the study was completed, and study sample size were extracted.

Quality Assessment

Quality appraisal was conducted to assess the research quality of the field, and not to exclude studies from the current review. Establishing methodological quality was especially important given the inclusion of non-published work in the current review. Methodological quality was evaluated using an adapted version of the Quality In Prognosis Studies (QUIPS) tool (Hayden et al., 2013). The tool assesses quality across six bias domains: study participation, study attrition, prognostic factor measurement, outcome measurement, study confounding, and statistical analysis and reporting. Each subscale is assigned a quality rating of low, moderate or high risk of bias. Two additional domains were added to the quality assessment tool, rating the risk of bias arising from

18

study design and intervention integrity, amended from the Effective Public Health Practice Project (EPHPP) quality assessment tool (Thomas et al., 2004). The full adapted version of the quality assessment tool used in the current review can be found in Appendix B. Study protocol or original trial papers were consulted for information regarding the quality assessment process, as needed.

The quality assessment process was completed by the author. Additionally, a random sample (20%, $k = 5$) of studies were reviewed by a second rater (AF). Disagreements in the ratings of quality ascribed to each study were resolved through discussion until a position of consonance was reached. A two-way mixed effects, average-rater absolute agreement intraclass correlation coefficient was calculated to assess inter-rater reliability for quality assessment (Koo & Li, 2016).

Data Analysis

Meta-Analysis of Effect Sizes

A random-effects estimator was utilised for the meta-analysis, based on the assumption that the studies included in the synthesis represent a random sample of the total population of studies (Borenstein et al., 2009). The analyses were conducted using Comprehensive Meta-Analysis (CMA; 2014; Version 3.3) developed by Borenstein et al. (2013). Raw data from the study was converted to a comparable effect size statistic (*d*; standardised difference between two means). Effect sizes were assigned a positive value if they were consistent with *a priori* predictions (i.e., lower levels of therapeutic alliance are associated with increased rates of premature termination). Independent groups of clients within a study (e.g., treatment groups, clients seen within different clinic settings), for which the alliance-dropout effect is reported separately, were treated as subgroups. When a study reported multiple measures or raters of therapeutic alliance, the data was modelled as multiple outcomes within a study, and each of the associations with dropout were combined into a weighted mean effect size. Individual study effect sizes were then combined into a weighted summary effect statistic. The magnitude of the effect size was interpreted based on the parameters ($d = .20$, small; $d = .50$, medium; $d = .80$, large) described by Cohen (1992).

Heterogeneity in study findings was assessed using Q and I^2 statistics. A significant Q-value indicates heterogeneity in the association between therapeutic alliance and premature termination between studies, suggesting that moderator analysis is indicated. The I^2 statistic was also computed, illustrating the percentage of true heterogeneity between studies relative to the total variation in observed effects (Borenstein et al., 2009). The I^2 statistic can be loosely interpreted as: 0-40% might not be important, 30-60% may represent moderate heterogeneity, 50-90% may represent substantial heterogeneity, 75-100% considerable heterogeneity (Ryan, 2016).

Moderator Analyses

A mixed-method approach was used to conduct moderator analyses. Categorical variables (psychotherapy orientation, treatment length, client diagnosis, rater of therapeutic alliance, timepoint of alliance rating, dropout definition and publication status) were tested as moderators of the alliance-dropout association. Three or more studies per subgroup were required in order to conduct categorical moderator analyses (Card, 2012). Studies within each subgroup of categorical moderator variables were aggregated using a random-effects model, assuming a common among-study variance component across subgroups, and then subgroups were compared using a Q-test (Borenstein et al., 2009). In subgroup analyses, a significant Q-value indicates a difference between groups in the reported alliance-dropout association. As reported in Sharf et al. (2010), the levels of the alliance rater and alliance timepoint moderator variables were considered independent in subgroup analyses. A meta-regression analysis was then conducted comparing the alliance-dropout association across the continuous moderator variable (client education). Due to the variance between studies, a random-effects model was used (method of moments). Given the number of variables included in the moderator analyses ($n = 8$), a Bonferroni-corrected p-value of 0.0063 was used.

Publication Bias

Although grey literature was included in the current review, publication bias was still assessed. Firstly, a funnel plot was used to visually inspect the relationship between study size (standard error) and the effect size. Publication bias is indicated when there is asymmetry in the pattern of study effect sizes around the summary effect size (Borenstein et al., 2009). Egger's regression test was used to quantify the relationship between sample size and effect size (Egger et al., 1997), with a significant *p*-value signifying asymmetry of the funnel plot, potentially indicative of publication bias. The trim and fill method was used to correct for bias arising from missing studies in the review, using an iterative procedure to remove the most extreme small studies, and recomputing the effect size at each iteration until the funnel plot is symmetric around the new unbiased effect size (Borenstein et al., 2009; Duval & Tweedie, 2000). As there is significant variability in the results of trim and fill models applied, both a fixed-random effects trim and fill model (where a fixed effects model is used to trim and fill the meta-analysis and a random effects model is used to calculate the adjusted pooled estimate) and a random-random effect trim and fill model were considered to adjust for publication bias (Peters et al., 2007).

Results

Twenty-five studies were included in the meta-analytic review, comprising of data from 3177 individual therapy clients, ranging from 16 to 524 per study, across nine countries. The sample of psychotherapy clients within the review had a weighted mean age of 37.36 years and were 66.35% female. Table 2 presents a summary of the study characteristics and Table 3 displays an overview of the individual study findings.

Meta-Analysis of Effect Sizes

A small-medium positive effect size was observed between therapeutic alliance and premature termination from psychotherapy ($d = 0.443$; 95% CIs: $[0.295, 0.591]$; $Z = 5.86$, $p < .001$). The findings provide support for H1 and indicate that lower levels of therapeutic alliance are associated with increased incidence of dropout. Effect sizes ranged from $d = -0.177$ to 1.786 (Figure 2). As predicted, there was evidence of significant heterogeneity between studies with about 75% of the observed variance reflecting real-world differences in effect size ($Q(25) = 104.28$, $p < .001$; $I^2 = 76.98$, $T^2 = 0.08$). As there was significant heterogeneity among study effect sizes, moderator analyses were conducted to examine sources of variance.

Moderator Analyses

Psychotherapy Orientation

The majority of studies within the review assessed the alliance-dropout association within cognitive behavioural therapy ($k = 10$) or across a mix of psychotherapy treatments ($k = 12$). Three additional studies reported effect sizes for a singular treatment: transference focused therapy (Wasserman, 2011), supportive therapy (Wasserman, 2011), integrative psychotherapy (Kegel & Flückiger, 2015) and counselling (Mahon et al., 2015). Due to the small number of studies in each of these therapy orientation groups ($k < 3$), the studies could not be included in the subgroup analysis. As expected, there were no significant differences between the study effect sizes based on treatment with a cognitive behavioural therapy ($d = 0.502$), or multiple therapies within the same

22

Table 2

Study Characteristics

Author(s)	Paper Type	Country	Sample Size	Client Presenting Problem	Psychotherapy Orientation	Therapy Length	Dropout Rate (%)	Dropout Category	Alliance Measure	Alliance Rater	Time Alliance Rated
Al-Jabari (2015)	Dissertation	USA	524	Mixed	Mixed	No limitation	63.2	5	WAI short revised)	C	Early
Anderson (2010)	Dissertation	USA	31	Low self-esteem	Cognitive modification or supportive therapy	Low - 6 sessions	N/A	2	WAI	O	Early
Arnow et al. (2007)	Journal Article	USA	451	Depression	Cognitive behavioural analysis system of psychotherapy (and combined pharmacotherapy)	High - 24 sessions	22.0	2	WAI (short)	C	Early
Cooper et al. (2016)	Journal Article	USA	176	Depression	Cognitive therapy	High - until remission of symptoms	17.0	6	WAI (short)	O	Early
Doran et al. (2017)	Journal Article	USA	47	Mixed	CBT, brief relational therapy or a combination of both orientations	High - 30 sessions	29.8	N/S	WAI (short)	C	Average

Table 2

Study Characteristics

Author(s)	Paper Type	Country	Sample Size	Client Presenting Problem	Psychotherapy Orientation	Therapy Length	Dropout Rate (%)	Dropout Category	Alliance Measure	Alliance Rater	Time Alliance Rated
Elkin et al. (2014)	Journal Article	USA	72	Depression	CBT and interpersonal psychotherapy	High - 20 sessions	31	1	VTAS (patient factor)	O	Early
Eubanks et al. (2019)	Journal Article	USA	42	Mixed	CBT	High - 30 sessions	N/A	2	WAI (short)	C & T	Early
Gibbons et al. (2019)	Journal Article	USA	237	Depression	Cognitive therapy and short-term dynamic psychotherapy	Low - 16 sessions	79.0	1	WAI	C	Early
Haug et al. (2016)	Journal Article	Norway	65/57	Social anxiety and panic disorder	CBT	Low - 12 sessions	20.7	2	WAI (short)	C	Early and Late
Howard (2017)	Dissertation	USA	36	Mixed	CBT	No limitation	38.8	3	CALPAS	C & T	Early
Huang et al. (2013)	Journal Article	USA	16	Mixed	Psychodynamic or interpersonal psychotherapy	No limitation	N/A	4	WAI (short revised)	C	Early

Table 2

Study Characteristics

Author(s)	Paper Type	Country	Sample Size	Client Presenting Problem	Psychotherapy Orientation	Therapy Length	Dropout Rate (%)	Dropout Category	Alliance Measure	Alliance Rater	Time Alliance Rated
Huppert et al. (2014)	Journal Article	USA	133	Panic disorder with agoraphobia	CBT	Low - 11-20 sessions	27.1	N/A	WAI (short)	C & T	Early
Jordan et al. (2017)	Journal Article	New Zeal and	56	Anorexia nervosa	Specialist supportive clinical management, CBT, or interpersonal therapy.	High - 20 sessions	37.5	1	VTAS (revised)	O	Early
Keefe et al. (2020)	Journal Article	USA	185	Panic disorder and agoraphobia	CBT, panic-focused psychodynamic therapy, or applied relaxation	High - 24 sessions	26.9	1	WAI	C	Early
Kegel and Flückiger (2015)	Journal Article	Swit zerla nd	296	Mixed	Integrative	No limitation	57.1	7	Global Alliance element of BPSR	C	Average
Mahon et al. (2015)	Journal Article	USA	124	Mixed	Counselling	No limitation	70.2	1	WAI (short)	C	Early

Table 2

Study Characteristics

Author(s)	Paper Type	Country	Sample Size	Client Presenting Problem	Psychotherapy Orientation	Therapy Length	Dropout Rate (%)	Dropout Category	Alliance Measure	Alliance Rater	Time Alliance Rated
McEvoy et al. (2014)	Journal Article	Australia	84	Mixed	CBT	No limitation	26.2	5	HAQ-II	C	Early
Muran et al. (2009)	Journal Article	USA	128	Personality disorder (cluster C or not specified)	Short-term dynamic therapy, CBT or brief relational therapy	High - 30 sessions	34.4	2	WAI (short)	C & T	Early
Prom et al. (2014)	Journal Article	Peru	60	Mixed	Mixed	No limitation	42.0	4	WAI (Spanish version)	C & T	Early
Saatsi et al. (2007)	Journal Article	UK	97	Depression	Cognitive therapy	Low - 12 sessions	24.7	2	CALPAS or ARM	C	Average
Samstag et al. (2008)	Journal Article	USA	48	Mixed	Dynamic psychotherapy, CBT, supportive or relational therapy.	High - 30 sessions	N/A	1	WAI (short)	C & T	Average
Santos (2016)	Dissertation	USA	36	Depression	BA for Latinos or therapy as usual	Low - 12 sessions	N/A	1	WAI (short)	O	Early

Table 2

Study Characteristics

Author(s)	Paper Type	Country	Sample Size	Client Presenting Problem	Psychotherapy Orientation	Therapy Length	Dropout Rate (%)	Dropout Category	Alliance Measure	Alliance Rater	Time Alliance Rated
Sijercic (2018)	Dissertation	Canada	112	PTSD	Cognitive processing therapy	Low - 12 sessions	25.9	2	WAI (short)	O	Early
Spinhoven et al. (2007)	Journal Article	Netherlands	62	BPD	Schema focused therapy and transference focused psychotherapy	High - maximum 3 years	5.8	2	WAI	T	Late
Wasserman (2011)	Dissertation	USA	59	BPD	Transference focused psychotherapy, dialectical behaviour therapy and supportive psychotherapy.	High - 50-100 sessions	N/A	N/A	WAI	O	Average

Note: US = United States of America. UK = United Kingdom. PTSD = post-traumatic stress disorder. BPD = borderline personality disorder. CBT = cognitive behavioural therapy. BA = behavioural activation. N/A = not applicable (e.g., dropout rates reported in a study were not a true representation of dropout from the treatment). Dropout category: 1 = attending less than a specified number of sessions; 2 = failure to complete a treatment protocol; 3 = missing a scheduled appointment without rescheduling or attending future appointments; 4 = termination or non-return after the initial intake sessions; 5 = therapist judgement of dropout status; 6 = discontinuance of treatment prior to reliable and/or clinically significant improvement in symptoms; 7 = other. WAI = working alliance inventory, VTAS = Vanderbilt therapeutic alliance scale, BPSR = Bern Post Session Report, HAQ-II = helping alliance questionnaire (version 2). CALPAS = California psychotherapy alliance scales. ARM = Agnew relationship measure. Alliance rater: C = client, T = therapist, O = observer.

Table 3

Summary of Findings

Author(s)	Sample Description	Summary of Findings	Effect Size (d)	Aggregated Effect Size (d)
Al-Jabari (2015)	Clients attending a university psychology training clinic receiving individual psychotherapy from trainee therapists.	A multinominal logistic regression showed the first completed client-rated alliance score, controlling for the session number this rating was completed, predicted dropout status (Wald statistic (1) = 7.35, $\beta = 0.05$, $p < 0.01$; $OR = 1.05$, 95% CIs [1.00-1.11])	0.027	
Anderson (2010)	University students with low self-esteem receiving six sessions of cognitive modification or supportive therapy.	No significant differences in observer-rated alliance scores in the last client session between completers ($M = 206.39$; $SD = 14.28$, $n = 14$) and dropouts ($M = 199.26$, $SD = 24.96$, $n = 17$; $F = .828$, $p = .37$).	0.342	
Arnow et al. (2007)	Clients with major depressive disorder randomly assigned to either nefazadone alone, cognitive behavioural analysis system of psychotherapy alone, or both treatments. The review only included the sample who received psychotherapy.	Dropouts had lower client-rated alliance total scores ($M = 63$; $SD = 13$, $n = 99$) compared to completers ($M = 67$, $SD = 12$, $n = 352$).	0.327	
Cooper et al. (2016)	Clients with major depressive disorder randomly assigned to cognitive therapy combined with anti-depressant medication or medication alone.	A survival analysis model found that early observer-rated alliance scores, controlling for client age and marital status, were predictive of dropout (Estimate = - .59, $SE = .21$, $HR = .55$, 95% CI: [.36, .84], $p = .004$). The effect size was extracted from the meta-analysis completed by Flückiger et al. (2018).	1.010	

Table 3

Summary of Findings

Author(s)	Sample Description	Summary of Findings	Effect Size (*d*)	Aggregated Effect Size (*d*)
Doran et al. (2017)	Clients receiving 30 sessions of cognitive behavioural therapy, brief relational therapy, or a combined treatment.	Dropouts had a trend-level difference in client-rated alliance total scores compared to completers t (43) = 19.8, p = .06. Descriptive statistics were not reported by the author.	0.632	
Elkin et al. (2014)	Clients taking part in a multi-site collaborative study of two psychotherapies, cognitive behavioural therapy and interpersonal psychotherapy for major depressive disorder.	A point-biserial correlation showed a non-significant negative association between alliance scores at session 3 and premature termination (r_{pb} = - .22, p > .05, n = 67).	0.451	
Eubanks et al. (2019)	Clients receiving 30 sessions of outpatient cognitive behavioural therapy.	Dropouts had lower client-rated alliance total scores (M = 5.23, SD = 1.22, n = 14) compared to completers (M = 5.97, SD = .70, n = 22; t (18.48) = 2.08, p = .052). Dropouts also had lower therapist-rated alliance scores (M = 4.64, SD = 1.05, n = 14) compared to completers (M = 5.43, SD = .73, n = 23; t (35) = 2.67, p = .01).	0.793 0.916	0.854
Gibbons et al. (2019)	Clients with depression randomised to 16 sessions of cognitive therapy or short-term dynamic therapy.	A multinominal logistic regression found that client-reported alliance (agreement of tasks domain only) at session 2 predicted dropout from treatment after 2 - 6 sessions (Wald statistic (1) = 3.25, p = .071; OR = 1.597, 95% CI: [.960, 2.656]). Client-reported alliance (agreement of tasks domain only) at session 2 also predicted dropout from treatment after 7 - 11 sessions (Wald statistic (1) = 9.45, p = .002; OR = 3.202, 95% CI: [1.525, 6.724]).	0.258 0.642	0.381

Table 3

Summary of Findings

Author(s)	Sample Description	Summary of Findings	Effect Size (*d*)	Aggregated Effect Size (*d*)
Haug et al. (2016)	Clients receiving individual cognitive behavioural therapy for social anxiety disorder or panic disorder in a multicentre trial.	Zero-order correlations showed a non-significant positive association between client-rated alliance scores at session 3 and premature termination ($r = .03$, $p > .05$, $n = 65$). Zero-order correlations showed a significant negative association between client-rated alliance scores at session 8 and premature termination ($r = -.35$, $p < .01$, $n = 57$).	-0.06 0.747	0.344
Howard (2017)	Clients presenting to two university training clinics providing cognitive behavioural therapy. Data from the Psychological Services Centre (PSC) and Anxiety and Stress Disorders Clinic (ASDC) were included in the review. Data from the clinic providing DBT treatment, which included a group skills component, is not included in the computed effect size.	Unable to calculate an effect size from the binary logistic regression as it included the DBT condition. Descriptive data was provided by the author. In the PSC clinic, dropouts had lower client-rated alliance scores ($M = 5.64$, $SD = 0.610$, $n = 5$) compared to completers ($M = 6.03$, $SD = .0.354$, $n = 9$). In the ASDC clinic, dropouts also had lower client-rated alliance scores ($M = 5.55$, $SD = 1.029$, $n = 9$) compared to completers ($M = 6.30$, $SD = 0.336$, $n = 13$). In the PSC clinic, dropouts also had lower therapist-rated alliance scores ($M = 4.98$, $SD = 0.484$, $n = 5$) compared to completers ($M = 5.32$, $SD = 0.663$, $n = 9$). In the ASDC clinic, dropouts also had lower therapist-rated alliance scores ($M = 5.05$, $SD = 0.650$, $n = 9$) compared to completers ($M = 5.27$, $SD = 0.588$, $n = 13$).	0.963 0.458	0.712
Huang et al. (2013)	Client attending a community outpatient psychology department clinic receiving psychodynamic or interpersonal therapy.	Dropouts had non-significantly higher client-rated alliance total scores ($M = 3.39$, $SD = 1.11$, $n = 8$) compared to completers ($M = 3.23$, $SD = 0.57$, $n = 7$; $F(1,10) = 1.04$, $p = .36$).	-0.177	

Table 3

Summary of Findings

Author(s)	Sample Description	Summary of Findings	Effect Size (*d*)	Aggregated Effect Size (*d*)
Huppert et al. (2014)	Clients taking part in a multicentre trial, receiving CBT treatment for panic disorder with agoraphobia.	A multi-level model supports that both client- and therapist-rated alliance scores at session 3 are predictive of the number of sessions to dropout. Additional descriptive statistics were provided by the authors to compute an effect size. Dropouts had lower client-rated alliance (*M* = 69.92, *SD* = 12.30, *n* = 39) than completers (*M* = 74.60, *SD* = 8.40, *n* = 106).	0.488	0.523
		Dropouts also had lower therapist-rated alliance (*M* = 66.15, *SD* = 8.36, *n* = 39) than completers (*M* = 70.53, *SD* = 7.67, *n* = 106).	0.557	
Jordan et al. (2017)	Clients taking part in a randomised trial of specialist supportive clinical management, cognitive behavioural therapy, and interpersonal therapy for anorexia nervosa.	Dropouts had non-significantly lower observer-rated alliance (*M* = 50.00, *SD* = 4.78, *n* = 19) compared to completers (*M* = 52.03, *SD* = 4.56, *n* = 31; *t* = 1.52, *p* = .14).	0.440	
Keefe et al. (2020)	Clients taking part in a randomised trial of cognitive behavioural therapy, panic-focused psychodynamic therapy, or applied relaxation treatment for panic disorder with/without agoraphobia.	A cox proportional hazards model, including client expectancies, showed that early alliance was not a significant predictor of subsequent dropout (β = -0.04, *HR* = 0.96, *Z* = -0.16, *p* = 0.873). Descriptive data provided by authors to calculate an effect size. Dropouts had lower client-rated alliance average scores (*M* = 5.34; *SD* = 0.99, *n* = 39), compared to completers (*M* = 5.59, *SD* = 0.80, *n* = 146).	0.297	
Kegel and Flückiger (2015)	Clients receiving integrative psychotherapy from a private outpatient clinic.	Hierarchical linear modelling showed that client rated global alliance scores were lower for dropouts, compared with completers, over the course of treatment (Cohen's *d* for intercept = - 0.31).	0.310	

Table 3

Summary of Findings

Author(s)	Sample Description	Summary of Findings	Effect Size (*d*)	Aggregated Effect Size (*d*)
Mahon et al. (2015)	University students receiving short-term individual counselling at a university counselling centre.	Dropouts had significantly lower client-rated alliance (*M* = 59.49; *SD* = 6.79, *n* = 87), compared to completers (*M* = 61.92, *SD* = 14.07, *n* = 37; *t* (122) = 2.22, *p* = .03)	0.255	
McEvoy et al. (2014)	Clients receiving group or individual cognitive behavioural therapy for unipolar depressive disorder or anxiety disorder. Only clients in the individual therapy treatment group were included in the review analyses.	Dropouts had significantly lower client-rated alliance (*M* = 94.41; *SD* = 10.22, *n* = 22), compared to completers (*M* = 99.73, *SD* = 8.97, *n* = 62; *t* (82) = 2.30, *p* < .05)	0.572	
Muran et al. (2009)	Clients randomly assigned to 30 sessions of cognitive behavioural therapy, brief relational therapy, or short-term dynamic psychotherapy for personality disorders.	Pearson correlations (with dropout status treated as a continuous variable ranging from 1 to 2) showed a significant negative association between client-rated alliance scores (*r* = -.03, *p* < .01, *n* = 128) and premature termination. A significant negative association was also reported between therapist-rated alliance scores (*r* = - .27, *p* < .01, *n* = 128) and premature termination.	0.629 0.561	0.595
Prom et al. (2014)	Clients receiving outpatient therapy at a public mental health clinic in rural Peru.	Dropouts had lower client-rated alliance total scores (*M* = 206.2, *SD* = 34.5, *n* = 26) compared to completers (*M* = 211.1, *SD* = 25.2, *n* = 31). Dropouts also had lower therapist-rated alliance scores (*M* = 202.0, *SD* = 17.6, *n* = 25) compared to completers (*M* = 214.2, *SD* = 27.3, *n* = 33). Descriptive statistics not reported in the original paper, provided by the author.	0.164 0.516	0.340

Table 3

Summary of Findings

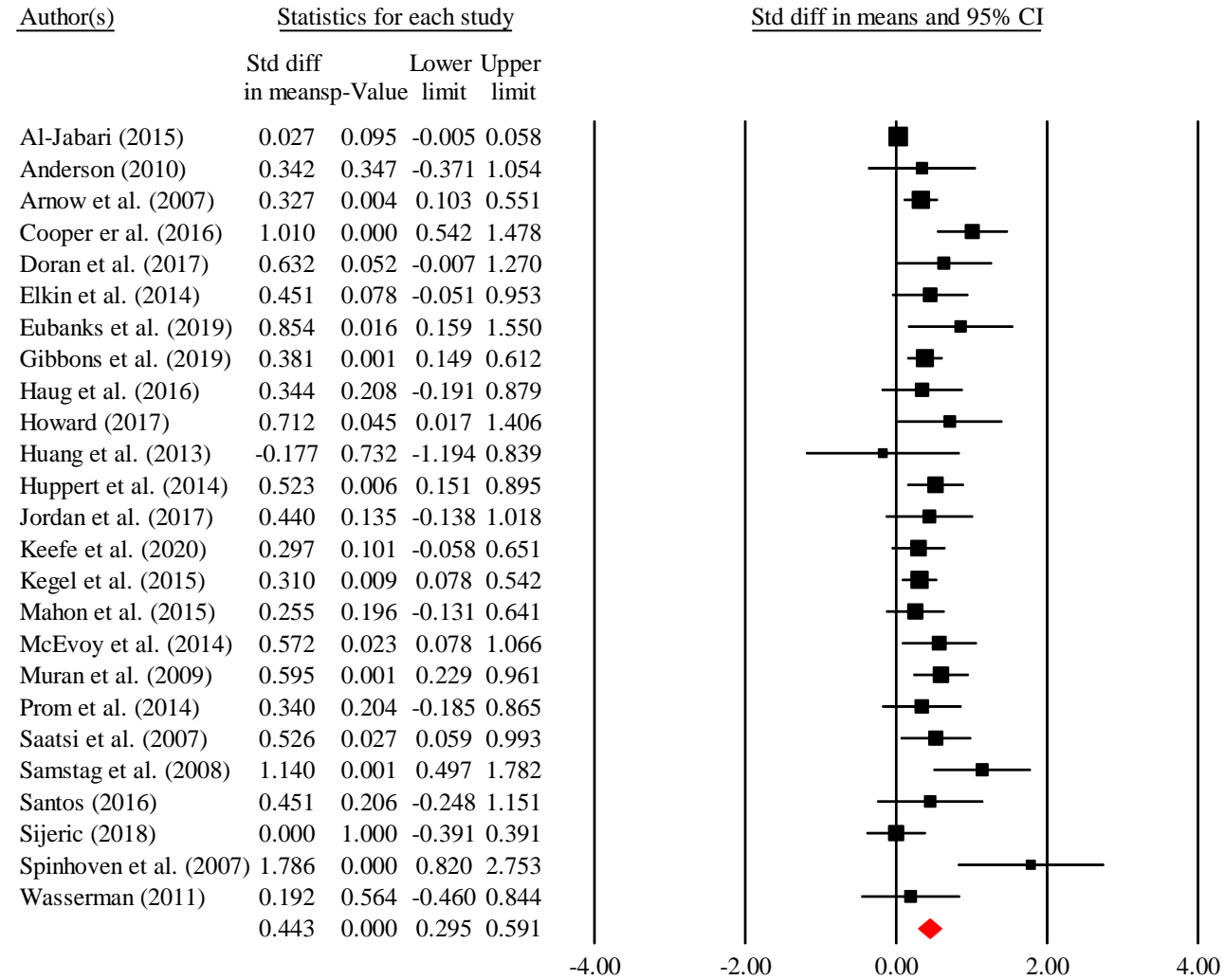| Author(s)	Sample Description	Summary of Findings	Effect Size (d)	Aggregated Effect Size (d)
Saatsi et al. (2007)	Clients receiving a minimum of 12 sessions (and up to 20 sessions) of cognitive therapy for depressive disorder.	Alliance scores were converted to z-scores for comparability. Dropouts had lower average client-rated alliance scores across sessions ($M = -0.37$; $SD = 1.00$, $n = 24$), compared to completers ($M = 0.14$, $SD = 0.96$, $n = 73$; $t(102) = 2.36$, $p < .05$).	0.526	
Samstag et al. (2008)	Clients randomly assigned to 30 sessions of dynamic psychotherapy, cognitive behavioural, supportive, or relational therapy.	Completers were allocated to two groups: good outcome and poor outcome. Dropouts had lower average client-rated alliance scores across sessions ($M = 4.19$; $SD = 0.69$, $n = 16$), compared to completers with a good outcome ($M = 5.22$, $SD = 0.80$, $n = 16$) and poor outcome ($M = 5.15$, $SD = 0.84$, $n = 16$). Dropouts also had lower average therapist-rated alliance scores across sessions ($M = 4.04$; $SD = 0.95$, $n = 16$), compared to completers with a good outcome ($M = 4.83$, $SD = 0.56$, $n = 16$) and poor outcome ($M = 4.80$, $SD = 0.81$, $n = 16$).	1.287 0.992	1.140
Santos (2016)	Clients receiving behavioural activation for Latinos or treatment as usual individual therapy for depression	A point-biserial correlation showed a non-significant negative association between alliance scores at session 2 and premature termination ($r_{pb} = -.22$, $p > .05$, $n = 36$).	0.451	
Sijercic (2018)	Clients receiving 12 sessions of cognitive processing therapy for post-traumatic stress disorder.	Dropouts had equivalent observer-rated alliance ($M = 4.89$; $SD = .74$, $n = 38$), compared to completers ($M = 4.89$, $SD = 0.77$, $n = 74$).	0.000	

Table 3

Summary of Findings

Author(s)	Sample Description	Summary of Findings	Effect Size (*d*)	Aggregated Effect Size (*d*)
Spinhoven et al. (2007)	Clients randomised to receive schema-focused therapy or transference focused psychotherapy for borderline personality disorder.	For those dropping out in the first three months of treatment, dropouts had lower therapist-rated alliance (*M* = 8.4, *SD* = 1.1, *n* = 5) compared to completers (*M* = 10.7, *SD* = 1.3, *n* = 57; *t* (60) = 3.67, *p* < .01). Cox regression analyses with pre-treatment borderline personality disorder severity index, treatment condition and therapeutic alliance revealed that client-rated alliance (Wald = 4.379, *p* < .05, *HR* = 0.775, 95% CIs [0.610,0.984]) and therapist-rated alliance (Wald = 8.171, *p* < .01, HR = 0.551, 95% CIs [0.367,0.984]) was predictive of time to dropout after three months of therapy.	1.786	
Wasserman (2011)	Clients diagnosed with a borderline personality disorder randomised to transference focused psychotherapy (TFP), dialectical behavioural therapy (DBT), and supportive psychotherapy (SPT). The TFP and SPT groups were only included in analysis due to group component of DBT.	Across all three treatments (including DBT), a cox proportional hazard regression model showed alliance had a significant negative relationship to early termination (*β* = -.835, *SE* = 0.386, *df* = 1, *p* = .03). To separate the individual treatment effect, the observed dropout frequencies for TFP and SPT were used, with alliance stratified into low and high groups by a mean split. In TFP, 25% of patients dropped out when alliance was low, while 21.4% dropped out when alliance was high. In SPT, 33.3% of patients dropped out when alliance was low, while 23.5% dropped out when alliance was high.	0.111 0.268	0.192

Note: *M* = mean, *SD* = standard deviation, *n* = sample size, *df* = degrees of freedom, *SE* = standard error, *HR* = hazard ratio, *CI* = confidence interval, *OR* = odds ratio

Figure 2

Forest Plot for the Therapeutic Alliance and Premature Termination Meta-Analysis

Author(s)	Statistics for each study				Std diff in means and 95% CI
	Std diff in means	p-Value	Lower limit	Upper limit	
Al-Jabari (2015)	0.027	0.095	-0.005	0.058	
Anderson (2010)	0.342	0.347	-0.371	1.054	
Arnow et al. (2007)	0.327	0.004	0.103	0.551	
Cooper er al. (2016)	1.010	0.000	0.542	1.478	
Doran et al. (2017)	0.632	0.052	-0.007	1.270	
Elkin et al. (2014)	0.451	0.078	-0.051	0.953	
Eubanks et al. (2019)	0.854	0.016	0.159	1.550	
Gibbons et al. (2019)	0.381	0.001	0.149	0.612	
Haug et al. (2016)	0.344	0.208	-0.191	0.879	
Howard (2017)	0.712	0.045	0.017	1.406	
Huang et al. (2013)	-0.177	0.732	-1.194	0.839	
Huppert et al. (2014)	0.523	0.006	0.151	0.895	
Jordan et al. (2017)	0.440	0.135	-0.138	1.018	
Keefe et al. (2020)	0.297	0.101	-0.058	0.651	
Kegel et al. (2015)	0.310	0.009	0.078	0.542	
Mahon et al. (2015)	0.255	0.196	-0.131	0.641	
McEvoy et al. (2014)	0.572	0.023	0.078	1.066	
Muran et al. (2009)	0.595	0.001	0.229	0.961	
Prom et al. (2014)	0.340	0.204	-0.185	0.865	
Saatsi et al. (2007)	0.526	0.027	0.059	0.993	
Samstag et al. (2008)	1.140	0.001	0.497	1.782	
Santos (2016)	0.451	0.206	-0.248	1.151	
Sijeric (2018)	0.000	1.000	-0.391	0.391	
Spinhoven et al. (2007)	1.786	0.000	0.820	2.753	
Wasserman (2011)	0.192	0.564	-0.460	0.844	
	0.443	0.000	0.295	0.591	

study ($d = 0.446$; $Q(1) = 0.11$; $p = .735$). The finding provides partial support for H2.

Treatment Length

The studies within the review had variation in treatment length/limitation; with 11 reporting longer treatment durations ($d = 0.586$), seven with short treatment durations ($d = 0.360$), and seven studies in which the number of psychotherapy sessions was not limited ($d = 0.256$). There was no significant variation in the association between therapeutic alliance and premature termination between studies of different treatment lengths/limitations ($Q(2) = 5.80$; $p = .055$), which provides evidence which does not support H3.

Client Diagnosis or Presenting Problem

Studies included in the review provided psychotherapy for a range of diagnoses/presenting problems, including anxiety disorders ($k = 3$), depression ($k = 6$), personality disorder ($k = 3$), eating disorders ($k = 1$), PTSD ($k = 1$) and other presenting problems ($k = 1$; self-esteem). Ten studies in the review examined the alliance-dropout association in clients with a range of presenting problems. A subgroup analysis comparing the effect sizes between studies of anxiety ($d = 0.391$), depression ($d = 0.500$), personality disorder ($d = 0.676$) and mixed presentations ($d = 0.405$) found no significant diagnosis-related differences in the effect sizes between alliance and dropout ($Q(3) = 1.37$, $p = .713$), therefore not supporting H4.

Client Education

Eleven studies in the review reported the level of client education (% completing higher level qualifications). A method of moments meta-regression analysis found that patient education was not a significant predictor of the alliance outcome association ($Q(1) = 0.92$, $\beta = 0.01$, $p = .338$, 95% CI: [-0.25, - 0.77]), contrary to H5.

Rater of Therapeutic Alliance

Raters of the alliance varied between studies, including client ($k = 17$), therapist ($k = 7$) and observer ($k = 7$). There was non-significant variation in the effect sizes between client- ($d = 0.421$), therapist- ($d = 0.713$), and observer-rated alliance ($d = 0.407$; $Q(2) = 3.33$, $p = .189$) and premature termination, contrary to H6.

Timepoint of Alliance Rating

The majority of studies ($k = 19$) rated alliance early in psychotherapy (within the first six sessions). Two studies rated alliance after session six, and six studies took an average alliance rating across the course of treatment. Comparison with studies rating alliance late in therapy could not be completed because of insufficient number of studies ($k = 2$). There was no significant variation between studies which rated alliance early ($d = 0.419$) or as an average across therapy ($d = 0.588$; $Q(1) = 0.98$, $p = .323$). Therefore, H7 is not supported.

Dropout Definition

Studies in the review largely defined dropout as attending less than a specified number of sessions ($k = 7$; $d = 0.429$), or as failure to complete a treatment protocol ($k = 8$; $d = 0.447$). Other definitions were operationalised less frequently: missing an appointment without attending future sessions ($k = 1$), not returning after the therapy intake session ($k = 2$), therapist judgement of dropout status ($k = 2$), and discontinuance of treatment prior to significant improvement in symptoms ($k = 1$). One study used an uncoded definition of dropout, and a further three studies did not specify their operationalisation of premature termination. A random effects subgroup analysis suggests no significant differences in effect sizes based on the two primary categorisations of dropout used in studies within the review ($Q(1) = 0.02$, $p = .903$). The finding provides no support for H8.

Publication Status

Studies included in the review were published peer-reviewed journal articles ($k = 19$) or unpublished dissertations/theses ($k = 6$). There was significant variation in the effect sizes based on the publication status of the research paper, with published work reporting significantly larger effect sizes ($d = 0.475$), compared to unpublished work ($d = 0.131$; $Q(1) = 8.93$, $p = .003$; Figure 3), supporting H9.

Publication Bias

Visual inspection of the funnel plot revealed a somewhat symmetrical pattern in the study effect sizes about the mean effect size (Figure 4). Egger's regression test, however, demonstrated significant asymmetry of the funnel plot ($\beta = 1.99$, 95% CI: [1.44, 2.53], $t(23) = 7.55$, $p < .001$). The trim and fill method was used to correct for bias arising from missing studies in the review. The random-random effects model resulted in no imputations and the summary effect size remained unchanged (Figure 4). The conservative fixed-random effects model resulted in the imputation of 13 studies to the left of the mean, corresponding with an adjusted unbiased effect size of $d = 0.112$ (95% CI: [-0.021, 0.253]; Figure 5).

Quality Assessment

Ratings of the risk of bias domains for the included studies were variable. None of the studies included in the meta-analytic review were excluded based on the quality assessment process. A summary of the quality assessment ratings can be found in Table 4.

A random sample of papers ($k = 5$) were independently rated for risk of bias by a second researcher (AF). A two-way mixed effects, average-rater ($n = 2$), absolute agreement intraclass correlation coefficient of .989 (95% CI: [.986, .992]) was reached, indicating "excellent" levels of inter-rater reliability for quality assessment (Koo & Li, 2016).
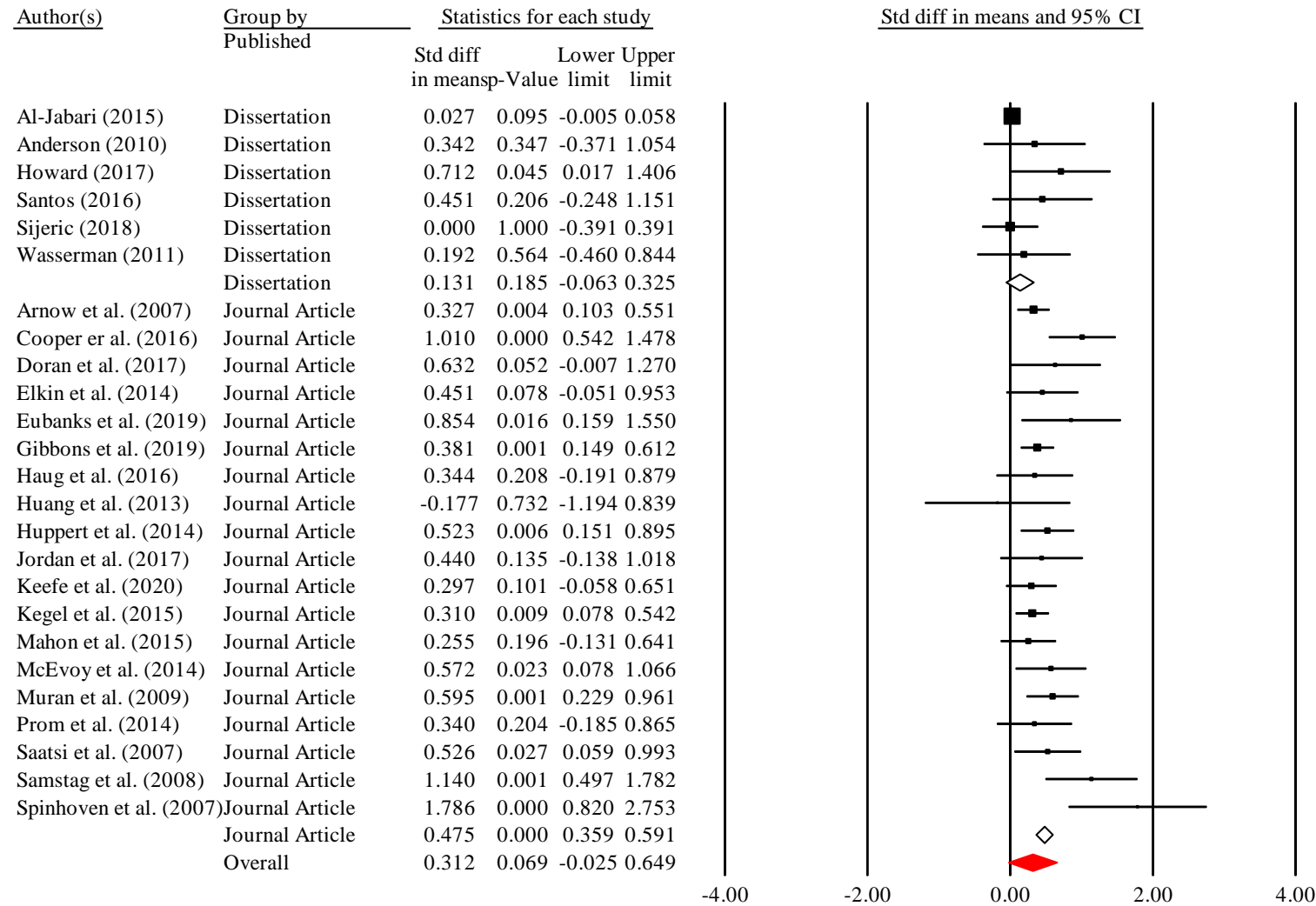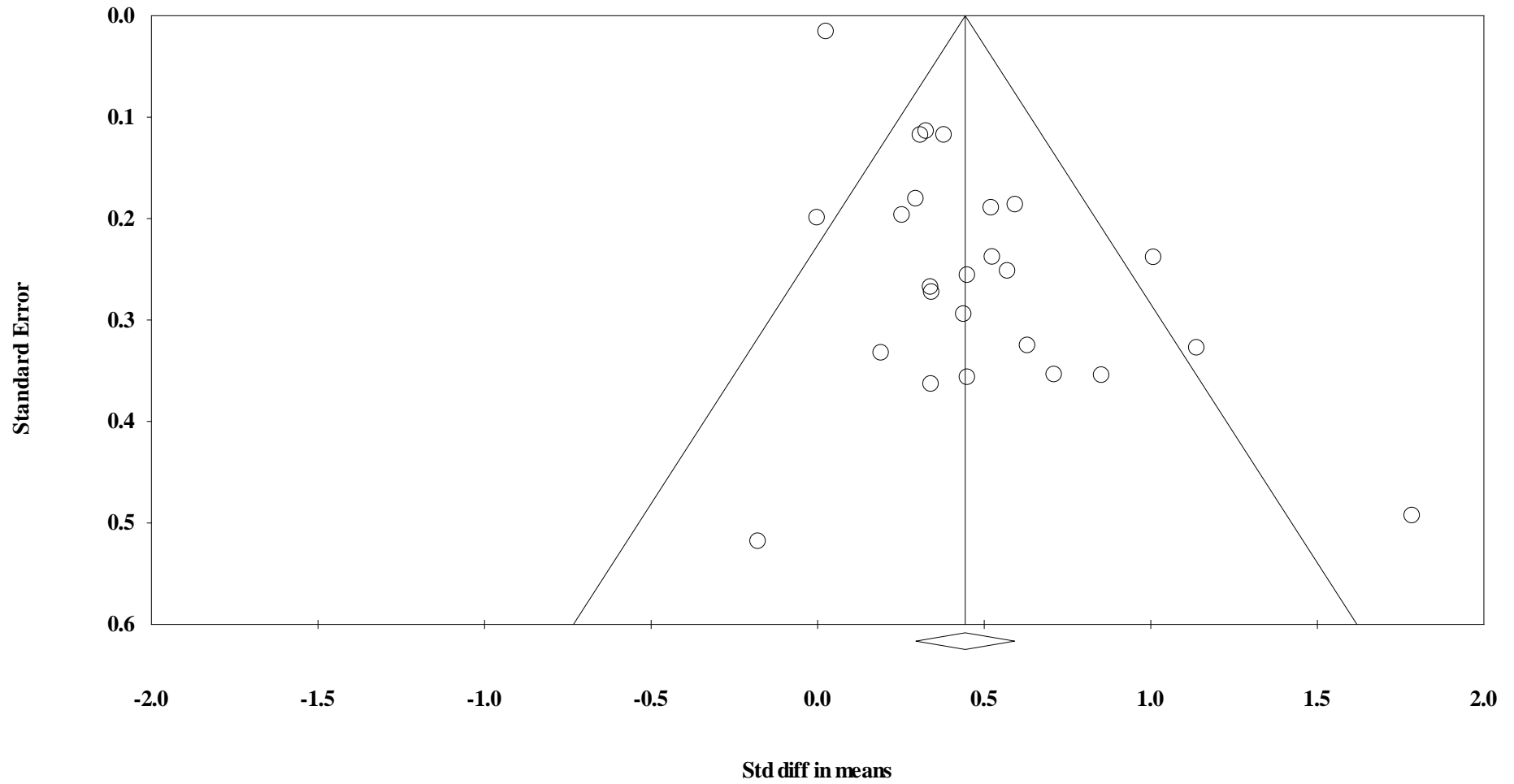
Figure 3

Subgroup Analysis of Publication Status

| Author(s)	Group by Published	Std diff in means	p-Value	Lower limit	Upper limit	Std diff in means and 95% CI
Al-Jabari (2015)	Dissertation	0.027	0.095	-0.005	0.058	
Anderson (2010)	Dissertation	0.342	0.347	-0.371	1.054	
Howard (2017)	Dissertation	0.712	0.045	0.017	1.406	
Santos (2016)	Dissertation	0.451	0.206	-0.248	1.151	
Sijeric (2018)	Dissertation	0.000	1.000	-0.391	0.391	
Wasserman (2011)	Dissertation	0.192	0.564	-0.460	0.844	
	Dissertation	0.131	0.185	-0.063	0.325	
Arnow et al. (2007)	Journal Article	0.327	0.004	0.103	0.551	
Cooper er al. (2016)	Journal Article	1.010	0.000	0.542	1.478	
Doran et al. (2017)	Journal Article	0.632	0.052	-0.007	1.270	
Elkin et al. (2014)	Journal Article	0.451	0.078	-0.051	0.953	
Eubanks et al. (2019)	Journal Article	0.854	0.016	0.159	1.550	
Gibbons et al. (2019)	Journal Article	0.381	0.001	0.149	0.612	
Haug et al. (2016)	Journal Article	0.344	0.208	-0.191	0.879	
Huang et al. (2013)	Journal Article	-0.177	0.732	-1.194	0.839	
Huppert et al. (2014)	Journal Article	0.523	0.006	0.151	0.895	
Jordan et al. (2017)	Journal Article	0.440	0.135	-0.138	1.018	
Keefe et al. (2020)	Journal Article	0.297	0.101	-0.058	0.651	
Kegel et al. (2015)	Journal Article	0.310	0.009	0.078	0.542	
Mahon et al. (2015)	Journal Article	0.255	0.196	-0.131	0.641	
McEvoy et al. (2014)	Journal Article	0.572	0.023	0.078	1.066	
Muran et al. (2009)	Journal Article	0.595	0.001	0.229	0.961	
Prom et al. (2014)	Journal Article	0.340	0.204	-0.185	0.865	
Saatsi et al. (2007)	Journal Article	0.526	0.027	0.059	0.993	
Samstag et al. (2008)	Journal Article	1.140	0.001	0.497	1.782	
Spinhoven et al. (2007)	Journal Article	1.786	0.000	0.820	2.753	
	Journal Article	0.475	0.000	0.359	0.591	
	Overall	0.312	0.069	-0.025	0.649	

Figure 4

Funnel Plot

Note: ○ observed studies; ◇ observed summary effect.

Figure 5

Funnel Plot with Fixed-Random Effect Trim and Fill Imputed Studies
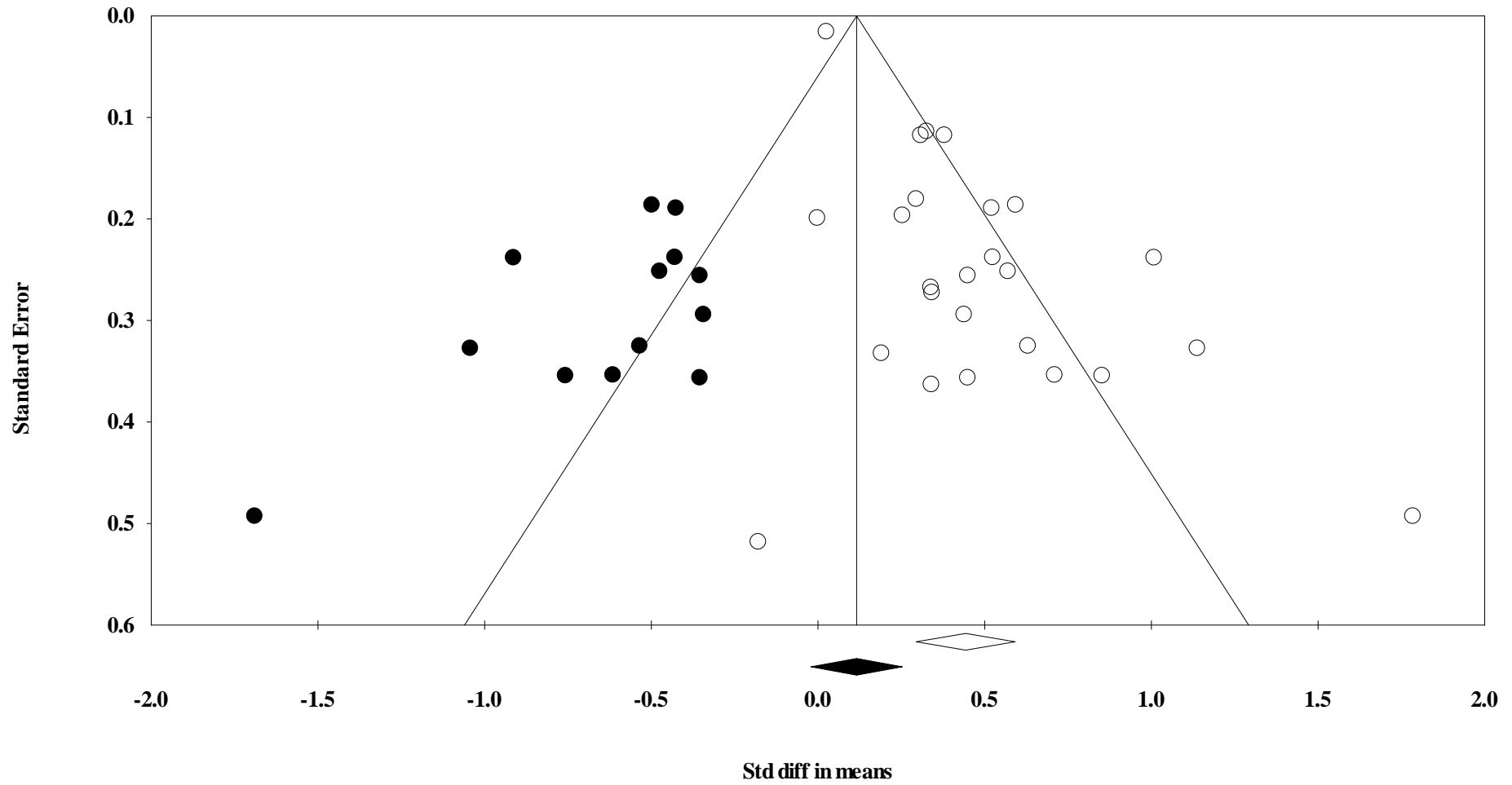
Note: ○ observed studies ● imputed studies; ◇ observed summary effect; ◆ adjusted summary effect.

Table 4

Quality Assessment

	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting	Study Design	Intervention Integrity
Al-Jabari (2015)	M	L	M	L	M	M	M	M
Anderson (2010)	L	M	L	M	L	L	M	L
Arnow et al. (2007)	L	L	M	M	H	M	L	L
Cooper et al (2016)	L	L	M	L	M	M	L	L
Doran et al. 2017)	M	M	L	H	M	M	M	M
Elkin et al. (2014)	L	M	M	M	H	M	L	L
Eubanks et al. (2019)	M	M	M	M	M	L	M	L
Gibbons et al (2019)	L	M	M	M	M	H	L	L
Haug et al (2016)	L	M	M	M	M	L	L	L
Howard (2017)	M	M	M	M	M	L	M	M
Huang et al. (2013)	M	H	M	M	M	M	M	M
Huppert et al. (2014)	M	M	M	H	H	L	L	L
Jordan et al. (2017)	M	M	M	M	M	M	M	L
Keefe et al. (2017)	M	L	M	M	M	M	L	L
Kegel & Flückiger (2015)	M	M	M	M	M	L	M	M
Mahon et al. (2015)	M	M	L	M	M	M	M	M
McEvoy et al. (2014)	M	M	L	M	M	M	M	M
Muran et al. (2009)	M	M	M	M	M	M	M	L
Prom et al. (2014)	H	M	L	M	H	M	M	M
Saatsi et al. (2007)	M	M	M	M	M	M	M	M
Samstag et al. (2008)	M	M	L	L	M	M	M	M
Santos (2016)	L	L	L	M	M	M	L	L
Sijercic (2018)	L	M	M	M	H	L	M	M
Spinhoven et al. (2007)	M	M	M	M	M	L	L	L
Wasserman (2011)	L	M	L	H	M	M	L	L

Note. Risk of bias rating: L = Low, M = Moderate, H = High.

Study Participation

Risk of bias arising from study participation factors was generally low to moderate, with studies largely recruiting participant samples consistent with the population of interest and providing adequate information on recruitment procedures and participant inclusion/exclusion criteria. Eighty-eight percent ($k = 21$) of the studies sufficiently described the sample for key demographic and clinical variables at baseline. One study (Prom et al., 2014) was assigned high risk of bias in this domain for failure to report characteristics of the client sample, participation rates of the intervention or participant inclusion/exclusion criteria.

Study Attrition

There was variability in the reporting of study attrition, including proportions of dropouts, reasons for dropout and description of characteristics of those lost to follow-up. Most papers were allocated a low ($k = 5$) or moderate ($k = 19$) risk of bias rating. For six of the included studies, the proportion of the baseline sample included in the analysis of the association between therapeutic alliance and dropout was less than 80%, and therefore may not be representative of the association in the wider sample.

Prognostic Factor Measurement

The measurement of the prognostic factor, therapeutic alliance, resulted in low ($k = 8$) and moderate ($k = 17$) risk of bias. The methods used to quantify therapeutic alliance were consistently valid and reliable, as this was a pre-requisite for review inclusion. Most papers ($k = 18$) used client- or therapist-reported prospective alliance ratings, however three of the seven observer-rated studies failed to confirm blinding of dropout/completer status at the time of alliance rating. Alliance was not always rated in the same setting (e.g. multicentre studies), or using a consistent measure for all participants (Saatsi et al., 2007), resulting in moderate risk of prognostic measurement bias.

Outcome Measurement

The measurement of dropout was assessed as contributing low ($k = 3$), moderate ($k = 19$) and high ($k = 3$) risk of bias. Studies using specified session limits or treatment protocol completion as methods to establish dropout status were noted for the risk of potential misclassification of early responders/remitters as dropouts. Three studies achieved a high risk of bias rating (Doran et al., 2017; Huppert et al., 2014; Wasserman, 2011) due to insufficient detail on the operationalisation of premature termination used.

Study Confounding

Papers were rated as resulting in high ($k = 5$), moderate ($k = 19$) and low ($k = 1$) risk of bias due to control of study confounding. For many included studies, the alliance-dropout analysis was not the primary study hypothesis, resulting in limited comparisons between groups in key characteristics that potentially required controlling for in statistical analyses.

Statistical Analysis and Reporting

There was variability in the robustness of statistical analysis, resulting in low ($k = 8$), moderate ($k = 16$) and high ($k = 1$) risk of bias. A pervasive issue in this area was the failure of some studies ($k = 14$) to use statistical analyses which account for the hierarchical structure of data, with clients nested within therapists. Even when this was used within studies, nested data was largely not sufficient to calculate an effect size for the meta-analytic synthesis. Furthermore, there was evidence of selective reporting, with three studies not presenting data for all subscales, timepoints or raters measured (Gibbons et al., 2019; Prom et al., 2014; Saatsi et al., 2007).

Study Design

The study design domain was added to the quality assessment tool to rate the robustness of the methodological design of the research. Eleven studies in review were randomised trials, or secondary analyses of participant data collected from randomised trials and were therefore rated as

having low risk of bias. Twelve studies were categorised as observational cohort designs and were assigned moderate risk of bias. Two studies (J. B. Anderson, 2010; Samstag et al., 2008) were categorised as case control methodologies, as the participant sample was selected retrospectively based on outcome status, and therefore assigned moderate risk of bias.

Intervention Integrity

The intervention integrity domain was added to the quality assessment tool to assess adherence to, and potential contamination of, the psychotherapy of interest, as this could pose a confound of therapeutic alliance and dropout. Fourteen studies were rated as having low risk of bias, due to high rates of participants receiving the intervention of interest, measurement of therapist consistency/adherence, and low likelihood of psychotherapy contamination (e.g., due to supervision procedures, fidelity checks, training). Eleven studies were found to have a moderate risk of bias, due to limited stringency in the assessment of fidelity or quality checking of psychotherapy.

Discussion

Summary of Findings

The findings of the current meta-analysis extend those of the previous review completed by Sharf et al. (2010) to support a small-medium association between therapeutic alliance and dropout from psychotherapy ($d = 0.443$, equivalent to $r = .199$). The association in the current review is somewhat smaller than the weighted effect size reported by Sharf et al. ($d = 0.55$), which may be explained by the inclusion of unpublished literature which was shown to report weaker alliance-dropout associations. Taken together, the findings suggest that individual psychotherapy dyads with a weak therapeutic alliance are more likely to end in client premature termination. The findings are consistent with previous meta-analytic reviews assessing the association between alliance and outcome ($r = .22$, Martin et al., 2000; $r = .275$, Horvath et al., 2011; $r = .278$, Flückiger et al., 2018). Although this association is not large, it has been argued to be robust for a variable measured within the complex interplay of processes occurring in the context of psychotherapy (Horvath & Bedi, 2002).

There was significant heterogeneity in the effect sizes of included studies, which was explored with *a priori* moderator analyses. In line with the review hypothesis, there was no significant moderator influence of treatment type, supporting the body of evidence purporting the alliance as a ubiquitous "common factor" and an essential ingredient across psychotherapy in promoting change. However, it must be noted that several studies in the review did not specify the therapy orientation of interest, or otherwise combined multiple therapy orientations within analyses, limiting the subgroups available for comparison of the alliance-dropout association between treatment types.

Despite some marginal differences in effect sizes between treatments of longer ($k = 11$, $d = 0.586$) and shorter (less than 20 sessions) durations ($k = 7$, $d = 0.360$), treatment length was not a significant moderator of the alliance-dropout association, contrary to the findings of Sharf et al.

(2010). However, on closer inspection the Sharf et al. (2010) subgroup analysis was based on six studies, with only two papers in the "medium" length subgroup, making subgroup comparisons unreliable.

Similarly, diagnosis was not found to be a significant moderator of the alliance-dropout association, contrary to findings in the wider alliance-outcome association research (Flückiger et al., 2018), perhaps due to the limited number of studies per subgroup underpowering the ability to find an effect.

The moderating influence of client education on the alliance-dropout association in the Sharf et al. (2010) meta-analysis was not replicated in the current review, which found no modulatory impact of education in the 11 studies which reported client educational history. The discrepancy may be in part due to differences in the coding of educational history, with the current review recording the percentage of clients with higher level qualifications, and Sharf et al. (2010) utilising the percentage of clients completing high school.

Furthermore, the alliance rater did not emerge as a significant moderator of the alliance-dropout relationship, in line with the findings of Sharf et al. (2010), Flückiger et al. (2018) and Horvath et al. (2011). Furthermore, the patterns of effect sizes in the current review varied from previous meta-analytic trends, which suggest that observer-rated alliance is associated with weaker outcome effect sizes than client-rated alliance (Flückiger et al., 2018), and weaker therapist-rated alliance associations with outcome (Horvath et al., 2011).

Contrary to the review hypothesis, there was no significant variability in effect sizes between studies which measure alliance early (before session 7) and those which computed an average alliance rating across psychotherapy. This is contrary to findings in the wider alliance-outcome literature in which the association between alliance and outcome is reportedly stronger when alliance is measured late in therapy, rather than early in the psychotherapy process (Flückiger et al., 2018; Horvath et al., 2011). Only two studies in the current review included an alliance rating

"late" in the therapy process, and therefore subgroup comparisons were not able to be made, which may explain the discrepant findings.

The operational definition of dropout did not significantly explain heterogeneity in study effect sizes, consistent with the findings of Sharf et al. (2010). It is important to note, however, that four of the categorisations of premature termination had insufficient studies ($k < 3$) to be included in the subgroup analyses.

Finally, within the current review, publication status was found to be a significant moderator of the alliance-dropout association. In line with *a priori* hypotheses, published research was found to report stronger alliance-dropout relationships than unpublished work, potentially reflective of publication bias. The importance of publication status in the outcome-alliance association varies from earlier studies (e.g. Martin et al., 2000) reporting no moderator effect of publication status.

Limitations of the Field and Directions for Future Research

The study selection and quality assessment process within the current review highlighted areas of development for the field of alliance research. Assessment of the field of literature to examine eligibility highlights the degree of psychotherapy intervention studies which do not report rates of dropout, or do not consider premature termination status as an outcome variable of interest. It is, therefore, recommended that future research within the field of alliance routinely report comparisons between premature terminators and completers (including alliance and other potentially confounding variables), with sufficient statistical data to enable meta-analysts to calculate a reliable effect size. Studies reporting weaker alliance-dropout effect sizes must be encouraged to publish findings to resolve the publication bias in the field. The review also highlights the diversity of operationalisation of premature termination. It is recommended that the field move towards greater specificity and consistency in operationalisation, with prioritisation given to definitions which take into account change in client symptoms (Hatchett & Park, 2003), to limit the potential for misclassification of early improvers as premature terminators. Finally, future

research with sufficient sample sizes to reliably employ robust statistical methods (e.g. Kegel &
Flückiger, 2015) accounting for the hierarchical structure of data often found in psychotherapy
intervention studies is needed to ensure that Type 1 error rates and effect sizes are not
inappropriately inflated.

A theoretical limitation across the alliance-outcome research is disagreement in the directional
relationship between alliance and clinical change. Researchers have commonly attributed change in
psychotherapy as a direct cause of the development of an alliance between the client and therapist;
an alternative interpretation proposes the alliance may develop as a consequence of the client and
therapists' perceptions of therapeutic change (Doran, 2016). The same dynamic interplay of factors
may be underlying the association with alliance and dropout, with perceived symptom change
enacting a mediatory or confounding role. The temporal primacy of alliance has been suggested by
ratings taken early in the psychotherapy process, however there has been inconsistent evidence for
the association between alliance and outcome when statistically accounting for prior symptom
change (Barber et al., 2000; Derubeis & Feeley, 1990; Feeley et al., 1999; Klein et al., 2003).
Within the wider alliance-outcome literature, the causal or dynamic nature of the relationship
between alliance and outcome is still under investigation, and the specific causal nature of the role
of alliance in premature termination, by controlling for prior symptom change in analyses, requires
further exploration.

Limitations of the Current Review

The search strategy was designed to comprehensively identify relevant papers; however, it is
recognised that the identified research may not represent the absolute extent of literature meeting
the review criteria. Although the current study included significantly more papers ($k = 25$) than the
previous review ($k = 11$) of this area, it may be that the small sample of studies was still
underpowered to reveal subgroup differences in effect sizes. Furthermore, the exclusion criteria
limited studies only to those published in English, excluding the contribution from non-English

language journals. Cross-reviewing of data extraction and quality assessment of all papers by multiple researchers would have increased the reliability of the study characteristics and effect sizes extracted and computed from information within primary papers. However, the second rating of data extraction and quality assessment procedures demonstrated a high degree of inter-rater agreement in the sample of papers independently rated by a second reviewer. The findings require replication in clinical populations that were considered beyond the scope of the current review, such as psychotherapy for children and adolescents, couples, families, groups, substance-misuse treatment and psychotherapy for violent and/or offending behaviour. Furthermore, with a recent exponential increase in the provision of video and telephone psychotherapy, the association between alliance and dropout also requires replication in remote psychotherapy modalities.

Theoretical arguments have been made for the reconceptualisation of alliance to focus on the processes of negotiation and repair of alliance ruptures, as the focus on client and therapist agreement and collaboration masks the important processes of repair of withdrawal ruptures and disagreement within the therapeutic relationship (Safran & Muran, 2006). Over recent years, there has been increasing empirical support for the importance of investigation of alliance ruptures and patterns of alliance (Doran, 2016; Stiles & Goldsmith, 2010). A limitation of the current meta-analytic review, therefore, is the limited focus on the *quality* of the therapeutic alliance (or alliance strength at one or multiple aggregated timepoints), which fails to capture temporal patterns reflecting the dynamic and reflexive nature of the relationship between therapeutic alliance and premature termination. As the current meta-analytic review confirms the association between alliance strength and premature termination, future systematic reviews are needed to further characterise the nuances of this relationship.

Clinical Implications

An implication of the review findings is the need for therapists to attend to, and continually monitor, the strength of therapeutic alliance with clients as soon as therapy begins, to reduce

likelihood of client unilateral termination. A range of validated, and brief, alliance measures (e.g., WAI) are recommended to facilitate routine alliance monitoring. The findings of the review underscore the importance of therapists actively cultivating the alliance from the first session, and remaining vigilant to, and prepared to address, alliance ruptures at the first sign of emergence. Training (e.g. Alliance Focused Training, Eubanks-Carter et al., 2015) aimed at bolstering clinicians' awareness of ruptures, developing metacommunicative interventions, and capabilities in fostering the alliance, may contribute to a reduction in client premature termination; potentially leading to improved therapy experiences and symptom improvement for clients, and better clinical effectiveness for clinicians and mental health services.

Conclusion

The current meta-analysis provides further evidence that establishing robust therapeutic alliance in psychotherapy is an important factor in reducing client unilateral termination of treatment. The association between therapeutic alliance and dropout is largely consistent across psychotherapy orientations, client presenting problems, treatment lengths and operational definitions of alliance and dropout. The effect sizes do, however, vary as a function of publication status, suggestive of a publication bias across the literature. The quality assessment process highlighted areas for methodological development in the field; including consistent consideration of premature termination as an outcome of interest, greater consistency in the operationalisation of dropout, and use of robust statistical methods to account for hierarchical data structures and establish temporal primacy/causal relationships. Furthermore, future research focussing beyond the strength of the alliance in predicting premature termination, and onto the role of alliance ruptures and temporal patterns of alliance is needed.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Aguinis, H., Gottfredson, R. K., & Wright, T. A. (2011). Best-practice recommendations for estimating interaction effects using meta-analysis. *Journal of Organizational Behavior, 32*(8), 1033–1043. https://doi.org/10.1002/job.719

*Al-Jabari, R. (2015). *Therapist (dis)continuity, therapeutic relationship, and (premature) termination in a psychology training clinic* [Doctoral dissertation]. University of North Texas.

Alexander, L. B., & Luborsky, L. (1986). The Penn helping alliance scales. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 325–356). Guildford Press.

*Anderson, J. B. (2010). *The predictive power of therapeutic alliance on treatment retention* [Doctoral dissertation]. Western Michigan University. https://scholarworks.wmich.edu/dissertations/497

*Arnow, B. A., Blasey, C., Manber, R., Constantino, M. J., Markowitz, J. C., Klein, D. N., Thase, M. E., Kocsis, J. H., & Rush, A. J. (2007). Dropouts versus completers among chronically depressed outpatients. *Journal of Affective Disorders, 97*(1–3), 197–202. https://doi.org/10.1016/j.jad.2006.06.017

Barber, J. P., Connolly, M. B., Crits-Christoph, P., Gladis, L., & Siqueland, L. (2000). Alliance predicts patients' outcome beyond in-treatment change in symptoms. *Journal of Consulting and Clinical Psychology, 68*(6), 1027–1032. https://doi.org/10.1037/0022-006X.68.6.1027

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

Björk, T., Björck, C., Clinton, D., Sohlberg, S., & Norring, C. (2009). What happened to the ones who dropped out? Outcome in eating disorder patients who complete or prematurely terminate treatment. *European Eating Disorders Review, 17*(2), 109–119. https://doi.org/10.1002/erv.911

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice, 16*(3), 252–260. https://doi.org/10.1037/h0085885

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2013). *Comprehensive Meta-Analysis* (Version 3) [Computer Software]. Biostat. https://meta-analysis.com/

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley. https://doi.org/10.1002/9780470743386

Cahill, J., Barkham, M., Hardy, G., Rees, A., Shapiro, D. A., Stiles, W. B., & Macaskill, N. (2003). Outcomes of patients completing and not completing cognitive therapy for depression. *British Journal of Clinical Psychology, 42*(2), 133–143. https://doi.org/10.1348/014466503321903553

Card, N. A. (2012). *Applied meta-analysis for social science research*. Guilford Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

*Cooper, A. A., Strunk, D. R., Ryan, E. T., DeRubeis, R. J., Hollon, S. D., & Gallop, R. (2016). The therapeutic alliance and therapist adherence as predictors of dropout from cognitive therapy for depression when combined with antidepressant medication. *Journal of Behavior Therapy and Experimental Psychiatry, 50*, 113–119. https://doi.org/10.1016/j.jbtep.2015.06.005

DeRubeis, R. J., & Feeley, M. (1990). Determinants of change in cognitive therapy for depression. *Cognitive Therapy and Research, 14*(5), 469-482. https://doi.org/10.1007/BF01172968

Doran, J. M. (2016). The working alliance: Where have we been, where are we going? *Psychotherapy Research, 26*(2), 146–163. https://doi.org/10.1080/10503307.2014.954153

*Doran, J. M., Safran, J. D., & Muran, J. C. (2017). An investigation of the relationship between the alliance negotiation scale and psychotherapy process and outcome. *Journal of Clinical Psychology, 73*(4), 449–465. https://doi.org/10.1002/jclp.22340

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

*Elkin, I., Falconnier, L., Smith, Y., Canada, K. E., Henderson, E., Brown, E. R., & McKay, B. M. (2014). Therapist responsiveness and patient engagement in therapy. *Psychotherapy Research, 24*(1), 52–66. https://doi.org/10.1080/10503307.2013.820855

Eubanks-Carter, C., Muran, J. C., & Safran, J. D. (2015). Alliance-focused training. *Psychotherapy, 52*(2), 169–173. https://doi.org/10.1037/a0037596

*Eubanks, C. F., Lubitz, J., Muran, J. C., & Safran, J. D. (2019). Rupture resolution rating system (3RS): Development and validation. *Psychotherapy Research, 29*(3), 306–319. https://doi.org/10.1080/10503307.2018.1552034

Feeley, M., DeRubeis, R. J., & Gelfand, L. A. (1999). The temporal relation of adherence and

    alliance to symptom change in cognitive therapy for depression. *Journal of Consulting and*

    Clinical Psychology, 67(4), 578–582. https://doi.org/10.1037/0022-006X.67.4.578

Flückiger, C., Del, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult

    psychotherapy: A meta-analytic synthesis. *Psychotherapy, 55*(4), 316–340.

    https://doi.org/10.1037/pst0000172

Flückiger, C., Del Re, A. C., Wlodasch, D., Horvath, A. O., Solomonov, N., & Wampold, B. E.

    (2020). Assessing the alliance-outcome association adjusted for patient characteristics and

    treatment processes: A meta-analytic summary of direct comparisons. *Journal of Counseling*

    Psychology, 67(6), 706-711. https://doi.org/10.1037/cou0000424

*Gibbons, M. B. C., Gallop, R., Thompson, D., Gaines, A., Rieger, A., & Crits-Christoph, P.

    (2019). Predictors of treatment attendance in cognitive and dynamic therapies for major

    depressive disorder delivered in a community mental health setting. *Journal of Consulting*

    and Clinical Psychology, 87(8), 745–755. https://doi.org/10.1037/ccp0000414

Hatcher, R. L., & Barends, A. W. (1996). Patients' view of the alliance in psychotherapy:

    Exploratory factor analysis of three alliance measures. *Journal of Consulting and Clinical*

    Psychology, 64(6), 1326–1336. https://doi.org/10.1037//0022-006x.64.6.1326.

Hatcher, R. L., & Gillaspy, J. A. (2006). Development and validation of a revised short version of

    the working alliance inventory. *Psychotherapy Research, 16*(1), 12–25.

    https://doi.org/10.1080/10503300500352500

Hatchett, G. T., & Park, H. L. (2003). Comparison of four operational definitions of premature

    termination. *Psychotherapy, 40*(3), 226–231. https://doi.org/10.1037/0033-3204.40.3.226

*Haug, T., Nordgreen, T., Öst, L. G., Tangen, T., Kvale, G., Hovland, O. J., Heiervang, E. R., &

    Havik, O. E. (2016). Working alliance and competence as predictors of outcome in cognitive

behavioral therapy for social anxiety and panic disorder in adults. *Behaviour Research and Therapy, 77*, 40–51. https://doi.org/10.1016/j.brat.2015.12.004

Hayden, J. A., van der Windt, D. A., Cartwright, J. L., Côté, P., & Bombardier, C. (2013). Assessing bias in studies of prognostic factors. *Annals of Internal Medicine, 158*(4), 280–286. https://doi.org/10.7326/0003-4819-158-4-201302190-00009

Holmes, S. C., Johnson, C. M., Suvak, M. K., Sijercic, I., Monson, C. M., & Wiltsey Stirman, S. (2019). Examining patterns of dose response for clients who do and do not complete cognitive processing therapy. *Journal of Anxiety Disorders, 68*, 102120. https://doi.org/10.1016/j.janxdis.2019.102120

Horvath, A. O., & Bedi, R. P. (2002). The alliance. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 37–69). Oxford University Press.

Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (pp. 25–70). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199737208.003.0002

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the working alliance inventory. *Journal of Counseling Psychology, 36*(2), 223–233. https://doi.org/10.1037/0022-0167.36.2.223

*Howard, K. P. (2017). *Therapist-relevant constructs predicting treatment* [Doctoral dissertation]. Ohio State University.

*Huang, T., Hill, C., & Gelso, C. (2013). Psychotherapy engagers versus non-engagers: Differences in alliance, therapist verbal response modes, and client attachment. *Psychotherapy Research, 23*(5), 568–577. https://doi.org/10.1080/10503307.2013.807378

*Huppert, J. D., Kivity, Y., Barlow, D. H., Shear, M. K., Gorman, J. M., & Woods, S. W. (2014).

Therapist effects and the outcome-alliance correlation in cognitive behavioral therapy for

panic disorder with agoraphobia. *Behaviour Research and Therapy, 52*(1), 26–34.

https://doi.org/10.1016/j.brat.2013.11.001

*Jordan, J., McIntosh, V. V. W., Carter, F. A., Joyce, P. R., Frampton, C. M. A., Luty, S. E.,

McKenzie, J. M., Carter, J. D., & Bulik, C. M. (2017). Predictors of premature termination

from psychotherapy for anorexia nervosa: Low treatment credibility, early therapy alliance,

and self-transcendence. *International Journal of Eating Disorders, 50*(8), 979–983.

https://doi.org/10.1002/eat.22726

*Keefe, J. R., Chambless, D. L., Barber, J. P., & Milrod, B. L. (2020). Predictors and moderators of

treatment dropout in cognitive-behavioral and psychodynamic therapies for panic disorder.

Psychotherapy Research, 31(4), 432–442. https://doi.org/10.1080/10503307.2020.1784487

*Kegel, A. F., & Flückiger, C. (2015). Predicting psychotherapy dropouts: A multilevel approach.

Clinical Psychology & Psychotherapy, 22(5), 377–386. https://doi.org/10.1002/cpp.1899

Klein, D. N., Schwartz, J. E., Santiago, N. J., Vivian, D., Vocisano, C., Arnow, B., Manber, R.,

Riso, L. P., McCullough, J. P., Borian, F. E., Castonguay, L. G., Blalock, J. A., Markowitz, J.

C., Rothbaum, B., Thase, M. E., Miller, I. W., & Keller, M. B. (2003). Therapeutic alliance in

depression treatment: Controlling for prior change and patient characteristics. *Journal of

Consulting and Clinical Psychology, 71*(6), 997–1006. https://doi.org/10.1037/0022-

006X.71.6.997

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation

coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163.

https://doi.org/10.1016/j.jcm.2016.02.012

Lambert, M. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research, 17*(1), 1–14. https://doi.org/10.1080/10503300601032506

Longo, D. A., Lent, R. W., & Brown, S. D. (1992). Social cognitive variables in the prediction of client motivation and attrition. *Journal of Counseling Psychology, 39*(4), 447–452. https://doi.org/10.1037/0022-0167.39.4.447

Luborsky, L., Crits-Christoph, P., Alexander, L., Margolis, M., & Cohen, M. (1983). Two helping alliance methods for predicting outcomes of psychotherapy: A counting signs vs. a global rating method. *Journal of Nervous and Mental Disease, 171*(8), 480–491. https://doi.org/10.1097/00005053-198308000-00005

*Mahon, M., Laux, J. M., McGuire Wise, S., Ritchie, M. H., Piazza, N. J., & Tiamiyu, M. F. (2015). Brief therapy at a university counseling center: Working alliance, readiness to change, and symptom severity. *Journal of College Counseling, 18*(3), 233–243. https://doi.org/10.1002/jocc.12017

Marmar, C. R., Horowitz, M. J., Weiss, D. S., & Marziali, E. (1986). The development of the therapeutic alliance rating system. In L. S. Green & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 367–390). Guilford Press.

Martin, D. J., Garske, J. P., & Katherine Davis, M. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 68*(3), 438–450. https://doi.org/10.1037/0022-006X.68.3.438

*McEvoy, P. M., Burgess, M. M., & Nathan, P. (2014). The relationship between interpersonal problems, therapeutic alliance, and outcomes following group and individual cognitive behaviour therapy. *Journal of Affective Disorders, 157*, 25–32. https://doi.org/10.1016/j.jad.2013.12.038

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

*Muran, J. C., Safran, J. D., Gorman, B. S., Samstag, L. W., Eubanks-Carter, C., & Winston, A. (2009). The relationship of early alliance ruptures and their resolution to process and outcome in three time-limited psychotherapies for personality disorders. *Psychotherapy, 46*(2), 233–248. https://doi.org/10.1037/a0016085

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine, 26*(25), 4544–4562. https://doi.org/10.1002/sim.2889

*Prom, M. C., Stovall, J., Bedregal, L. E., Phillips, J., & Davidson, M. A. (2014). Therapeutic alliance: Satisfaction and attrition of patients from a mental health clinic in Ayacucho, Peru. *International Psychiatry, 11*(4), 95–97. https://doi.org/10.1192/s1749367600004707

Roos, J., & Werbart, A. (2013). Therapist and relationship factors influencing dropout from individual psychotherapy: A literature review. *Psychotherapy Research, 23*(4), 394–418. https://doi.org/10.1080/10503307.2013.775528

Ryan, R. (2016). *Heterogeneity and subgroup analyses in Cochrane consumers and communication group reviews: Planning the analysis at protocol stage.* Cochrane Consumers and Communication Review Group. https://cccrg.cochrane.org/sites/cccrg.cochrane.org/files/public/uploads/heterogeneity_subgroup_analyses_revising_december_1st_2016.pdf

*Saatsi, S., Hardy, G. E., & Cahill, J. (2007). Predictors of outcome and completion status in cognitive therapy for depression. *Psychotherapy Research, 17*(2), 185–195. https://doi.org/10.1080/10503300600779420

Safran, J. D., & Muran, J. C. (2006). Has the concept of the therapeutic alliance outlived its usefulness? *Psychotherapy, 43*(3), 286–291. https://doi.org/10.1037/0033-3204.43.3.286

*Samstag, L. W., Muran, J. C., Wachtel, P. L., Slade, A., Safran, J. D., & Winston, A. (2008). Evaluating negative process: A comparison of working alliance, interpersonal behavior, and narrative coherency among three psychotherapy outcome conditions. *American Journal of Psychotherapy, 62*(2), 165–194. https://doi.org/10.1176/appi.psychotherapy.2008.62.2.165

*Santos, M. M. (2016). *Is the therapeutic alliance associated with and predictive of treatment retention and outcome among latinos?: A secondary analysis of an RCT of behavioral activation for Latinos with depression versus treatment-as-usual* [Doctoral dissertation]. University of Wisconsin-Milwaukee. https://dc.uwm.edu/etd/1306

Sharf, J. (2008). *Psychotherapy dropout: A meta-analytic review of premature termination.* [Doctoral dissertation]. Icahn School of Medicine, Mount Sinai.

Sharf, J., Primavera, L. H., & Diener, M. J. (2010). Dropout and therapeutic alliance: A meta-analysis of adult individual psychotherapy. *Psychotherapy, 47*(4), 637–645. https://doi.org/10.1037/a0021175

*Sijercic, I. (2018). *The effect of early therapeutic alliance on treatment dropout in cognitive processing therapy and client client factors as moderators of this relationship.* [Doctoral dissertation]. Ryerson University. https://digital.library.ryerson.ca/islandora/object/RULA%3A8093/datastream/OBJ/download/The_effect_of_early_therapeutic_alliance_on_treatment_dropout_in_cognitive_processing_therapy_and_client_factors_as_moderators_of_this_relationship.pdf

*Spinhoven, P., Giesen-Bloo, J., Van Dyck, R., Kooiman, K., & Arntz, A. (2007). The therapeutic alliance in schema-focused therapy and transference-focused psychotherapy for borderline

personality disorder. *Journal of Consulting and Clinical Psychology, 75*(1), 104–115.

https://doi.org/10.1037/0022-006X.75.1.104

Stiles, W. B., & Goldsmith, J. Z. (2010). The alliance over time. In J. C. Muran & J. P. Barber

(Eds.), *The therapeutic alliance: An evidence-based guide to practice* (pp. 44–62). The

Guildford Press.

Suh, C. S., Strupp, H. G., & O'Malley, S. S. (1986). The Vanderbilt process measures: The

psychotherapy process scale (VPPS) and the negative indicators scale (VNIS). In L. S.

Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp.

285–323). Guilford Press.

Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-

analysis. *Journal of Consulting and Clinical Psychology, 80*(4), 547–559.

https://doi.org/10.1037/a0028226

Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically

reviewing the literature: Providing the research evidence for public health nursing

interventions. *Worldviews on Evidence-Based Nursing, 1*(3), 176–184.

https://doi.org/10.1111/j.1524-475X.2004.04006.x

Todd, D. M., Deane, F. P., & Bragdon, R. A. (2003). Client and therapist reasons for termination: A

conceptualization and preliminary validation. *Journal of Clinical Psychology, 59*(1), 133–

147. https://doi.org/10.1002/jclp.10123

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the working alliance inventory.

Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1(3), 207–210.

https://doi.org/1040-3590/89/S00.75

*Wasserman, R. H. (2011). *The role of working alliance in the treatment of borderline personality

disorder* [Doctoral dissertation]. The Pennsylvania State University.

Wierzbicki, M., & Perarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice, 24*(2), 190–195. https://doi.org/10.1037/0735-7028.24.2.190

Williams, S. L., Ketring, S. A., & Salts, C. J. (2005). Premature termination as a function of intake data based on ethnicity, gender, socioeconomic status, and income. *Contemporary Family Therapy, 27*(2), 213–231. https://doi.org/10.1007/s10591-005-4040-8

Appendices

Appendix A. Studies Excluded due to Insufficient Data

Citation	Reason and Actions
Ryan, E. (2009). *The therapeutic alliance in cognitive therapy for depression in combination with antidepressant medication: relations to subsequent symptom change and treatment retention* (Doctoral dissertation, The Ohio State University).	Unable to calculate effect size as there were no means and standard deviations reported. Author completed a logistic regression and reported odds ratios but not confidence intervals (which are needed to calculate an effect size). An email address for the author or a ResearchGate profile could not be found. Therefore, the paper was excluded from the review.
Green, J. (2019). *Adult Attachment Style and the Therapeutic Alliance as Predictors of Premature Therapy Termination: A Retrospective Chart Review.* (Doctoral dissertation, University of Connecticut).	The results section reports only one premature terminator completing alliance measures. Therefore, it was not possible to compute an effect size from the data. The paper was excluded from the review.
Walker, R. N. (2012). *Therapeutic Factors and Premature Termination in a University-based Counseling Training Clinic* (Doctoral dissertation, University of Georgia).	The paper reports the results of a binary logistic regression; however, no odds ratios and confidence intervals were reported. There were also no means and standard deviations of therapeutic alliance for completers/ terminators reported. An email address for the author or a ResearchGate profile could not be found. Therefore, the paper was excluded from review.

Marmarosh, C. L., Gelso, C. J., Markin, R. D., Majors, R., Mallery, C., & Choi, J. (2009). The Real Relationship in Psychotherapy: Relationships to Adult Attachments, Working Alliance, Transference, and Therapy Outcome. *Journal of Counseling Psychology, 56*(3), 337-350. https://doi.org/10.1037/a0015169	The results section of the paper reports an ANOVA revealing no significant differences between third-session WAI scores between clients who completed the third-session measures but dropped out ($n = 17$) and those who participated to termination ($n = 31$). However, no means, standard deviations or F-statistics are reported for this analysis. Email contact made with Cheri Marmarosh, who confirms that she no longer has access to the data and therefore cannot provide the necessary descriptive statistics to calculate an effect size. Therefore, the paper was excluded from the review.
Choi, G. (2014). *Client Attachment as a Predictor of Therapy Outcome and Premature Termination.* (Doctoral dissertation, University of Tennessee).	The results section and Table 5 report point-biserial correlation between fifth session working alliance bond, task and goal subscales and premature termination. However, there are no sample sizes reported with the correlations to compute an effect size for these analyses. Tables 2 and 3 in the results section do report the n-value for analyses relating to premature termination and working alliance, however these vary ($n = 55$ and $n = 46$, respectively), suggesting that there is potentially missing data. The authors also report in the results section that "Whereas working alliance at Session 3 and 5 had no significant correlation with premature termination, every subscale

	of working alliance at termination were significantly negatively associated with premature termination" (Page 24). However, they do not report the correlations between the total WAI scores at session 3 and 5 and premature termination. An email was sent to the author through ResearchGate to request additional data, however no response was received. A reliable effect size cannot be calculated from the available data, therefore the paper was excluded from the review.
Coutinho, J., Ribeiro, E., Fernandes, C., Sousa, I., & Safran, J. D. (2014). The development of the therapeutic alliance and the emergence of alliance ruptures. *Anales de Psicologia, 30*(3), 985-994. https://doi.org/10.6018/analesps.30.3.168911	The results section reports the results of non-parametric smooth spline analyses to model therapeutic alliance over the time course of treatment and are used to predict outcome (including premature termination). The data presented was not sufficient to calculate an effect size. Attempted contact was made with the author by email and through ResearchGate, requesting descriptive statistics for the analyses. However, no response was received. Therefore, the paper was excluded from the review.

Appendix B. Quality Assessment Tool

Domains 1-6 amended items from Quality in Prognostic Studies tool (Hayden et al., 2013)

Domains 7-8 amended items from the Effective Public Health Practice Project Quality Assessment

Tool for Quantitative Studies (Thomas et al., 2004)

Domain	Prompts
1) Study Participation	
Source of target population	The sample of interest is likely to represent the population of interest based for key characteristics (e.g., age, gender, baseline presenting problems).
Method used to identify problem	The sampling frame and recruitment are adequately described, possibly including methods to identify the sample, place of recruitment, and period of recruitment
Inclusion and exclusion criteria	Inclusion and exclusion criteria for study participation are adequately described
Adequate study participation	There is adequate participation in the study by eligible individuals (e.g., more than 75% of those eligible proceed to intervention)
Baseline characteristics	The baseline study sample is adequately described for key characteristics (e.g., age, gender, baseline presenting problems).
Summary Study Participation Rating	High risk of bias - The relationship between therapeutic alliance and dropout is very likely to be different for participants and eligible nonparticipants
	Moderate risk of bias- The relationship between therapeutic alliance and dropout may be different for participants and eligible nonparticipants

	Low risk of bias - The relationship between the therapeutic alliance and dropout is unlikely to be different for participants and eligible nonparticipants
2) Study Attrition	
Proportion of baseline sample available for analysis	Proportion of baseline sample available for therapeutic alliance/dropout analysis is more than 80%
Reasons and potential impact of subjects lost to follow up	Reasons for loss to follow up are described
Outcome and prognostic factor information on those lost to follow up	Participants lost to follow up are adequately described for key characteristics (age, gender, baseline presenting problems).
Summary Study Attrition Rating	High risk of bias - The relationship between therapeutic alliance and dropout is very likely to be different for completing and noncompleting participants
	Moderate risk of bias- The relationship between therapeutic alliance and dropout may be different for completing and noncompleting participants
	Low risk of bias - The relationship between the therapeutic alliance and dropout is unlikely to be different for completing and non-completing participants
3) Prognostic Factor (Therapeutic Alliance) Measurement	
Definition of the PF	A clear definition or description of therapeutic alliance is provided

Valid and reliable measurement of PF	Method used to measure therapeutic alliance is adequately valid and reliable to limit misclassification bias
	The measurement of therapeutic alliance is blind to dropout/completer status at the time of completing
	Continuous variables are reported for therapeutic alliance or appropriate cut-offs are used
Method and setting of PF measurement	The method and setting of measurement of therapeutic alliance is the same for all study participants
Proportion of data on PF available for analysis	More than 80% of the study sample (including dropouts and completers) has completed data for therapeutic alliance
Method used for missing data	Appropriate methods are used for missing therapeutic alliance data
Summary Prognostic Factor Measurement Rating	High risk of bias - The measurement of therapeutic alliance is very likely to be different for different levels between dropout and completers
	Moderate risk of bias- The measurement of therapeutic alliance may be different for different levels between dropout and completers
	Low risk of bias - The measurement of therapeutic alliance is unlikely to be different for different levels between dropout and completers

4) Outcome (Premature Termination) Measurement

Definition of the Outcome	A clear definition of dropout is provided
Valid and reliable measurement of Outcome	The method of dropout used in valid and reliable to limit misclassification bias

Method and setting of Outcome Measurement	The method and setting of dropout/completer status is the same for all study participants
Summary Outcome Measurement Rating	High risk of bias - The measurement of dropout/completer status is very likely to be different related to the baseline level of the therapeutic alliance
	Moderate risk of bias- The measurement of dropout/completer status may be different related to the baseline level of the therapeutic alliance
	Low risk of bias - The measurement of dropout/completer status may be different related to the baseline level of the therapeutic alliance
5) Study Confounding	
Important Confounders measured	All important confounders are measured (e.g., client characteristics including age and gender, therapist characteristics including age, gender and experiences, baseline severity of presenting problems, treatment type)
Method and setting of Confounding Measurement	The method and setting of confounding measurement are the same for all study participants
Appropriate accounting for confounding	Important potential confounders are accounted for in the study design
	Important potential confounders are accounted for in the analysis
Summary Study Confounding Rating	High risk of bias - The observed effect of therapeutic alliance on the dropout/completer status is very likely to be distorted by another factor.

	Moderate risk of bias- The observed effect of therapeutic alliance on the dropout/completer status may be distorted by another factor.
	Low risk of bias - The observed effect of therapeutic alliance on the dropout/completer status is unlikely to be distorted by another factor.

6) Statistical Analysis and Reporting	
Presentation of analytical strategy	There is sufficient presentation of data to assess the adequacy of the analysis
Model development strategy	The selected statistical model is adequate for the design of the study
Reporting of results	There is a description of the association between therapeutic alliance and dropout, including information about the statistical significance
	Continuous variables are reported, or cut-off points are used
	There is no selective reporting of results
Summary Statistical Analysis and Reporting Rating	High risk of bias - The reported results are very likely to be spurious or biased related to analysis or reporting
	Moderate risk of bias- The reported results may be likely to be spurious or biased related to analysis or reporting
	Low risk of bias - The reported results are unlikely to be spurious or biased related to analysis or reporting

7) Study Design	
Indicate study design	Randomised controlled trial, controlled clinical trial, cohort analytic, case-control, cohort, interrupted time series, other.

Randomisation	Was the study described as randomised?
Method of randomisation	If yes, was the method of randomisation described?
	High risk of bias - will be assigned to studies using any other method, or not stating the method used.
	Moderate risk of bias- will be assigned to those that described a cohort analytic study, a case control study, a cohort design, or an interrupted time series.
	Low risk of bias - will be assigned to those articles described as RCTs and CCTs.
8) Intervention Integrity	
Exposure of interest	What percentage of participants received the allocated intervention or exposure of interest? 1) 80-100%, 2) 60-79%, 3) less than 60%, 4) can't tell
Consistency of intervention	Was the consistency of the intervention measured? (e.g., adherence to therapy)
Unintended intervention	Is it likely that subjects received an unintended intervention (contamination or co-intervention) that may influence the results?
Summary of Intervention Integrity	High risk of bias - The participants are very likely to have received an unintended intervention
	Moderate risk of bias - the participants may have received an unintended or inconsistent intervention
	Low risk of bias - the participants are unlikely to have received an unintended or inconsistent intervention.

Part 2: Research Report

The role of therapists' interpersonal skills during "not on track" therapy sessions in predicting response

to counselling person-centred experiential and cognitive behavioural therapy interventions for

depression

Abstract

Objectives

The facilitative interpersonal skills (FIS) performance task prospectively rates therapists' abilities across eight interpersonal domains and has been associated with client outcome. The current study translated the methodology to rate therapists' FIS in genuine therapy sessions deemed to be "not on track" (NOT) towards recovery, with the aim to assess the association with client outcome and compare the strength of the relationship between therapy orientations.

Design

The research comprised a quantitative proof-of-concept study, re-examining data collected as part of the PRaCTICED trial, comparing effectiveness of cognitive behavioural and person-centred experiential therapy for depression.

Methods

The FIS methodology was adapted for rating therapists' FIS in a sample of genuine therapy sessions ($n = 59$) identified to be NOT. Treatment outcome was assessed using symptom measures collected 12-months post-randomisation to treatment.

Results

In the full sample, therapist FIS during NOT sessions was not a significant predictor of outcome at 12-month follow-up, and the association was not significantly moderated by therapy orientation. Post-hoc analyses suggest an interaction between FIS and depression severity in predicting outcome at 12-months post-randomisation.

Conclusion

The research establishes in-session FIS as a promising methodology for measuring therapists' interpersonal qualities in "problematic" therapy sessions. A differential impact of therapists' FIS was observed between depression severity subgroups, providing preliminary evidence that clients with

greater baseline clinical severity of depression symptoms may respond less favourably to therapists with high levels of FIS during "problematic" therapy sessions. Future research with larger samples comparing FIS in distinct clinical subgroups is needed to substantiate the finding.

Practitioner Points

- Therapists' use of FIS during "problematic" therapy sessions has a complex relationship with client outcome.
- Therapist modulation of interpersonal skills in line with client characteristics, such as baseline depression severity, may be linked to the outcome of therapy, although replication is needed to confirm this finding.

Limitations

- The in-session FIS methodology is limited in the inability to isolate therapists' contributions to the facilitative environment and assesses FIS in one session in the context of a whole treatment.
- Future research with a larger sample size is needed to employ robust multilevel modelling analyses and examine FIS domain items with sufficient statistical power.

Introduction

Therapists have been found to play a significant role in clients' responses to psychological therapies (Barkham et al., 2017). A strong body of evidence supports the demonstrable effectiveness of common relational process variables such as empathy, positive regard, and the therapeutic alliance, in improving outcomes of individual psychotherapy for adults (John C. Norcross & Lambert, 2018). Therapeutic alliance, often defined in the process literature as agreement between the client and therapist on the goals and tasks of therapy, in the presence of a dyadic bond (Bordin, 1979), is one of the most frequently investigated process variables related to success in psychotherapy, across psychotherapy orientations (Flückiger et al., 2018). Therapeutic alliance has been found to have a moderate association with client outcome (Flückiger et al., 2018; Horvath et al., 2011), including premature termination (Bentham, 2021; Sharf et al., 2010). Therapist empathy, the capacity to understand what the client is experiencing or trying to express, has also been found to be a moderately strong predictor of a range of client outcomes, across presentations and therapeutic orientations (Elliott et al., 2011, 2018). A second Rogerian capacity, positive regard (non-possessive warmth), has also been found to have a small-moderate association with outcome in comprehensive meta-analytic reviews (Farber et al., 2018; Farber & Doolin, 2011). Additional facilitative capacities of the therapist, including emotional expression (Peluso & Freund, 2018), alliance rupture-repair (Eubanks et al., 2018; Safran et al., 2011) and cultivating positive expectations (Constantino et al., 2018) have been shown to be "probably effective" elements of the psychotherapeutic relationship associated with improved clinical outcomes for clients (Norcross & Lambert, 2018). Further therapist characteristics such as verbal fluency and expressiveness (Greenberg & Paivio, 2003), and persuasiveness (Truax et al., 1968) have also emerged as important interpersonal capacities of effective therapists.

Facilitative Interpersonal Skills

The Facilitative Interpersonal Skills (FIS; Anderson et al., 2006) research builds on the robust findings of the process-outcome research by operationalising several common and intercorrelated

76

relational processes in a single measure. FIS are defined as an individual's ability to perceive, decode, and express a variety of interpersonal messages (Anderson et al., 2006). The FIS methodology is pan-theoretical, and can assess therapists' interpersonal abilities across treatment orientations and forms of psychotherapy (Anderson, Finkelstein et al., 2020). The FIS tool was originally developed as a performance task to prospectively rate therapists' FIS, aiming to isolate therapist responses and standardise the interpersonal contributions of the client. The FIS task involves a therapist being presented with audio-visual clips of client interactions and responding as if they were the therapist. The stimulus clips simulate challenging moments in therapy, representing presentations at varying points on the interpersonal circumplex (e.g., friendly-hostile, controlling-submissive; Anderson et al., 2020). Responses are recorded and evaluated by trained raters on eight relational capacities. A description of each of the FIS domains is presented in Table 1.

Table 1

FIS Domains and Descriptions

FIS Domain	Description
Verbal Fluency	This item rates the extent to which the therapist is verbally comfortable and at ease in communicating. High verbal fluency responses sound relaxed, well-paced and feature a rhythmical quality. Low verbal fluency is often indicated by expressions of anxiety in communicating e.g., stammering, awkward/extended pauses.
Hope and Positive Expectations	This item rates expressions of hope, optimism, and positive expectations for change. A response high in hope and positive expectations will demonstrate believable, personalised pathways for the client to recover and make changes. Responses low in hope and positive expectations will often show overt pessimism and hopelessness or describe an issue as out of the client's control.
Persuasiveness	This item rates the therapist's capacity to induce the other to accept a view that may be different from his or her own view. Highly persuasive responses are appropriate and specific, use clear language, and feel convincing. Responses low in persuasiveness feel generic, lack credibility, or may be incoherent.

Emotional Expression	This item rates the extent to which the therapist's response is delivered with effective expressions of emotion. Highly emotionally expressive responses are engaging, show clear affect, and have a strong prosody. Responses lacking emotional expression appear dull, flat and feature little affect.
Warmth, Acceptance, and Understanding	This item is a rating of the ability of the therapist to care for and accept the other. Highly rated responses show the therapist's ability to demonstrate genuine care for the client, acceptance for them as they are, and compassion. Responses low in warmth/acceptance/understanding include a judgemental attitude, guilt-inducing comments, and signs of frustration/exasperation with a client.
Empathy	This item rates the therapist's capacity to respond with an expressed understanding of the subjective experience of the client. Highly empathic responses demonstrate active listening, use of the client's language, and inference of implicit aspects of the client's experience. Responses low in empathy disregard or distort the client's experience, or otherwise misidentify obvious aspects of the client's situation.
Alliance Bond Capacity	This item rates the therapist's capacity to provide a collaborative environment, one in which there is recognition of the need to work with the client jointly on problems. Responses demonstrating strong alliance bond capacity will ensure the problems are worked on collaboratively and engage the client as part of the therapy process. Low alliance bond capacity responses will undermine fostering a collaboration by putting all the responsibility on the client or engaging in moralistic lecturing.
Problem/Rupture-Repair Responsiveness	In the performance task, each simulated client had a clear interpersonal "rupture" which required an individual solution. The rupture-repair responsiveness item therefore rated the extent to which the therapist appears responsive to the interpersonal issue. As interpersonal ruptures were not guaranteed in the current study, the more general "problem responsiveness" rating system was used (developed by De Jong and colleagues). This item measures the extent to which the therapist appears to offer solutions to the specific client problem, which could include an interpersonal rupture, and whether said solutions are likely to be helpful.

Note. Adapted from Anderson, Finkelstein et al., 2020; Anderson & Patterson, 2013; Anderson & Perlman, 2019).

In a series of studies, Anderson and colleagues have demonstrated the ability of the FIS performance task in predicting client therapeutic outcomes (Anderson et al., 2009; Anderson, Crowley et al., 2016; Anderson, McClintock et al., 2016). An early study by the research team found that FIS predicts therapy outcome over and above therapist self-reported social skills, age and

demographic factors (Anderson et al., 2009). A second study, comparing therapists' training status, found that FIS skills, and not therapists' level of training, predicted their therapeutic alliance and outcome with clients in therapy (Anderson, Crowley et al., 2016). Furthermore, the FIS performance task has been shown to have prospective predictive ability, forecasting client symptom change in therapy occurring over a year after the performance task was completed by trainee therapists (Anderson, McClintock et al., 2016). It is important to note, however, that the differences between high- and low-rated FIS therapists appear to diminish over longer treatment durations (e.g., 16 sessions; Anderson, McClintock et al., 2016). These findings suggest that therapists' interpersonal characteristics are likely to optimise the facilitative environment and impactful therapy processes, leading to improved working alliance and clinical outcomes for clients. The link between FIS and therapeutic outcome for clients is made all the more important by recent evidence that therapist FIS is amenable to improvement and development through simple modelling, feedback, and more substantive training (Anderson, Perlman et al., 2020; Jones, 2019; Perlman et al., 2020).

Establishing the validity of FIS in actual therapy sessions is important as interpersonal abilities are intrinsically interactive and will be dynamically influenced by the client in the interpersonal/therapy context. Furthermore, it is important to highlight how FIS combine and interact within the therapy room to promote the conditions for clinical change. To date, there is limited evidence that therapists' interpersonal skills and behaviours in actual therapy sessions are predictive of client outcome. Uhlin (2011) adapted the original FIS manual to create a protocol for rating therapists' FIS in naturalistic therapy sessions. The protocol was used to rate 45 video-taped therapy sessions and a moderate association between FIS and working alliance was reported, but no relationship with client outcome was observed (Uhlin, 2011). The inconsistent finding could be accounted for by the rating of the third therapy session for each client, potentially limiting the observation of interpersonally challenging client interactions that the original FIS simulates. The session choice is particularly salient given the most significant correlations between outcome and

interpersonal skills are during periods of time identified as "critical incidents" in therapy (Janzen, 2007).

Interpersonal Skills Across Therapy Orientations

Although FIS are proposed to be transtheoretical common factors across psychotherapy orientations, it may be that interpersonal abilities are emphasised and utilised to different degrees across psychotherapies to promote clinical change. It is argued the interpersonal/relational abilities of the therapist cannot be fully independent from the therapeutic modality used, suggesting an interaction between technical and relational factors to elicit client change (Anderson, Crowley, et al., 2016). Treatment modalities with a relational focus, such as brief relational therapy (Newhill et al., 2003), time-limited dynamic psychotherapy (Strupp & Binder, 1984) and person-centred experiential therapy (Hill, 2011; Pearce et al., 2012; Sanders & Hill, 2014), are likely to provide the conditions in which high utilisation of a therapist's FIS would enhance the therapeutic techniques and methods to optimise the effectiveness of therapy (Anderson, Crowley, et al., 2016). Currently, there is no extant literature, using the FIS method, to consider the interactions between psychotherapy orientations and therapist interpersonal skills. An area requiring exploration is whether a therapists' ability to use interpersonal skills to create the facilitative environment is more central to therapeutic frameworks with an interpersonal, emotion-focussed, or humanistic emphasis, in which the primary mechanism of change is often within the therapeutic relationship.

The Current Study

The current study translated the FIS performance task methodology to the in-session assessment of FIS in genuine therapy sessions. To address the limitations of Uhlin's (2011) methodology, and to maximise the potential that the treatment sessions included "critical incidents" and relationally challenging moments in therapy which elicit the use of therapists' interpersonal skills, the current study identified client sessions which were deemed to be "not on track" (NOT) towards recovery. Therapists' FIS were rated using audio-recordings of therapy sessions from a randomised non-

inferiority trial comparing two common and contrasting psychotherapy treatments for depression, namely, cognitive behavioural therapy (CBT) and person-centred experiential therapy (PCET). As such, the current study is the first to employ the in-session FIS methodology, with increased specificity to potentially "problematic" sessions, and examine associations with client outcome across two distinct psychotherapy treatments. In summary, the two aims and hypotheses for the current study are:

1) To assess if therapists' facilitative interpersonal skills during NOT client sessions are predictive of client outcome. In line with the findings of previous research, it is hypothesised that clients engaging in therapy with therapists with higher utilisation of FIS will have greater clinical improvement over the duration of treatment (Hypothesis 1 – H1).

2) To consider interactions between psychotherapy orientations and therapist skills by assessing if FIS are more important for client outcome in PCET or CBT. It is hypothesised that higher scores on the FIS will be more predictive of client symptomatic improvement in PCET than in CBT (Hypothesis 2 – H2).

<center>**Method**</center>

Design

The current study used a newly developed methodology to examine previously untested research hypotheses, and therefore the research is best framed as exploratory to assess proof-of-concept. The research used a quantitative methodology to re-examine data collected in a randomised controlled non-inferiority trial.

Aspects of the project, primarily the coding of therapist FIS, were completed collaboratively with another researcher (AF, Trainee Clinical Psychologist). Information regarding the shared and distinct aspects of the projects can be found in Appendix A.

Ethical Approval

The PRaCTICED trial attained ethical approval from the Health Research Authority (REC 14/YH/0001). Ethical approval was also granted by the University of Sheffield Ethics Board for the current study (Reference: 032774; Appendix B).

The PRaCTICED Trial

The current study constituted a quantitative analysis of the data collected as part of the PRaCTICED trial (Barkham et al., 2021; Saxon et al., 2017). The trial was embedded within the local Improving Access to Psychological Therapies (IAPT) service and involved randomisation of participants, meeting diagnostic criteria for moderate or severe depression, to PCET or CBT. Five-hundred and ten clients, aged 18 years and older, were randomised into the trial and received therapy from 50 experienced therapists.

Exclusion criteria in the trial included: presence of an organic condition, diagnosis of a personality disorder, bipolar disorder, psychosis or schizophrenia, drug or alcohol dependence, a long-term physical health condition, or elevated clinical risk. Participants meeting inclusion to the trial were invited to a second stage screening interview, at which they completed the Clinical Interview Schedule Revised (CIS-R) to evaluate the presence of moderate or severe depression (Lewis et al., 1992).

The primary outcome measure was the PHQ-9 (Patient Health Questionnaire; Spitzer et al., 1999). Secondary sessional measures comprised the GAD-7 (Generalised Anxiety Disorder; Spitzer et al., 2006) and WSAS (Work and Social Adjustment Scale; Mundt et al., 2002). Additional secondary outcome measures included the CORE-OM (Clinical Outcomes in Routine Evaluation – Outcome Measure; Evans et al., 2002), BDI-II (Beck Depression Inventory, Second Edition; Beck et al., 1996) and the EQ-5D-5L (The EuroQol Group, 1990), which were completed at baseline and 6- and 12-months after randomisation. A range of client and therapist demographic and clinical variables were also collated. Further details on the trial participants, procedure or treatments can be found in the original trial papers (Barkham et al., 2021; Saxon et al., 2017).

Treatment

Treatment was delivered up to a maximum of 20 sessions. Those randomised to receive PCET (Hill, 2011; Pearce et al., 2012; Sanders & Hill, 2014) completed a form of person-centred experiential therapy for depression derived from humanistic approaches. The modality of PCET is termed Counselling for Depression (CfD) in the IAPT model but this latter term does not convey the theoretical basis for the intervention and becomes confused with the general phrase counselling for depression. Hence, the specific term identifying its theoretical base is now used (Murphy, 2019). All PCET therapists completed training prior to the trial commencement; including a five-day taught module and 80 hours supervised practice.

The comparative treatment was CBT, combining techniques from Beckian cognitive therapy (A. T. Beck et al., 1979; J. S. Beck, 2011) and behavioural activation (Martell et al., 2001). All therapists completed training in the CBT approach as part of high intensity practitioner IAPT training, as well as additional regular supplementary training provision. All client sessions were audio-recorded for the purposes of establishing treatment fidelity and adherence to therapeutic protocols, competency monitoring and as a potential data source for research focussed on furthering understanding of patient outcomes.

Outcome Measures

The sessional PHQ-9 (Appendix C) was utilised in the current study to identify client sessions that are NOT. The PHQ-9 is a nine item, self-report screening measure, which asks the respondent to rate, on a four-point Likert scale, how often they have been affected by each of the symptoms listed as diagnostic criteria for depression according to the DSM-5 (Diagnostic and Statistical Manual; American Psychiatric Association, 2013). Total scores can range from 0 to 27, with scores of 10 and over suggesting the presence of clinical levels of depression. The measure has been found to have high sensitivity (92%) and specificity (80%) at detecting clinical levels of depression (Gilbody et al., 2007), and has been found to have good construct validity and internal reliability (Kroenke et al., 2001).

The current study used the BDI-II and CORE-OM as outcome measures to attain independence from the PHQ-9 data used to establish NOT session status. The BDI-II (Appendix D) is a 21-item measure of the symptoms of depression. A total score is derived by summing the highest rating for each of the 21 symptoms, which are rated on a scale of 0 to 3. The BDI-II has been found to have good internal consistency, retest reliability and good sensitivity and specificity for detecting depression, dependent on the clinical characteristics of the sample (Wang & Gorenstein, 2013). The CORE-OM (Appendix E) is a 34-item self-report instrument assessing clinical outcomes across four domains: subjective well-being, symptoms, functioning, and risk. A domain score and total composite score is calculated by summing the responses to each item. The measure has been found to have good internal and test–retest reliability and good sensitivity to change (Evans et al., 2002).

Participant Selection

Sample Size

The use of the FIS tool as a therapy session process measure is a novel methodology. Therefore, ideal sample size for the study was guided by generalised rule of thumb principles, however, was ultimately constrained by the participant selection criteria which were designed to

increase the sensitivity of the study. A sample size calculation estimated that a minimum sample size of 84 clients would be required for sufficient statistical power to evidence a medium effect size, with 80% power and a critical *p*-value of .05 (Cohen, 1992). Alternative guidance suggests a minimum ratio of participants to predictor variables of 10:1 for regression analyses (Harrell, 2001), suggesting that a minimum sample size of 30 participants would be adequate.

Participant Selection Criteria

The session selection procedure was completed by a researcher (MSB) independent from the FIS rating process, to minimise bias by ensuring that FIS raters were blinded to client and therapist demographic and outcome variables. The following inclu

sion criteria were used to identify therapy sessions from the PRaCTICED dataset for the current study.

1. The client completed a minimum of four sessions.
2. The client remained in the same therapy arm for the duration of treatment.
3. The client completed the PHQ-9 measure within the first session to identify baseline severity, and an additional three further PHQ-9 measures to establish NOT session status.
4. The client baseline PHQ-9 scores were equal to or greater than a score of 13.
5. The client session had been identified as NOT using the algorithm developed by Delgadillo et al. (2018).
6. A maximum of five client sessions treated by the same therapist were considered for inclusion. If more than five sessions in the available NOT session pool were treated by the same therapist, five of the sessions were randomly selected for inclusion.

Identifying "Not on Track" Sessions

To identify client sessions that were NOT, the expected treatment response (ETR) algorithm developed by Delgadillo et al. (2018) was utilised. The procedure involved an automated system which used growth curve models to identify average treatment trajectories for clients, based on baseline clinical severity (using the PHQ-9 scores at the start of treatment). Distinct algorithms

were developed for nine baseline depression severity groups using routine outcome data from a large treatment dataset to estimate expected treatment response for each group. Growth curve modelling identified average treatment trajectories, representing a loglinear trend in reduction of depressive symptoms over the course of treatment. Confidence intervals were applied to the treatment trajectories.

The algorithm was applied to the PRaCTICED trial data. Sessional PHQ-9 scores found to exceed the upper boundary of treatment trajectories, identified clients that were NOT towards recovery. As the PHQ-9 measure was completed at the start of a therapy session and was rated in reference to depressive symptoms over the previous two weeks, a NOT signal indicated the previous session was NOT towards recovery. The first and last sessions in treatment were excluded. In the instance that a client was identified as having multiple sessions that were NOT, the first NOT session with an available recording was identified. The selection was stratified so that within each treatment, half of the cases were NOT sessions followed by another NOT session, and the other half were NOT sessions that were followed by a session where the client was back on track towards recovery. The NOT sessions that were not selected for the study sample were used as training material to develop the FIS rating procedure and increase consistency between raters.

The sample of NOT client sessions included 68 cases ($n = 40$ CBT; $n = 28$ PCET) by 27 therapists ($n = 17$ CBT; $n = 10$ PCET). The audio-recordings of nine NOT sessions were not available, or the quality was too poor to reliably rate FIS. Therefore, 59 NOT client sessions were included in the final study sample.

Participants

The sample comprised 59 clients aged between 19-73 (mean $= 39.9$; $SD = 12.13$). Clients were assessed as meeting criteria for moderate ($n = 20$) and severe ($n = 38$) depression on the CIS-R, with one missing client rating. Twenty-eight (47.5%) of the clients in the sample received PCET and 31 (52.5%) received CBT. Treatment duration was an average of 15 sessions ($SD = 4.85$), with no significant differences in session length between PCET (mean $= 14.8$, $SD = 5.81$) and CBT

(mean $= 15.45$, $SD = 3.82$; $t = -0.48$, $p = .64$). Therapy was delivered by 24 therapists, 19 female therapists/counsellors (79.2%) and five male therapists/counsellors (20.8%), with an average of 13.1 years ($SD = 6.48$) of therapy experience and 9.4 years ($SD = 3.5$) in their current role. Additional client and therapist characteristics can be found in Table 2.

Table 2

Demographic Descriptors of the Client and Therapist Sample

Demographic Variable	*n*	%
Client Ethnicity		
White British	56	94.9
Other	3	5.08
Client Employment Status		
Employed/in education	35	59.3
Unemployed	14	23.7
Missing data	10	17.0
Therapist Age Band		
30-39	6	25.0
40-49	4	16.7
50-59	9	37.5
60+	5	20.8

Facilitative Interpersonal Skills

The FIS rating manual (Anderson & Patterson, 2013) was the primary measure used in the current study, measuring eight items relating to a therapist's skill in fostering a facilitative environment (Table 1; Appendix F). The FIS manual (Anderson & Patterson, 2013) was adapted to be applied to rating therapy sessions. The eight FIS items have been confirmed to represent a single

underlying construct and are internally consistent (McClintock et al., 2012). The FIS performance task (Anderson et al., 2009, 2019; Anderson & Patterson, 2013) and previous research rating real therapy sessions (Uhlin, 2011) have achieved good levels of inter-rater reliability. The FIS performance task has been found to have concurrent validity, and is strongly correlated with measures of social skills, empathy and sociability (Anderson & Patterson, 2013). The two forms of the FIS: the performance task and in-session rating method have been found to be moderately related ($r = .49$; Uhlin, 2011).

Training

To maximise precision and agreement, two expert raters (CB and AF) jointly rated therapist FIS. The raters were provided with extensive training prior to commencement of FIS rating. The researchers had access to the PCET and CBT treatment manuals to familiarise with therapy characteristics with potential influence on the interpersonal responsiveness of the therapist. Researchers also accessed consultation from an expert therapist (MH), to ensure that interpersonal features of the sessions were understood correctly in the context of the intervention. Training on the FIS rating procedure was facilitated by experienced FIS researchers (KDJ and KS). Two initial training sessions aimed to increase familiarity in rating therapist responses to the simulated client interactions from the performance task. Training involved discussion of the FIS rating procedure, detailed analysis of each FIS item in the coding manual, and live collaborative rating of stimulus videos. Between training sessions, raters independently coded a selection of stimulus videos, and scores were then calibrated until consensus between raters was reached.

Once the initial training was completed, a further two training sessions focused on applying the FIS methodology to in-session therapy tapes. In-session therapy recordings had been rated to consensus by KDJ and KS and were compared and discussed with those allocated by the primary raters. Following completion of training, the two FIS raters independently rated a sample of calibration tapes until a set intraclass correlation coefficient (ICC) criterion was reached. A two-way mixed effects, average-rater ($n = 2$), absolute agreement ICC model was used, with a suggested

criterion level of >.70, indicating "moderate to good" levels of inter-rater reliability (Koo & Li, 2016). Following independent rating of five training tapes from the PRaCTICED dataset, an ICC of .71 (95% CI: [.57, .81]) was reached.

Rating Procedure

The study sample recordings were randomly allocated into blocks of 10 by an independent researcher (GH). Each block contained an equal number of PCET and CBT sessions to ensure that practice and/or fatigue effects did not disproportionately influence ratings of one therapeutic orientation. The raters were blinded to client demographic, clinical, and outcome data until the rating of the FIS was completed, although some characteristics were discernible from the audio-recordings. Within each block, two therapy tapes (one from each therapeutic orientation) were randomly selected to be independently rated by both primary raters to prevent rater drift. A calibration meeting took place after each block of ratings to discuss issues arising from the rating procedures, and to calibrate the double-rated sessions to ensure consistent application of the manual. Anonymised minutes of the calibration meetings can be found in Appendix G. For the sample of double-rated tapes, the calibrated FIS scores were used in the statistical analyses.

Each client session was divided into three segments, as shorter sections were easier to hold in mind to consistently apply the manual and capture the potential variability in FIS skill use over the session duration. The vignettes in the FIS performance task were less than two minutes in duration and elicited short therapist responses, therefore splitting sessions into segments was more comparable to the original procedure than rating whole 45-50 minute sessions (e.g. Uhlin, 2011). The sessions were variable in length but were generally around 50 minutes, therefore sessions were split into two 17 minutes segments, with a final segment covering the remainder of the recording. Each segment was rated consistently by the same researcher.

The coding procedure involved the rater listening to a segment and noting specific key moments, before listening to the key moments for a second time. The FIS manual (Appendix F) was then used to assign a score for each of the eight FIS domains for the segment. Within the manual, a

five-point Likert scale was used to rate each item from one (skill deficit) to five (optimal presence of the skill). A score of three was initially assumed, and raters adjusted the score to match qualitative descriptors of each level of competence at applying that skill. Half marks were allocated when raters felt a score between two descriptors was most representative of therapist skill. The final FIS domain was rated using the "problem responsiveness" criteria developed for use with more "benign", less interpersonally difficult, sessions than the initial performance task stimulus videos (developed by De Jong and colleagues, Appendix H). The original, rupture-repair responsiveness item was seen as a specific form of problem responsiveness, and as such raters highlighted therapists' sensitivity and responsiveness to problems arising within the session, including, but not exclusive to, ruptures in the therapeutic relationship. It was noted that real-world therapy sessions may consist of multiple "problems", and therefore multiple problem responsiveness scores could be allocated for each segment.

For the purposes of the current study, the domain scores were summed across the three segments to compute a total score for each of the eight FIS domains. If multiple ratings were made in the problem responsiveness domain, the average of the problem responsiveness ratings was first taken for each segment, before summing into a total domain score. The domain scores were then summed to create a total FIS score for each dyad session.

Interrater Agreement

In total, 13.6% ($n = 8$) of the sample tapes received two independent ratings. A two-way mixed effects, average-rater ($n = 2$), absolute agreement ICC level of .79 (95% CI: [.72 , .84]) was reached, reflecting "good" levels of inter-rater agreement (Koo & Li, 2016). Furthermore, independent-samples t-tests revealed no significant differences in the FIS domain and total scores between raters (Appendix I).

Statistical Analysis

Statistical analyses were conducted in Statistical Package for Social Sciences (SPSS; Version 26.0; IBM Corp., 2019). The current study investigates previously untested hypotheses; therefore,

the analyses were exploratory, and aimed to determine proof-of-concept of the study methodology and in-session rating of FIS.

Missing outcome measures in the dataset were identified and dealt with using complete case analysis with covariate adjustment, following the recommendations of Groenwold et al. (2012). Prior to conducting primary analyses, independent-samples *t*-tests and correlation analyses were performed to identify important variables associated with outcome that would require controlling for in the main analyses. Therapy orientation, baseline depression severity, client age, gender and ethnicity were examined for relationships with the change in CORE-OM or BDI-II scores from baseline to 12 months. All significant predictors were retained as covariates in the primary analyses.

Given the hierarchical structure of the data, with FIS scores from a client-therapist dyad ($n = 59$), nested within therapists ($n = 24$), a multilevel modelling approach would have been beneficial. However, simulation studies have demonstrated a sample size of less than 30 at the highest level (e.g., therapist level) increases the likelihood of biased estimations (Maas & Hox, 2005). It was therefore decided the sample size and structure of data in the current study was insufficient to reliably employ multilevel modelling procedures. Single-level regression analyses were therefore used to model the data. The results of the analyses must be interpreted with caution due to the violation of the assumption of independence of observation which has the potential to inflate the Type 1 error rate and estimates of effect size.

To examine H1, the predictive ability of therapists' FIS in explaining variability in client outcome, the current study used a single-level hierarchical linear regression model. The dependent variables in the analyses were the BDI-II and CORE-OM scores at 12 months. Predictor variables were added to the model in two blocks. In the initial block, the baseline BDI-II or CORE-OM scores were added, as well as additional covariates. The FIS total score was then added in a second block, to assess the unique proportion of variance in outcome explained.

To examine H2, positing an interaction between FIS skills and therapy orientation on treatment outcome, two moderated regression analyses were conducted. The analyses were conducted using the

PROCESS macro for SPSS (Version 3.5.2; Hayes, 2020). Again, the dependent variables in the analyses were BDI-II and CORE-OM scores at 12-month timepoints, controlling for baseline BDI-II and CORE-OM scores and additional covariates. As the moderator variable of interest (therapy orientation) was dichotomous, the FIS total score variable was mean centred to enable computation of the interaction. An interaction term, between therapy orientation (PCET or CBT) and FIS total score was added to the model to examine the unique proportion of variance explained and determine if therapy orientation was a significant moderator of the FIS-outcome relationship.

Subgroup analyses in the PRaCTICED trial (Barkham et al., 2021) revealed differences in treatment effect dependent on depression severity groups. Therefore, exploratory post-hoc analyses were completed to examine if the impact of FIS skills on treatment outcome varied between baseline depression severity groups. Two-way full-factorial analysis of covariance (ANCOVA) models were conducted, with change in BDI-II and CORE-OM scores between baseline and 12-month follow-up timepoints (12-month score minus baseline score) as the dependent variable. For the current analyses, total FIS scores were stratified into two groups (high and low FIS therapists), based on a mean split. Baseline depression severity and FIS category were included in the model as full-factorial predictor variables to assess for significant interactions. Significant interactions were followed-up with post-hoc simple main effects analyses.

<center>**Results**</center>

Descriptive Data

FIS ratings were completed for 59 therapy sessions. Within a possible range of 24-120, the mean FIS total score in the sample was 90.1 ($SD = 16.6$, range = 53-116). Means and standard deviations of the FIS domain scores are presented in Table 3. The client sessions used for FIS rating ranged from the 2nd to the 17th session (mean = 7.49, $SD = 4.10$). Across the full sample, the average improvement in CORE-OM and BDI-II scores was 6.22 ($SD = 6.91$, $n = 39$) and 12.34 ($SD = 11.83$, $n = 38$) points, respectively, between baseline and 12-month time-points. Subgroup differences in treatment outcome can be found in Table 4.

Missing Outcome Data

There was a high rate of missing outcome data in the current study, with 34% of the CORE-OM measures ($n = 20$), and 36% ($n = 21$) of the BDI-II measures missing at the 12-month timepoint. Missing data on the CORE-OM and BDI-II was disproportionately found in the severe depression level group (36.8% and 39.5%, respectively) compared to the moderate depression group (30.0% and 30.0%, respectively). Gender also seemed to be associated with data missingness, with a higher rate of missing data on the CORE-OM and BDI-II in male clients (42.9% and 42.9%, respectively) compared to female clients (28.9% and 31.6%, respectively). Completed cases were used in further analyses, with gender and depression severity as covariates to reduce the bias arising from missing data. There were no significant differences in the mean FIS total scores between the participants with completed CORE-OM measures (mean = 89.18, $SD = 16.13$, $n = 39$) and those with missing data (mean = 91.91, $SD = 17.66$, $n = 20$; $t(57) = 0.60$, $p = .553$). There were also no significant differences in the mean FIS total scores between those with completed BDI-II measures at 12 months (mean = 88.71, $SD = 16.07$, $n = 38$) and those with missing measures (mean = 92.63, $SD = 17.53$, $n = 21$; $t(57) = .87$, $p = .389$).

Table 3

FIS Domain and Total Scores in PCET and CBT

	Mean (*SD*)		
	PCET	**CBT**	**Total**
FIS Domains and Total Score	**(*n* = 28)**	**(*n* = 31)**	**(*n* = 59)**
Verbal Fluency	11.73 (2.23)	11.47 (31.11)	11.59 (2.71)
Hope/Positive Expectations	9.73 (1.41)	11.53 (1.79)	10.68 (1.85)
Persuasiveness	10.38 (1.59)	11.35 (2.18)	10.89 (1.97)
Emotional Expression	12.82 (2.43)	11.19 (3.07)	11.97 (2.88)
Warmth	13.18 (1.91)	10.98 (2.40)	12.03 (2.43)
Empathy	12.05 (2.05)	10.42 (2.04)	11.19 (2.19)
Alliance Bond Capacity	11.82 (2.61)	10.97 (3.06)	11.37 (2.87)
Problem Responsiveness	10.54 (2.60)	10.25 (2.98)	10.39 (2.79)
FIS Total Score	92.25 (14.25)	88.17 (18.42)	90.11 (16.56)

Note: FIS = Facilitative Interpersonal Skills, PCET = person-centred experiential therapy, CBT = cognitive behavioural therapy, SD = Standard Deviation.

Preliminary Analyses

Independent-samples *t*-tests revealed no significant differences in treatment outcome (change in CORE-OM and BDI-II scores from baseline to 12-month timepoints) dependent on baseline depression severity, client gender or client ethnicity (Table 4). Significant associations were observed between client age and change in CORE-OM ($r = .371$, $p = .020$) and BDI-II ($r = .491$, $p = .002$) scores from baseline to 12 months. Furthermore, outcome significantly differed between

therapy orientation, with greater improvements on the CORE-OM and BDI-II outcome measures observed in CBT, when compared to PCET, between baseline and 12-month timepoints (Table 4). Client age and therapy orientation were retained as covariates in further analyses.

Table 4

Differences in Outcome by Therapy Orientation, Depression Severity, Client Gender and Client Ethnicity

Therapy Outcome	Mean (*SD*)		Independent-Samples *T*-Test
Therapy Orientation	PCET (*n* = 18)	CBT (*n* = 21)	
Δ CORE-OM	3.50 (6.45)	8.56 (6.56)	*t* (37) = 2.42, *p* = .021
Δ BDI-II	7.94 (11.20)	16.10 (11.28)	*t* (36) = 2.23, *p* = .032
Baseline Depression Severity	Moderate (*n* = 14)	Severe (*n* = 24)	
Δ CORE-OM	4.33 (7.04)	7.22 (6.89)	*t* (36) = 1.24, *p* = .224
Δ BDI-II	8.29 (11.39)	14.57 (11.96)	*t* (35) = 1.58, *p* = .124
Client Gender	Female (*n* = 27)	Male (*n* = 12)	
Δ CORE-OM	6.63 (7.39)	5.29 (5.87)	*t* (37) = -.55, *p* = .583
Δ BDI-II	13.08 (12.99)	10.42 (9.07)	*t* (36) = -.64, *p* = .527
Client Ethnicity	White (*n* = 37)	BAME (*n* = 2)	
Δ CORE-OM	6.63 (6.97)	4.26 (7.69)	*t* (37) = -.41, *p* = .687
Δ BDI-II	12.31 (12.04)	11.00 (9.90)	*t* (36) = - .15, *p* = .882

Note: Δ CORE-OM and Δ BDI-II refers to change in scores between baseline and 12-month time-points. CORE-OM = Clinical Outcomes in Routine Evaluation Outcome Measure, BDI-II = Beck Depression Inventory, *SD* = standard deviation, BAME = Black, Asian, and Minority Ethnic group.

Primary Analyses

The Role of FIS in Predicting Therapy Outcome

Hierarchical multiple regression models were conducted to determine if therapist FIS improved the prediction of outcome of psychotherapy (change in CORE-OM and BDI-II scores between baseline and 12 months) over and above therapy orientation, client age, client gender and baseline depression severity (Table 5). Assumptions of linearity, homoscedasticity, multicollinearity, and normality of residuals were met (Appendix J). As previously stated, some therapists were represented multiple times in the study sample; therefore, the assumption of independence of observations was not met, requiring cautious interpretation of significant findings.

When predicting treatment outcome at 12 months on the CORE-OM measure, the full regression model, including baseline CORE-OM score, depression severity, client age, client gender, therapy orientation, and FIS total score (Table 5, Model 2) was statistically significant ($R^2 = .42$, $F(6, 31) = 3.726$, $p = .007$, adjusted $R^2 = .307$). The addition of FIS total score to the model accounted for an additional 0.8% of the variance in 12-month CORE-OM scores and did not significantly improve the model fit (F change $(1,31) = 0.44$, $p = .513$, R^2 change $= .008$), contrary to H1.

When predicting treatment outcome at 12 months on the BDI-II measure, the full regression model, including baseline BDI-II score, depression severity, client age, client gender, therapy orientation (Table 5, Model 2) was also statistically significant ($R^2 = .58$, $F(6, 30) = 7.27$, $p < .001$, adjusted $R^2 = .511$). Similarly, the addition of FIS total score to the model accounted for an additional 1.1% of the variance in 12-month BDI-II scores and did not significantly improve the model fit (F change $(1,30) = 0.79$, $p = .381$, R^2 change $= 0.011$), contrary to H1.

Table 5

Summary of Hierarchical Regression Analyses for Variables Predicting CORE-OM and BDI-II

Scores at 12 months

	Model 1				Model 2			
Variable	*B*	*SE B*	*β*	*p*	*B*	*SE B*	*β*	*p*
Predicting CORE-OM Scores at 12 months (n = 38)								
Baseline CORE-OM	0.33	0.24	0.21	.174	0.37	0.25	0.24	.146
Depression Severity	-.1.53	1.98	-0.11	.444	-1.59	2.00	-0.12	.433
Client Age	0.26	0.08	0.43	.004	0.27	0.08	0.44	.004
Client Gender	-0.91	2.08	-0.06	.663	-0.22	2.34	-0.02	.926
Therapy orientation	-5.34	1.82	-0.41	.006	-5.04	1.89	-0.39	.012
FIS Score					0.04	0.06	0.10	.513
R²			0.41				0.42	
F for change in R²			4.46	.003			0.44	.513
Predicting BDI-II Scores at 12 months (n = 37)								
Baseline BDI-II	0.66	0.22	0.44	.005	0.71	0.23	0.48	.004
Depression Severity	-4.81	3.62	-0.19	.194	-5.13	3.65	-0.20	.171
Client Age	0.59	0.14	0.51	< .001	0.60	0.14	0.52	< .001
Client Gender	0.95	3.30	0.04	.776	2.40	3.69	0.09	.520
Therapy orientation	-9.19	3.04	-0.37	.005	-8.31	3.20	-0.33	.014
FIS Score					0.93	0.10	0.12	.381
R²			0.58				0.59	
F for change in R²			8.62	< .001			0.79	.381

Note. CORE-OM = Clinical Outcomes in Routine Evaluation Outcome Measure, BDI-II = Beck

Depression Inventory, FIS = Facilitative Interpersonal Skills, SE = Standard Error

Therapy Orientation Moderator Analysis

A moderator regression analysis was run to assess the statistical significance of the interaction

term between FIS total score and therapy orientation (PCET or CBT) in predicting CORE-OM

scores at 12-months, controlling for baseline CORE-OM scores, depression severity, client age and client gender. The model predicted treatment outcome on the CORE-OM (F (7, 30) = 3.08, p = .014, R^2 = .418; Table 6), however there was a non-statistically significant moderator effect of therapy orientation, as evidenced by the addition of the interaction term explaining an additional 0.72% of the total variance (p = .546). A similar pattern was observed when the BDI-II was examined as the dependent variable. The model (Table 6) was a significant predictor of BDI-II scores at 12 months (F (7, 29) = 6.07, p < .001, R^2 = . 595), however the interaction between FIS total score and therapy orientation was not significant (p = .514), accounting for only an additional 0.61% of the variability in treatment outcome.

Table 6

Summary of Moderator Regression Analysis for Variables Predicting CORE-OM and BDI-II Scores at 12 months

Variable	Predicting CORE-OM Scores at 12 months (n = 38)			Predicting BDI-II Scores at 12 months (n = 37)		
	B	SE B	p	B	SE B	p
Baseline CORE-OM/BDI-II	0.30	0.25	.255	0.59	0.23	.016
Depression Severity	-1.09	2.16	.619	-3.87	3.96	.337
Client Age	0.26	0.09	.005	0.58	0.14	< .001
Client Gender	-0.75	2.21	.737	0.85	3.48	.809
Therapy Orientation	-5.35	1.86	.007	-9.30	3.10	.006
FIS Score	0.04	0.09	.630	0.13	0.14	.346
FIS*Therapy Orientation	-0.07	0.12	.546	-0.13	0.19	.514

Note. CORE-OM = Clinical Outcomes in Routine Evaluation Outcome Measure, BDI-II = Beck Depression Inventory, FIS = Facilitative Interpersonal Skills, *SE* = Standard Error

Subgroup Analyses

Post-hoc exploratory analyses examining the impact of FIS skills on treatment outcome, considering the interaction with baseline depression severity groups were conducted using ANCOVA. Testing of the assumptions of ANCOVA, including linearity of relationships, homogeneity of regression slopes, homoscedasticity, homogeneity of variances and normality of residuals, was completed (Appendix K). As previously mentioned, the sample data violates the independence of observations assumption, and therefore the findings must be interpreted cautiously.

Table 7

Means and Adjusted Means for Change in CORE-OM and BDI-II Scores by Baseline Depression Severity for High and Low FIS Therapists

	Moderate Depression Severity		Severe Depression Severity	
	Low FIS	High FIS	Low FIS	High FIS
Δ CORE-OM				
M (*SD*)	1.08 (6.44)	-8.38 (4.31)	-11.38 (5.80)	-4.24 (6.14)
M$_{adj}$ (*SE*)	0.06 (2.28)	-7.18 (1.92)	-11.09 (1.78)	-4.71 (1.54)
Δ BDI-II				
M (*SD*)	-0.83 (12.53)	-13.88 (6.75)	-21.50 (9.92)	-9.23 (10.83)
M$_{adj}$ (*SE*)	-3.17 (3.53)	-11.00 (3.01)	-21.03 (2.78)	-10.28 (2.52)

Note: Δ CORE-OM and Δ BDI-II refers to change in scores between baseline and 12-month time-points. Minus scores indicate improvement in symptoms. *M* = mean scores, *M*$_{adj}$ = adjusted means, FIS = Facilitative Interpersonal Skills, SD = Standard Deviation, SE = Standard Error.

A two-way full factorial ANCOVA revealed a statistically significant interaction between depression severity and FIS category on change in CORE-OM score from baseline to 12-month timepoint, whilst controlling for client age, client gender, and therapy orientation ($F(1,31) = 11.61$, $p = .002$, Figure 1, Table 7). Therefore, an analysis of simple main effects for FIS category and depression severity was performed with a Bonferroni adjusted critical *p*-value ($p < .025$) to account

for multiple comparisons. The effect of FIS category on change in CORE-OM scores in the moderate depression severity group was not significant ($F(1, 31) = 5.41$, $p = .027$). However, in the severe group, the difference in adjusted mean change in CORE-OM score was significantly greater in the low FIS group, than the high FIS group ($F(1,31) = 6.40$, $p = .017$). Within the low FIS group, those with severe baseline depression severity had significantly greater reductions in the CORE-OM measure than those with moderate depression severity ($F(1, 31) = 15.19$, $p < .001$). However, in the high FIS group, there were no significant differences in CORE-OM change scores between baseline depression severity groups ($F(1,31) = .973$, $p = .030$).

Figure 1

Two-Way ANCOVA Showing Adjusted Mean Change in CORE-OM Score from Baseline to 12 Months by Baseline Depression Severity and FIS Category
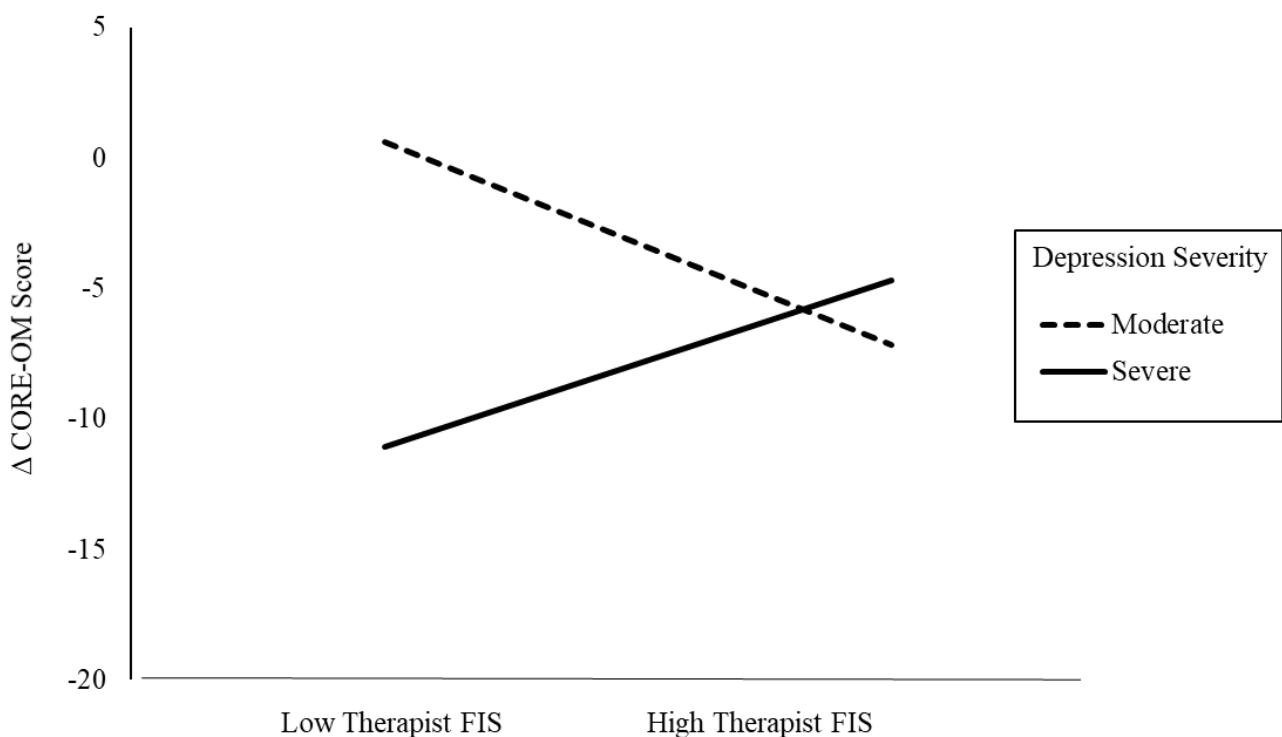
Note: Δ CORE-OM score refers to adjusted mean change in CORE-OM scores between baseline and 12 months, controlling for client age, client gender and therapy orientation. CORE-OM = Clinical Outcomes in Routine Evaluation Outcome Measure. Minus scores indicate improvement in symptoms. FIS = Facilitative Interpersonal Skills.

Table 8

Adjusted Means and Standard Deviations for Change in CORE-OM and BDI-II Scores by Baseline

Depression Severity for High and Low FIS Therapists within PCET and CBT

Depression Severity	FIS Category	Δ CORE-OM		Δ BDI-II	
		CBT	PCET	CBT	PCET
Moderate	Low FIS	-4.78 (4.10)	2.46 (2.70)	-10.23 (7.11)	1.42 (4.47)
	High FIS	-7.76 (2.35)	-2.99 (3.93)	-12.60 (4.08)	-6.10 (6.50)
Severe	Low FIS	-13.68 (2.00)	-5.79 (3.17)	-23.70 (3.44)	-18.24 (5.24)
	High FIS	-1.33 (2.78)	-5.52 (1.49)	-7.45 (5.53)	-10.00 (2.46)

Note: Δ CORE-OM and Δ BDI-II refers to change in scores between baseline and 12-month time-points. Minus scores indicate improvement in symptoms. FIS = Facilitative Interpersonal Skills. CBT = Cognitive Behavioural Therapy. PCET = Person Centred Experiential Therapy. SD = Standard Deviation.

Post-hoc exploratory analyses replicated the above procedure separately for the two therapy orientations to examine if the interaction was observed in PCET, CBT, or across both therapy orientations. For the clients receiving PCET, a two-way full-factorial ANCOVA revealed a nonsignificant interaction between FIS category and depression severity on client change in CORE-OM scores, whilst controlling for client age and gender ($F(1, 12) = 0.84$, $p = .378$). A significant interaction between FIS category and depression severity was, however, observed in the clients receiving CBT ($F(1, 14) = 7.17$, $p = .018$, Table 8, Figure 2). Analysis of simple main effects for the CBT group only, revealed that the effect of FIS category on change in CORE-OM scores in the moderate depression severity group was not significant ($F(1,14) = .360$, $p = .558$). However, in the severe group receiving CBT, the difference in adjusted mean change in CORE-OM score was

significantly greater in the low FIS group, than the high FIS group ($F(1,14) = 12.18$, $p = .004$). Within both the low ($F(1,14) = 4.09$, $p = .063$) and high ($F(1,14) = 3.37$, $p = .088$) FIS groups who were randomised to CBT treatment, there were no significant differences in treatment outcome between those with moderate and severe depression severity. The findings of the current analysis must be interpreted with caution due to the limited number of clients within some of the subgroups (Table 9).

Figure 2

Two-way ANCOVA Showing Adjusted Mean Change in CORE-OM Score from Baseline to 12 Months by Baseline Depression Severity and FIS Category for the CBT Group Only
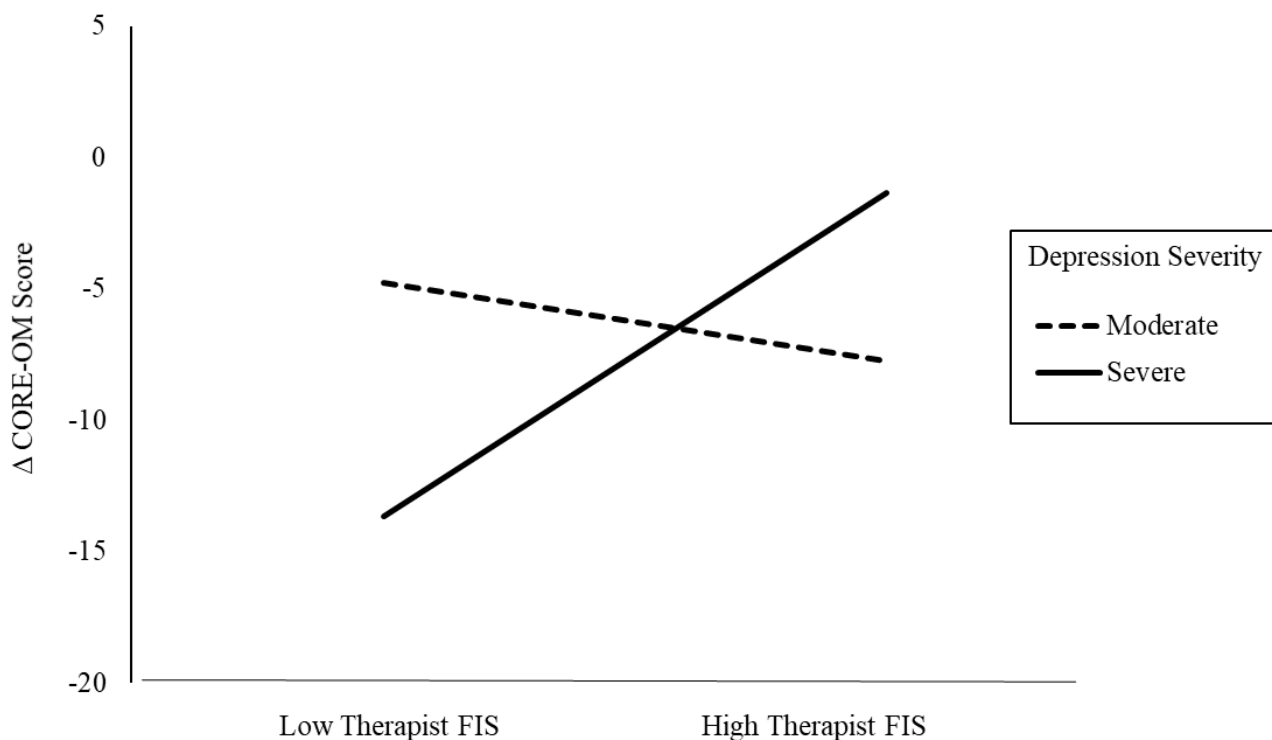
Note: Δ CORE-OM score refers to adjusted mean change in CORE-OM scores between baseline and 12 months, controlling for client age and gender. Minus scores indicate improvement in symptoms. FIS = Facilitative Interpersonal Skills.

A similar pattern of results was observed in the BDI-II outcome measure. A two-way full factorial ANCOVA revealed a statistically significant interaction between depression severity and FIS category on change in BDI-II score from baseline to 12-months, whilst controlling for client age, client gender, and therapy orientation ($F(1, 30) = 8.49$, $p = .007$, Figure 3, Table 7). Simple main effects analyses were again conducted. The effect of FIS category on BDI-II change score in the moderate depression severity group was not significant ($F(1, 30) = 2.62$, $p = .116$). However, in the severe depression group the difference in adjusted mean change in BDI-II score was significantly greater in the low FIS group, than the high FIS group ($F(1,30) = 7.015$, $p = .013$). Within the low FIS group, those with severe baseline depression severity had significantly greater reductions in the BDI-II measure than those with moderate depression severity ($F(1,30) = 16.054$, $p < .001$). However, in the high FIS group, there were no significant differences in BDI-II change scores between baseline depression severity groups ($F(1,30) = .032$, $p = .859$).

Table 9

Table of Frequencies of Clients Belonging to the Study Subgroups

Depression Severity	FIS Category	Δ CORE-OM		Δ BDI-II	
		CBT	PCET	CBT	PCET
Moderate	Low FIS	2	4	2	4
	High FIS	6	2	6	2
Severe	Low FIS	8	2	8	2
	High FIS	4	10	3	10

Note: Δ CORE-OM and Δ BDI-II refers to change in scores between baseline and 12-month timepoints. Minus scores indicate improvement in symptoms. FIS = Facilitative Interpersonal Skills.

As with the CORE-OM, post-hoc exploratory analyses replicated the two-way ANCOVAs predicting BDI-II scores at 12 months separately for the two therapy orientations. For the clients randomised to PCET, a nonsignificant interaction between FIS category and depression severity on client change in BDI-II scores, whilst controlling for client age and gender, was observed ($F(1,12) = 2.34$, $p = .152$). A nonsignificant interaction between FIS category and depression severity was also observed in the clients receiving CBT ($F(1, 13) = 3.32$, $p = .092$). Adjusted mean change in BDI-II scores can be found in Table 8. The sample sizes in some of the groups are small ($n = 2$) and therefore the findings must be interpreted with caution (Table 9).

Figure 3

Two-way ANCOVA Showing Adjusted Mean Change in BDI-II Score from Baseline to 12 Months by Baseline Depression Severity and FIS Category
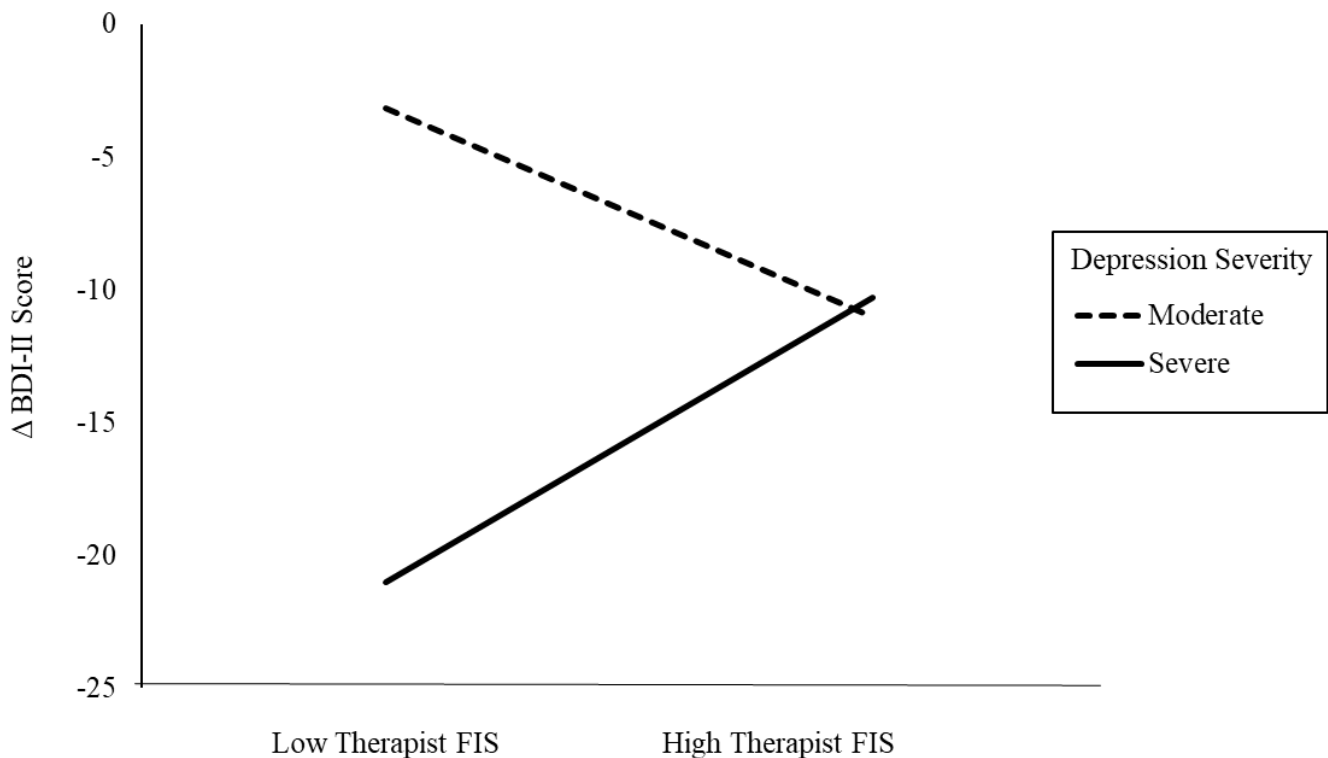
Note: Δ BDI-II score refers to adjusted mean change in BDI-II scores between baseline and 12 months, controlling for client age, client gender and therapy orientation. Minus scores indicate improvement in symptoms. BDI-II = Beck Depression Inventory. FIS = Facilitative Interpersonal Skills.

<center>**Discussion**</center>

Summary of Findings

The primary aim of the current study was to examine the predictive ability of therapists' FIS during NOT therapy sessions in client outcome at 12-months post-randomisation to treatment. Across the full sample of clients included in the study, regression analyses revealed a non-significant influence of FIS on therapy outcome at 12 months, assessed on the CORE-OM and BDI-II. The findings are contrary to the extant body of literature demonstrating the ability of the FIS performance task in predicting client therapeutic outcomes (Anderson et al., 2009; Anderson, Crowley et al., 2016; Anderson, McClintock et al., 2016), as well as the wider psychotherapy process literature which posits robust relationships between therapist interpersonal abilities and therapy outcome (Norcross & Lambert, 2018). However, the findings are consistent with the only other extant research rating FIS in the context of client sessions, which reported a non-significant association between FIS and psychotherapy outcome (Uhlin, 2011). A potential explanation for the null finding may be the length of the treatment under investigation in the current study (15 sessions on average), as high- and low-rated FIS therapists reportedly have near equivalence in client outcomes over longer treatment durations (Anderson, McClintock et al., 2016).

The current study was the first to use the FIS methodology to rate and compare two psychotherapy orientations, namely PCET and CBT. The current study hypothesised that psychotherapy orientation may constitute a moderator of the relationship between FIS and outcome, with a stronger association between interpersonal skills and client symptom improvement in PCET, than in CBT. The hypothesis was not supported by the data. Within the same sample, Faulds (2021) did observe significant differences in the FIS domain items between CBT and PCET, in that CBT therapists were observed to demonstrate higher levels of hope than PCET therapists, and PCET therapists demonstrated higher capacity in emotional expression, warmth and empathy than CBT therapists. Despite the domain level differences in FIS, Faulds (2021) reported no significant difference in FIS skills at the composite level. The parity in FIS skills between therapies at the

composite level provides justification for the lack of an observed modulatory role of treatment in the FIS-outcome association.

The post-hoc subgroup analyses revealed an association between high and low FIS therapists and symptom outcome 12-months post-randomisation to treatment, which was observed to have a differential pattern dependent on client baseline depression severity. As such, the difference in the direction of associations between FIS and outcome between the moderate and severe depression severity groups may have masked the overall predictive ability of FIS in the entire sample. The observation of an interaction must be interpreted with caution as the analyses were conducted *post hoc* and may be a statistical artefact of fitting an interaction term in an underpowered dataset.

A trend in line with the study hypothesis, although one which does not reach significance, is shown in the clients with moderate depression. It was observed that clients with moderate baseline depression severity demonstrate more symptom improvement at 12 months when the NOT session was completed with a therapist with higher-than-average FIS, comparative to those seen by a therapist with lower-than-average FIS skills. This trend is in line with previous literature that has demonstrated that therapists with a higher capacity to create the facilitative interpersonal environment are linked to greater improvements in client outcomes than those with lower interpersonal capabilities (Anderson et al., 2009; Anderson, Crowley et al., 2016; Anderson, McClintock et al., 2016).

Contrary to the study hypothesis, individuals in the severe baseline depression group, where the NOT session was with a therapist demonstrating lower-than-average FIS showed a greater improvement in symptom measures over the course of treatment, compared to those with a high FIS therapist. Additional analyses revealed that this finding was particularly relevant for clients with severe depression receiving CBT. Speculative hypotheses can be proposed to make sense of the findings. It may be considered that the experience of severe depression increases clients' sensitivity, and perhaps intolerance, to the manifestations of a high degree of interpersonal responsiveness and emotionally expressive communication from the therapist. Otherwise, severely depressed clients

may struggle to connect with or relate to an empathic, hopeful, warm and emotionally rich interpersonal style, potentially due to their own symptom experiences of apathy, hopelessness and/or self-criticism. As such, it may be, that in an attempt to attune to clients' interpersonal communication, effective therapists down-regulate their FIS capabilities to bolster a connection. Otherwise, clients with severe depression may demonstrate a preference for therapies with less of an interpersonal focus, such as CBT, which employs tangible behavioural and cognitive strategies. An alternative explanation posits that therapists working with severely depressed clients, in the knowledge of clinical deterioration/non-improvement from the sessional PHQ-9 measures, may intentionally enhance the use of FIS in an attempt to bolster the therapeutic relationship and get the client "back on track" towards recovery. However, the focus on the alliance may limit the attention to other possible sources of impediment to therapeutic progress, such as areas measured in the assessment of signal cases instrument, including: difficulties with social support, readiness to change, diagnostic formulation, life events, and need for medication referral (Lambert, 2017; Probst et al., 2020). The findings highlight the potential role of interpersonal processes underlying the outcome of the wider PRaCTICED trial which reported that, for the severe group only, results favoured CBT to PCET in client outcome on the PHQ-9 at 12-month follow-up (Barkham et al., 2021). The explanatory hypotheses are in line with theory in the wider psychotherapy literature, which suggests that "common factors" of therapy can influence positive change in the context of low levels of symptoms, but that greater symptom severity can diminish the effectiveness of the foundational relational elements of psychotherapy (Lambert et al., 2004; Lorenzo-Luaces et al., 2014). Future research using the FIS methodology, investigating distinct subgroups of clients, is needed to replicate and substantiate this finding in a larger sample to help uncover client populations where utilising high levels of FIS capabilities optimises psychological treatment.

Overall, the current study provides support for the application and utility of the FIS methodology in rating genuine therapy sessions. The current study, in collaboration with the findings of Faulds (2021), has demonstrated the utility of selecting NOT sessions to maximise the

potential of isolating interpersonal ruptures or events in the therapy dyad. The study has also

evidenced that high inter-rater agreement in FIS coding can be achieved in rating longer, and more

complex sections of genuine therapy sessions. Finally, the current study provides some evidence in

support of the association between in-session ratings of FIS and client outcome 12-months post-

randomisation to treatment.

Limitations and Directions for Future Research

The in-session rating of FIS has associated limitations. Firstly, therapists' FIS within therapy

sessions constitute reflexive and relational acts and are somewhat jointly constructed with the client.

Therefore, this methodology, although an insight into therapists' behaviours within a setting with

ecological validity, does not allow for the isolation the therapists' contributions to the

psychotherapy processes and outcomes. Furthermore, the client contributions were not standardised

or explicitly measured in the current project. Secondly, the project by Faulds (2021) demonstrates

therapists' variability in the use of FIS skills, therefore rating only one session per dyad may

overlook interpersonal transactions in other sessions influencing client outcome. Thirdly, the use of

audio-recordings to rate in-session FIS may have limited the use of non-verbal interpersonal

behaviours of the therapist that contribute to the facilitative interpersonal environment, and limits

comparability with the original performance task using video-recordings of therapists' responses.

A further limitation is the relatively small sample size, which was below the minimum

number of cases required from the power calculation. The final study sample size was limited by

the number of cases meeting sample inclusion/exclusion criteria and the extent of missing outcome

data from the PRaCTICED trial. Furthermore, rating in-session FIS is time and labour intensive,

however, training additional raters would have potentially introduced additional error variance. As

the research constituted a proof-of-concept study, it was decided that prioritising the quality of the

FIS rating and sample selection, with a smaller sample, would provide the best conditions in which

to assess the reliability and promise of the in-session FIS methodology. Further research, using

large datasets, would be useful to extend the current findings. A recent study reported on the

development of an automated system for FIS rating using machine learning, which could increase the scalability of the FIS performance task (Goldberg et al., 2021) and potentially extend to in-session rating in future iterations.

Other limitations of the current study would also be addressed by future research utilising larger datasets, allowing for more robust statistical analysis methods to be employed. Due to the structure of the current data, with the client session FIS ratings nested within therapists, a multilevel modelling approach would have been beneficial. However, the size and structure of the current sample was not sufficient for reliable modelling (Maas & Hox, 2005). As multilevel modelling analysis was not viable, the statistical analyses used in the current study must be interpreted with caution, as the observations within therapists were not independent, which may have inflated the Type 1 error rate and estimates of effect sizes. Additionally, given the significant differences in FIS domain items between PCET and CBT (Faulds, 2021), the current study may have benefitted from assessment of the FIS-outcome association at an individual skill level. However, a larger sample size would be required to run the multiple analyses with sufficient statistical power to reveal an underlying effect.

Clinical Implications

The current study adds to existing evidence demonstrating that clients are helped differentially by components of psychotherapy, and highlights how the research differentiating distinct client subgroups can help elucidate what works best for whom. Understanding how FIS can be used consciously by therapists to create optimal therapy environments to promote clinical change directly influences clinical practice and training. Although requiring replication, the current findings may suggest that considered modulation of interpersonal skills may be needed, dependent on a complex interplay of client characteristics, such as baseline depressions severity, in order to provide the foundation for optimally effective psychotherapy treatments. The benefit of the in-session FIS methodology is its potential utility as a supervision or training tool to increase clinician's self-awareness of their use of interpersonal skills in sessions with clients to develop responsive and

reflexive use based on client interpersonal style. Furthermore, FIS skills have been found to be amenable to development through modelling, deliberate practice and training (T. Anderson, Perlman, et al., 2020; Jones, 2019; Perlman et al., 2020) and may pose a beneficial addition to clinical training programmes to cultivate interpersonally-effective therapists.

Conclusion

The current research establishes rating in-session FIS as a promising methodology for measuring therapists' interpersonal qualities in "problematic" therapy sessions. In the full sample, therapist FIS during NOT sessions was not a significant predictor of outcome at 12 months, and the association was not significantly moderated by therapy orientation. A differential impact of FIS was observed in depression severity subgroups, providing preliminary indication that clients with greater baseline clinical severity of depression symptoms may respond less favourably to therapists with high levels of FIS during "problematic" therapy sessions. Future research with larger samples comparing FIS in distinct clinical subgroups is needed to substantiate the study findings.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders.* American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Anderson, T., Crowley, M. E. J., Himawan, L., Holmberg, J. K., & Uhlin, B. D. (2016). Therapist facilitative interpersonal skills and training status: A randomized clinical trial on alliance and outcome. *Psychotherapy Research, 26*(5), 511–529. https://doi.org/10.1080/10503307.2015.1049671

Anderson, T., Finkelstein, J., & Horvath, S. A. (2020). The facilitative interpersonal skills method: Difficult psychotherapy moments and appropriate therapist responsiveness. *Counselling and Psychotherapy Research, 20*(3), 463–469. https://doi.org/10.1002/capr.12302

Anderson, T., McClintock, A. S., Himawan, L., Song, X., & Patterson, C. L. (2016). A prospective study of therapist facilitative interpersonal skills as a predictor of treatment outcome. *Journal of Consulting and Clinical Psychology, 84*(1), 57–66. https://doi.org/10.1037/ccp0000060

Anderson, T., Ogles, B. M., Patterson, C. L., Lambert, M. J., & Vermeersch, D. A. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist success. *Journal of Clinical Psychology, 65*(7), 755–768. https://doi.org/10.1002/jclp.20583

Anderson, T., & Patterson, C. (2013). *Facilitative interpersonal skill task and rating method*. http://clinica.ispa.pt/ficheiros/areas_utilizador/user11/facilitative_interpersonal_skill_manual_and_rating_scale.pdf

Anderson, T., Patterson, C., McClintock, A. S., McCarrick, S. M., Song, X., & The Psychotherapy and Interpersonal Skills Lab Team. (2019). *Facilitative interpersonal skills task and rating manual* [Unpublished]. Ohio University. https://www.fisresearch.com/rating.html

Anderson, T., Patterson, C., & Weiss, A. C. (2006). *Facilitative interpersonal skills performance analysis rating method* [Unpublished]. Ohio University, Ohio.

Anderson, T., & Perlman, M. R. (2019). Therapeutic interpersonal skills for facilitating working alliance. In J. N. Fuertes (Ed.), *Working alliance skills for mental health professionals* (pp. 43-68). Oxford University Press. https://doi.org/10.1093/med-psych/9780190868529.001.0001

Anderson, T., Perlman, M. R., McCarrick, S. M., & McClintock, A. S. (2020). Modeling therapist responses with structured practice enhances facilitative interpersonal skills. *Journal of Clinical Psychology, 76*(4), 659–675. https://doi.org/10.1002/jclp.22911

Barkham, M., Lutz, W., Lambert, M. J., & Saxon, D. (2017). Therapist effects, effective therapists, and the law of variability. In L. G. Castonguay & C. E. Hill (Eds.), *How and why are some therapists better than others?: Understanding therapist effects*. (pp. 13–36). American Psychological Association. https://doi.org/10.1037/0000034-002

Barkham, M., Saxon, D., Hardy, G. E., Bradburn, M., Galloway, D., Wickramasekera, N., Keetharuth, A. D., Bower, P., King, M., Elliott, R., Gabriel, L., Kellett, S., Shaw, S., Wilkinson, T., Connell, J., Harrison, P., Ardern, K., Bishop-Edwards, L., Ashley, K., … Brazier, J. E. (2021). Person-centred experiential therapy versus cognitive behavioural therapy delivered in the English Improving Access to Psychological Therapies service for the treatment of moderate or severe depression (PRaCTICED): A pragmatic, randomised, non-inferiority trial. *The Lancet Psychiatry*. https://doi.org/10.1016/S2215-0366(21)00083-3

Beck, A. T., Rush, J. A., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. The Guilford Press.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck depression inventory-II*. Psychological Corporation.

Beck, J. S. (2011). *Cognitive therapy: Basics and beyond* (2nd ed.). Guildford Press.

Bentham, C. (2021). *Therapeutic alliance and premature termination: An updated meta-analysis of adult individual psychotherapy* [Unpublished doctoral dissertation]. University of Sheffield.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice, 16*(3), 252–260. https://doi.org/10.1037/h0085885

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Constantino, M. J., Visla, A., Coyne, A. E., & Boswell, J. F. (2018). A meta-analysis of the association between patients' early treatment outcome expectation and their posttreatment outcomes. *Psychotherapy, 55*(4), 473–485. https://doi.org/10.1037/pst0000169

Delgadillo, J., de Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., Ali, S., Aguirre, E., Appleton, M., Nevin, J., O'Hayon, H., Patel, U., Sainty, A., Spencer, P., & McMillan, D. (2018). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: A multisite, open-label, cluster randomised controlled trial. *The Lancet Psychiatry, 5*(7), 564–572. https://doi.org/10.1016/S2215-0366(18)30162-7

Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. *Psychotherapy, 48*(1), 43–49. https://doi.org/10.1037/a0022187

Elliott, R., Bohart, A. C., Watson, J. C., & Murphy, D. (2018). Therapist empathy and client outcome: An updated meta-analysis. *Psychotherapy, 55*(4), 399–410. https://doi.org/10.1037/pst0000175

Eubanks, C. F., Muran, J. C., & Safran, J. D. (2018). Alliance rupture repair: A meta-analysis. *Psychotherapy, 55*(4), 508–519. https://doi.org/10.1037/pst0000185

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE–OM. *British Journal of Psychiatry, 180*(1), 51–60. https://doi.org/10.1192/bjp.180.1.51

Farber, B. A., & Doolin, E. M. (2011). Positive regard. *Psychotherapy, 48*(1), 58–64. https://doi.org/10.1037/a0022141

Farber, B. A., Suzuki, J. Y., & Lynch, D. A. (2018). Positive regard and psychotherapy outcome: A meta-analytic review. *Psychotherapy, 55*(4), 411–423. https://doi.org/10.1037/pst0000171

Faulds, A. (2021). *An examination of the relationship between client and therapist variables and therapists' facilitative interpersonal skills (FIS) during "not on track" therapy sessions* [Unpublished doctoral dissertation]. University of Sheffield.

Flückiger, C., Del, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy, 55*(4), 316–340. https://doi.org/10.1037/pst0000172

Gilbody, S., Richards, D., Brealey, S., & Hewitt, C. (2007). Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): A diagnostic meta-analysis. *Journal of General Internal Medicine, 22*(11), 1596–1602. https://doi.org/10.1007/s11606-007-0333-y

Goldberg, S. B., Tanana, M., Imel, Z. E., Atkins, D. C., Hill, C. E., & Anderson, T. (2021). Can a computer detect interpersonal skills? Using machine learning to scale up the facilitative interpersonal skills task. *Psychotherapy Research, 31*(3), 281–288. https://doi.org/10.1080/10503307.2020.1741047

Greenberg, L., & Paivio, S. (2003). *Working with emotions in psychotherapy*. Guildford Press.

Groenwold, R. H. H., Donders, A. R. T., Roes, K. C. B., Harrell, F. E., & Moons, K. G. M. (2012). Dealing with missing outcome data in randomized trials and observational studies. *American Journal of Epidemiology, 175*(3), 210–217. https://doi.org/10.1093/aje/kwr302

Harrell, F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. Springer.

Hayes, A. F. (2020). *The PROCESS macro for SPSS* (Version 3.5.2) [Computer software]. https://processmacro.org/index.html

Hill, A. (2011). *Curriculum for counselling for depression: Continuing professional development for qualified therapists delivering high intensity interventions.* https://webarchive.nationalarchives.gov.uk/20160302160209/http://www.iapt.nhs.uk/silo/files/curriculum-for-counselling-for-depression.pdf

Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (pp. 25–70). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199737208.003.0002

IBM Corp. (2019). *Statistical Package for the Social Sciences* (Version 27.0) [Computer software]. IBM Corp.

Janzen, J. (2007). *Attachment and facilitative interpersonal skill in critical therapy incidents* [Doctoral dissertation]. McGill University. https://escholarship.mcgill.ca/downloads/hx11xj96d?locale=en

Jones, A. (2019). *The role of feedback in therapist training of facilitative interpersonal skills* [Doctoral dissertation]. Texas Tech University. https://ttu-ir.tdl.org/handle/2346/85947

Kim, D. M., Wampold, B. E., & Bolt, D. M. (2006). Therapist effects in psychotherapy: A random-effects modeling of the National Institute of Mental Health Treatment of Depression

Collaborative Research Program data. *Psychotherapy Research, 16*(2), 161–172.

https://doi.org/10.1080/10503300500264911

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation

coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163.

https://doi.org/10.1016/j.jcm.2016.02.012

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression

severity measure. *Journal of General Internal Medicine, 16*(9), 606–613.

https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lambert, M. J. (2017). Maximizing psychotherapy outcome beyond evidence-based medicine.

Psychotherapy and Psychosomatics, 86(2), 80-89. https://doi.org/10.1159/000455170

Lambert, M. J., Garfield, S. L., & Bergin, A. E. (2004). Overview, trends, and future issues. In M. J.

Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behaviour Change*

(pp. 805–821). Wiley.

Lewis, G., Pelosi, A. J., Araya, R., & Dunn, G. (1992). Measuring psychiatric disorder in the

community: A standardized assessment for use by lay interviewers. *Psychological Medicine,*

22(2), 465–486. https://doi.org/10.1017/S0033291700030415

Lorenzo-Luaces, L., Derubeis, R. J., & Webb, C. A. (2014). Client characteristics as moderators of

the relation between the therapeutic alliance and outcome in cognitive therapy for depression.

Journal of Consulting and Clinical Psychology, 82(2), 368–373.

https://doi.org/10.1037/a0035994

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology,*

1(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

Martell, C. R., Addis, M. E., & Jacobson, N. S. (2001). *Depression in context: Strategies for guided*

action. W. W. Norton & Co.

McClintock, A. S., Song, X., Jameson, G., Anderson, T., Patterson, C., Goldman, G. A., Uhlin, B., MacFarlance, P., & Goldman, E. D. (2012, June 22). *Reliability of therapist facilitative interpersonal skills for brief responses to "thin slice" client video simulations*. Annual Meeting of the International Society for Psychotherapy Research, Virginia Beach, Virginia, USA.

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

Murphy, D. (2019). *Person-centred experiential counselling for depression: A manual for training and practice.* Sage Publications Ltd.

Newhill, C., Safran, J., & Muran, J. (2003). *Negotiating the therapeutic alliance: A relational treatment guide.* Guilford Press.

Norcross, J. C., & Lambert, M. J. (2018). Psychotherapy relationships that work III. *Psychotherapy, 55*(4), 303–315. https://doi.org/10.1037/pst0000193

Pearce, P., Sewell, R., Hill, A., & Coles, H. (2012). Counselling for depression. *Therapy Today*, 20–23. https://www.researchgate.net/profile/Peter_Pearce2/publication/256936041_Counselling_for_depression_TTFeb121/links/0deec52415769e509e000000.pdf

Peluso, P. R., & Freund, R. R. (2018). Therapist and client emotional expression and psychotherapy outcomes: A meta-analysis. *Psychotherapy, 55*(4), 461–472. https://doi.org/10.1037/pst0000165

Perlman, M. R., Anderson, T., Foley, V. K., Mimnaugh, S., & Safran, J. D. (2020). The impact of alliance-focused and facilitative interpersonal relationship training on therapist skills: An

RCT of brief training. *Psychotherapy Research, 30*(7), 871–884.

https://doi.org/10.1080/10503307.2020.1722862

Probst, T., Kleinstäuber, M., Lambert, M. J., Tritt, K., Pieh, C., Loew, T. H., Dahlbender, R. W., & Delgadillo, J. (2020). Why are some cases not on track? An item analysis of the assessment for signal cases during inpatient psychotherapy. *Clinical Psychology & Psychotherapy, 27*(4), 559-566. https://doi.org/10.1002/cpp.2441

Safran, J. D., Muran, J. C., & Eubanks-Carter, C. (2011). Repairing alliance ruptures. *Psychotherapy, 48*(1), 80–87. https://doi.org/10.1037/a0022140

Sanders, P., & Hill, A. (2014). *Counselling for depression: A person-centred and experiential approach to practice*. Sage.

Saxon, D., Ashley, K., Bishop-Edwards, L., Connell, J., Harrison, P., Ohlsen, S., Hardy, G. E., Kellett, S., Mukuria, C., Mank, T., Bower, P., Bradburn, M., Brazier, J., Elliott, R., Gabriel, L., King, M., Pilling, S., Shaw, S., Waller, G., & Barkham, M. (2017). A pragmatic randomised controlled trial assessing the non-inferiority of counselling for depression versus cognitive-behaviour therapy for patients in primary care meeting a diagnosis of moderate or severe depression (PRaCTICED): Study protocol for a randomised controlled trial. *Trials, 18*(1), 93. https://doi.org/10.1186/s13063-017-1834-6

Sharf, J., Primavera, L. H., & Diener, M. J. (2010). Dropout and therapeutic alliance: A meta-analysis of adult individual psychotherapy. *Psychotherapy, 47*(4), 637–645. https://doi.org/10.1037/a0021175

Spitzer, R. L., Kroenke, K., & Williams, J. B. W. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *Journal of the American Medical Association, 282*(18), 1737–1744. https://doi.org/10.1001/jama.282.18.1737

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

Strupp, H. H., & Binder, J. L. (1984). *Psychotherapy in a new key: A guide to time limited psychotherapy.* Basic Books.

The EuroQol Group. (1990). EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy, 16*(3), 199–208. https://doi.org/10.1016/0168-8510(90)90421-9

Truax, C. B., Fine, H., Moravec, J., & Millis, W. (1968). Effects of therapist persuasive potency in individual psychotherapy. *Journal of Clinical Psychology, 24*(3), 359–362. https://doi.org/10.1002/1097-4679(196807)24:3<359::AID-JCLP2270240325>3.0.CO;2-O

Uhlin, B. (2011). *Therapist in-session rated facilitative interpersonal skills (FIS-IS) in the psychotherapy process* [Doctoral dissertation]. Ohio University. https://etd.ohiolink.edu/pg_10?::NO:10:P10_ETD_SUBID:61851

Wang, Y. P., & Gorenstein, C. (2013). Psychometric properties of the Beck depression inventory-II: A comprehensive review. *Revista Brasileira de Psiquiatria, 35*(4), 416–431. https://doi.org/10.1590/1516-4446-2012-1048

Appendices

Appendix A: Shared and Distinct Aspects of the Project

Researchers CB and AF completed a collaborative project focusing on therapists' use of Facilitative

Interpersonal Skills (FIS) during "not on track" client sessions.

The shared components of the project included:

1. The identification of the "not on track" subsample from the PRaCTICED dataset.

2. The completion of training on the FIS rating method and sample of practice ratings until "good" levels of

   interrater agreement was reached.

3. The completion of FIS ratings on the PRaCTICED data subsample.

The individual projects have distinct aims, hypotheses, and proposed analyses:

Aims of project by CB:

1. To assess if therapists' facilitative interpersonal skills during "not on track" client sessions are predictive

   of client outcome (using the BDI-II and CORE-OM).

2. To consider interactions between treatment modalities and therapist skills by assessing if facilitative

   interpersonal skills are more important for client outcome in PCET or CBT.

Aims of project by AF:

1. To determine the relationship between therapy/therapist variables (therapy model, therapist adherence

   and competence, therapist demographics: age, gender, years of experience) and FIS during NOT therapy

   sessions.

2. To determine the relationship between client variables (client symptom severity, client resilience, client

   expectancy/credibility, client demographics: age, gender, employment status, multiple deprivation score)

   and FIS during NOT therapy sessions.

3. To consider the stability of therapists' interpersonal communication style in NOT therapy sessions, by

   examining patterns of FIS variability within a session and across clients.

Appendix B: Ethics Application and Approval Letter

The
University
Of
Sheffield.

Application 032724 Self-declaration

Section A: Applicant details

Date application started:
Fri 31 January 2020 at 09:53

First name:
Charlotte

Last name:
Bentham

Email:
cbentham1@sheffield.ac.uk

Programme name:
Doctorate in Clinical Psychology

Module name:
Research Thesis
Last updated:
31/01/2020

Department:
Psychology

Applying as:
Postgraduate research

Research project title:
The role of therapists' interpersonal skills during "not on track" client sessions in predicting client response to counselling personcentred experiential and cognitive behavioural therapy interventions for depression.

Similar applications:
Concurrently running project: An examination of the relationship between client and therapist variables and Facilitative Interpersonal Skills (FIS) during "not on track" therapy sessions.

Section B: Basic information

Supervisor

Name	Email
Gillian Hardy	g.hardy@sheffield.ac.uk

Proposed project duration

Proposed start date:
Sat 1 February 2020

Proposed end date:
Fri 28 May 2021

Section C: Summary of research

1. Aims & Objectives

Therapists have been found to play a significant role in clients' responses to psychological therapies. Facilitative interpersonal skills (FIS) are defined as an individual's ability to perceive, decode, and express a variety of interpersonal messages. Extant research, which assesses therapist FIS using a performance task, suggests that clients treated by therapists with higher levels of FIS have better treatment outcomes. However, there is limited evidence replicating this finding when a therapist's interpersonal skills are rated from actual therapy sessions. Furthermore, there is no extant literature to consider if a therapists' ability to use interpersonal skills are more central to particular therapeutic frameworks, such as cognitive behavioural therapy (CBT) or person-centred experiential therapy (PCET).

The current study aims:
1) To assess if therapists' facilitative interpersonal skills during "not on track" client sessions are predictive of client outcome. In line with the findings of previous research, it is hypothesised that clients engaging in therapy with therapists with higher utilisation of FIS will have greater clinical improvement over the duration of treatment.
2) To consider interactions between treatment modalities and therapist skills by assessing if facilitative interpersonal skills are more important for client outcome in PCET or CBT. It is hypothesised that higher scores on the FIS will be more predictive of client symptomatic improvement in PCET than in CBT.

2. Methodology

The proposed project will comprise a retrospective analysis of clinical data collected as part of the PRaCTICED trial; a non-inferiority randomised controlled trial comparing the outcomes of CBT and PCET.

A sample will be selected from the trial dataset meeting the inclusion and exclusion criteria. One of the criterion stipulates that a client session must be deemed "not on track" when compared to average treatment response curves, using the algorithm proposed by Delgadillo et al. (2018).

Session recordings will be extracted for the sample. The audio-recordings of identified sessions will then be used to rate therapists' interpersonal skills using the FIS measure. The ratings will be completed by trained raters (CB and AF), who will conduct frequent checks of inter-rater agreement.

The Beck Depression Inventory (BDI-II) and Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM) collected prior to treatment, and at 6- and 12-months follow-up time points will be extracted from the trial dataset, and used as measures of treatment effectiveness.

To examine the predictive ability of therapists' FIS in explaining variability in client outcome and the distinctions between therapies, the current study proposes to initially use a single-level regression model. The dependent variables in the analysis will be BDI-II and CORE-OM scores at 6- and 12-month follow-up time points. The regression model will include the following independent variables: 1) baseline BDI-II or CORE-OM score to assess treatment effect, and 2) therapist FIS score. To examine the second study hypothesis, an interaction term will be added to assess the relationship between FIS scores and treatment type. Additional exploratory descriptive statistics may also be used to examine the impact, and differences in utilisation between therapies, of specific interpersonal skills. Significant predictors that emerge from this analysis, and additional prognostic features, may then be added to an exploratory multi-level model, including the therapist as a random intercept, as a sensitivity analysis to check the robustness of the primary results.

The initial PRaCTICED trial involved participants experiencing mental health problems, and also involved highly sensitive topics, including information about participants' mental health problems. These risks were managed in the initial trial; using a robust consent process and active monitoring of risk and distress of the participants throughout the intervention stage. As a retrospective analysis of data collected from the trial, the proposed research does not present any additional risk to the participants involved.

All data made available to the researchers will be anonymised, with no inclusion of patient identifiable information. It is not possible, however, to anonymise the recordings of therapy sessions, as client and therapist voices may be identifiable. All data arising for the recordings will be anonymised. In the unlikely event that the researchers (CB & AF) recognise the therapist or client from a session recording, they will stop listening and ask the other researcher to rate the tape in order to protect confidentiality. All recordings will be stored on a University of Sheffield encrypted device. Only a limited number of recordings will be held on the device at any one time. Once the ratings have been completed the recordings will be deleted from the devices and retained centrally with the PRaCTICED trial team.

Section F: Supporting documentation

Additional Documentation

Document 1074987 (Version 1)   All versions
Research Protocol

External Documentation

- not entered -

Section G: Declaration

Signed by:
Charlotte Bentham
Date signed:
Fri 31 January 2020 at 10:05

Offical notes

- not entered -

Charlotte Bentham
Registration number: 180156933
Psychology
Programme: Doctorate in Clinical Psychology

Dear Charlotte

PROJECT TITLE: The role of therapists interpersonal skills during not on track client sessions in predicting client response to counselling personcentred experiential and cognitive behavioural therapy interventions for depression.
APPLICATION: Reference Number 032724

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

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

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

Yours sincerely

Thomas Webb
Departmental Ethics Administrator

125

Appendix C: The PHQ-9 Measure

Developed by Spitzer, Kroenke & Williams (1999)

Copyrighted information redacted for upload to the White Rose Repository.

Appendix D: The BDI-II Measure

Developed by Beck, Steer & Brown (1996)

Appendix E: CORE-OM Measure

Copyrighted information redacted for upload to the White Rose Repository.

Appendix F: The FIS Manual

Anderson and Patterson (2013)

Appendix G: Minutes of Calibration Meetings for FIS Ratings

Confidential information redacted for upload to the White Rose Repository.

Appendix H: Problem Responsiveness Item Rating System

Copyrighted information redacted for upload to the White Rose Repository.

Appendix I: Rater Differences in FIS Scores

Table 1.

Average FIS scores and independent-samples t-tests between FIS raters

FIS Item	Rater 1 (Mean)	Rater 2 (Mean)	t	p
Verbal Fluency	11.5	11.7	-0.23	0.82
Hope	10.5	10.9	-0.73	0.47
Persuasiveness	11.0	11.0	-0.001	0.99
Emotional Expression	11.7	12.3	-0.83	0.41
Warmth	11.0	12.1	-0.14	0.89
Empathy	10.9	11.4	-0.71	0.48
Alliance Bond Capacity	11.7	11.3	0.47	0.64
Problem Responsiveness	10.1	10.6	-0.73	0.47
FIS Total	89.4	91.3	-0.42	0.68

Appendix J: Assessment of Assumptions for Regression Analyses

There was linearity of relationships between variables, as assessed by visual inspection of partial regression plots and a plot of studentised residuals against the predicted values. Visual inspection of plots of studentised residuals versus unstandardised predicted values also indicated there was homoscedasticity. There was no evidence of multicollinearity, as indicated by tolerance values greater than 0.1. The assumption of normality of residuals was also met, as assessed by P-P plots of the regression standardised residuals.

Regression Predicting CORE-OM at 12 months

Figure 1

Scatterplot of Studentised Residual and Unstandardised Predicted Variable of CORE-OM Score at 12 Months
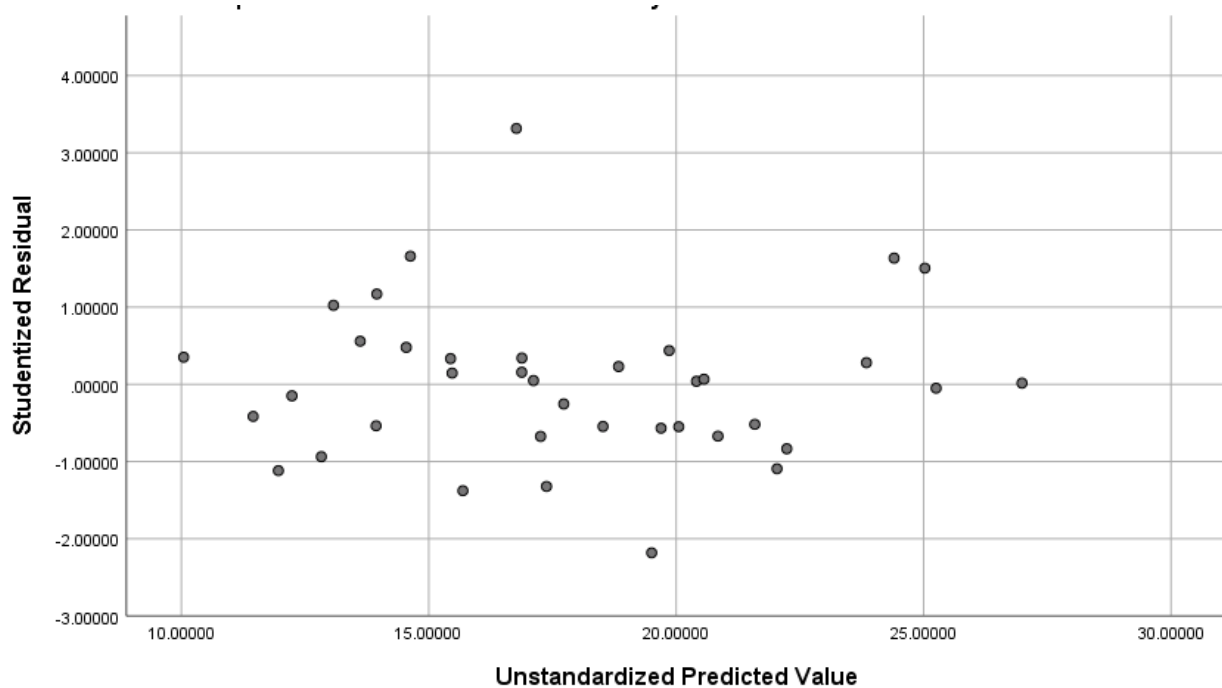
Figure 2

Partial Regression Plot of Relationship between CORE-OM at Baseline at 12 Months
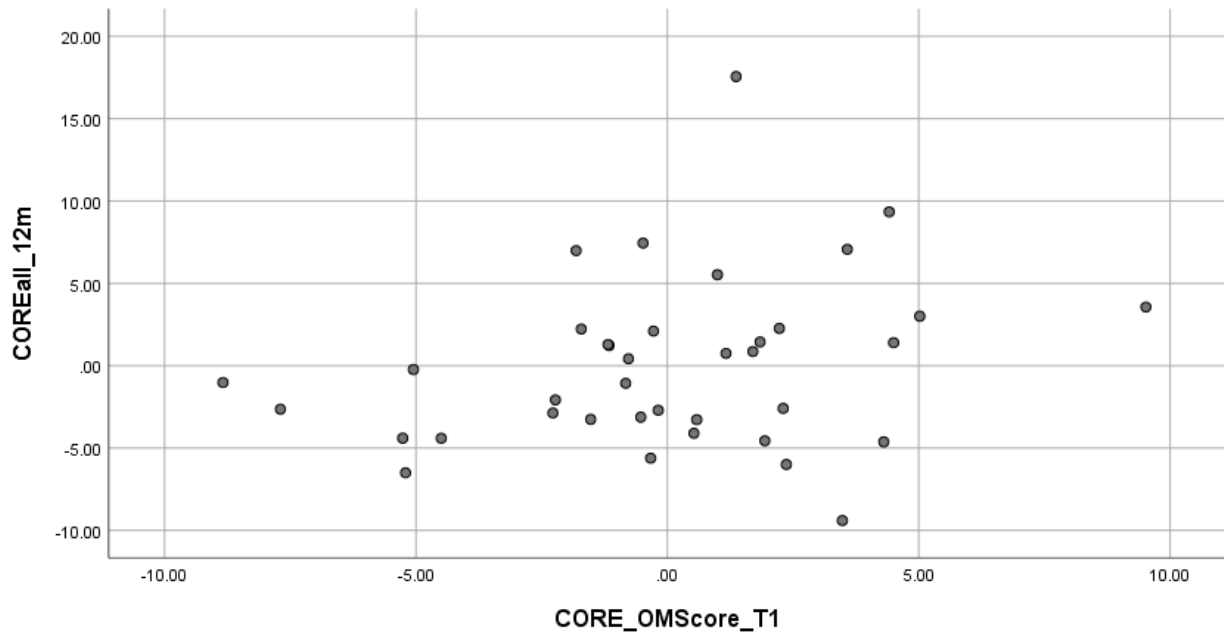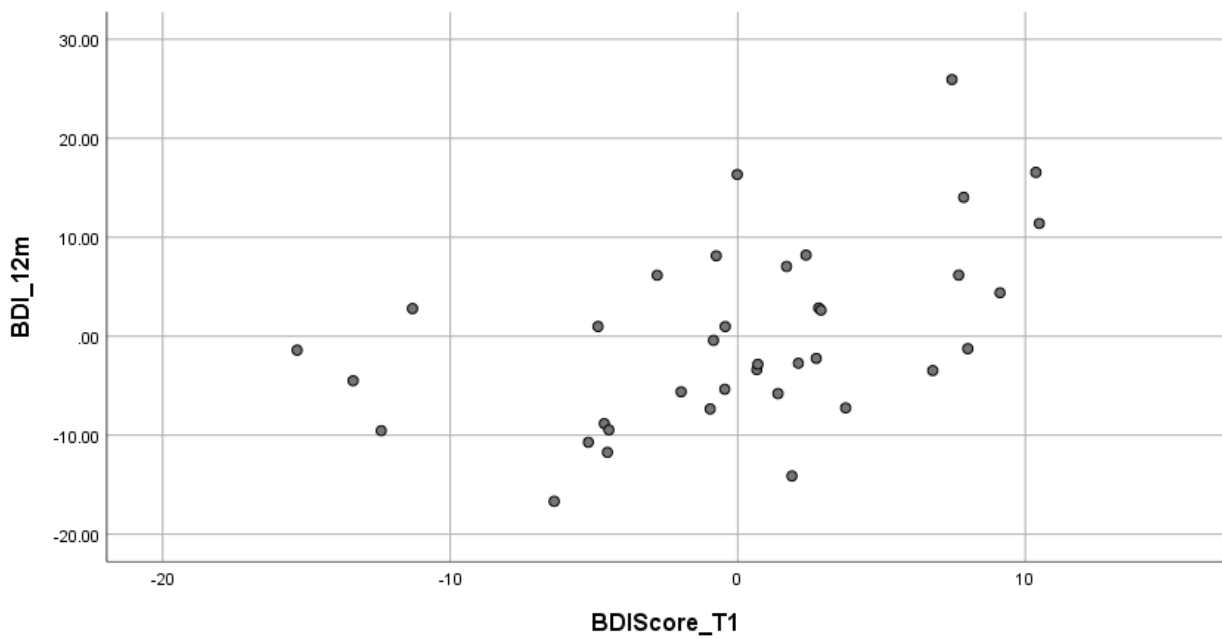

Figure 3

Partial Regression Plot of Relationship Between Client Age and CORE-OM at 12 Months
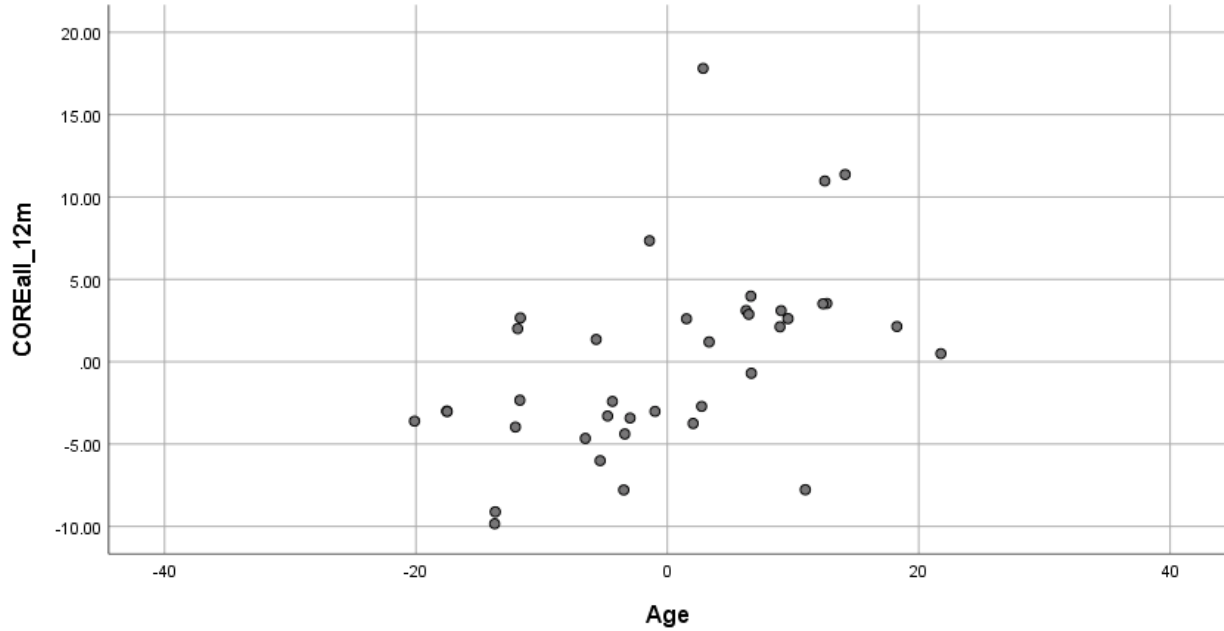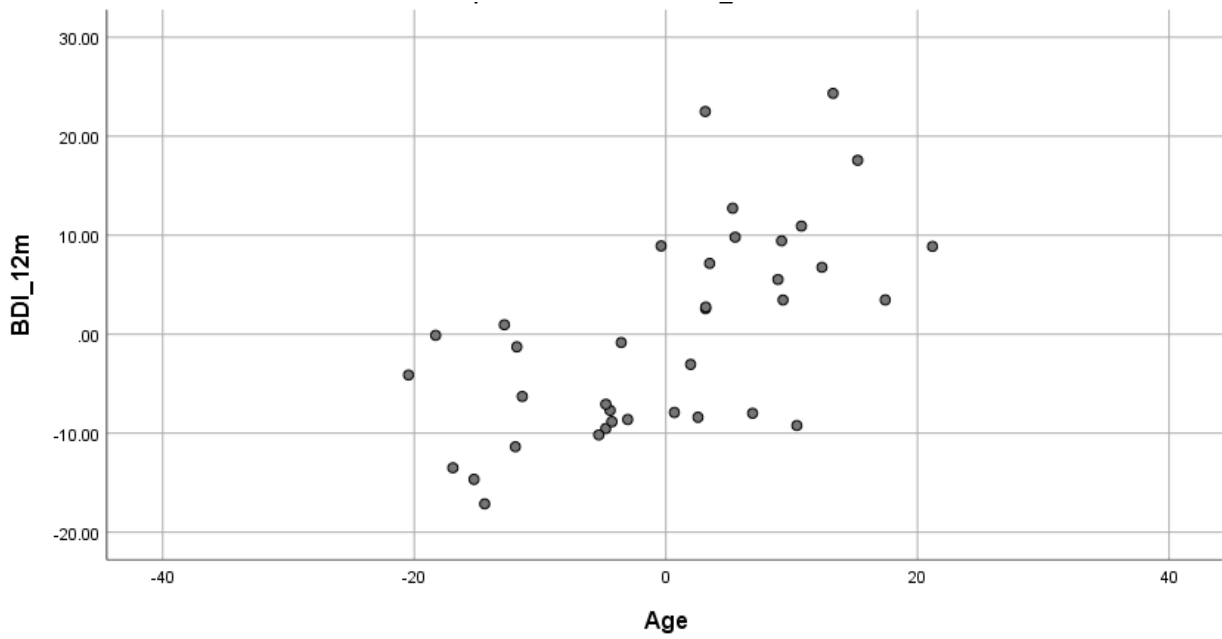
Figure 4

*Partial Regression Plot of Relationship Between FIS Total and CORE-OM at 12 Months*
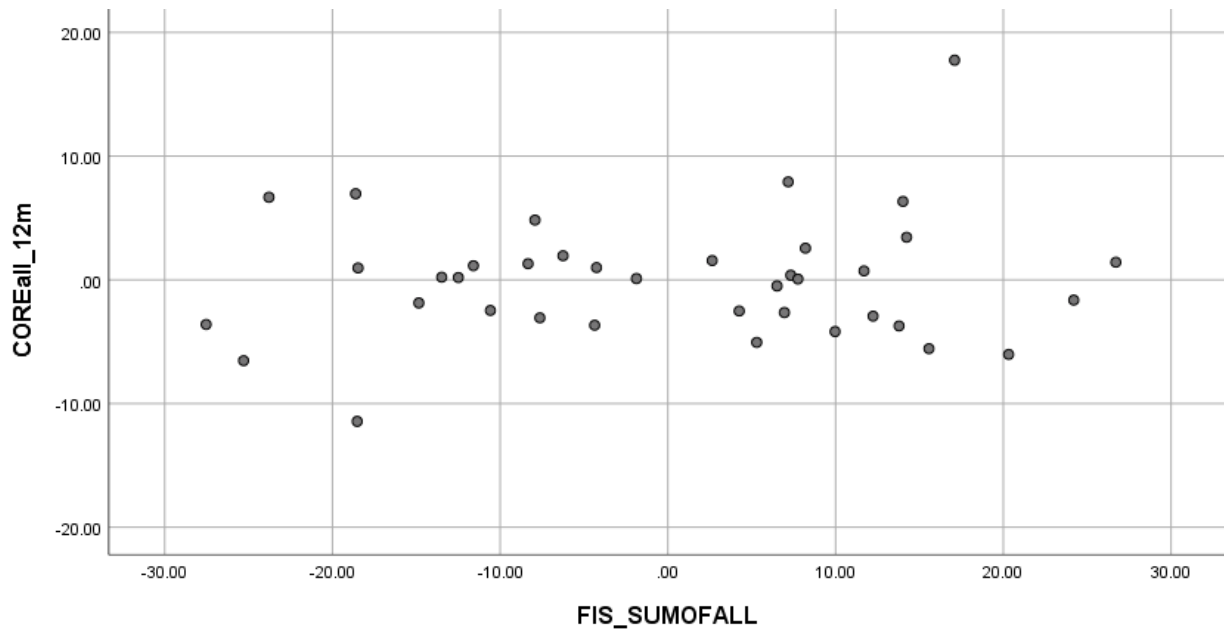
Figure 5

Normal P-P Plot of Regression Standardised Residuals for Analysis Predicting CORE-OM at 12 Months

Figure 6

Scatterplots of Studentised Residual and Unstandardised Predicted Variable of BDI-II Score At 12 Months
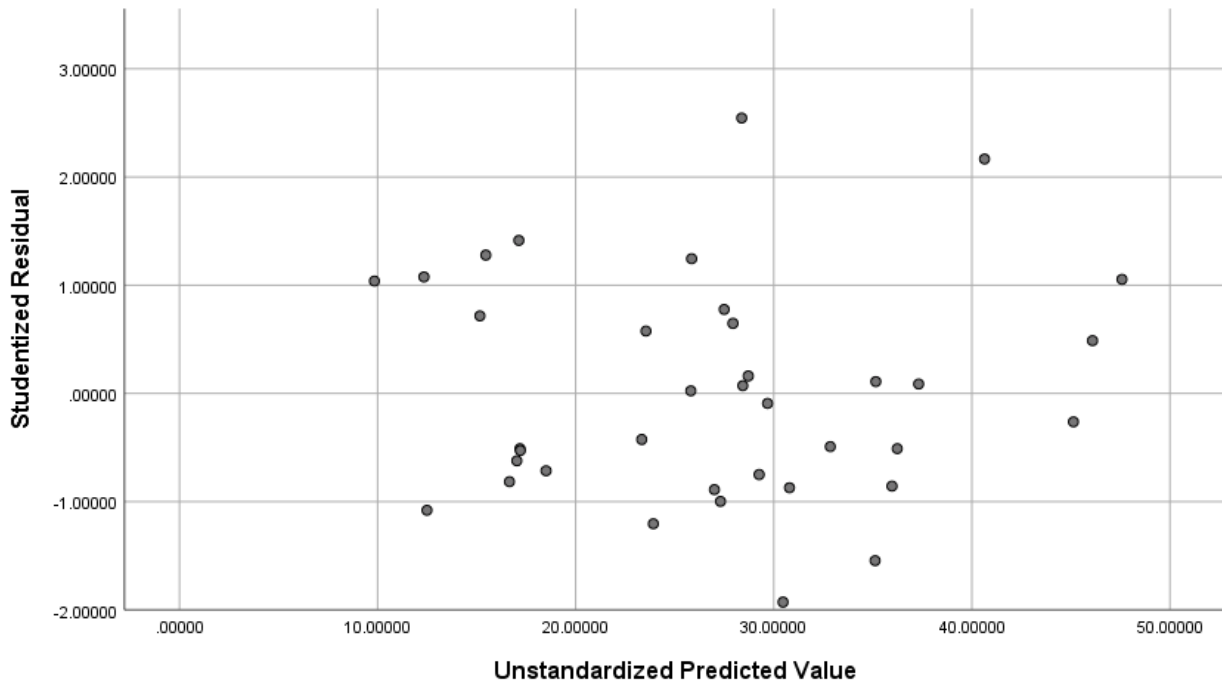
Figure 7

Partial Regression Plot of Relationship Between BDI-II at Baseline at 12-Months

Figure 8

Partial Regression Plot of Relationship Between Client Age and BDI-II at 12-Months

Figure 9

Partial Regression Plot of Relationship Between FIS Total and BDI-II at 12-Months
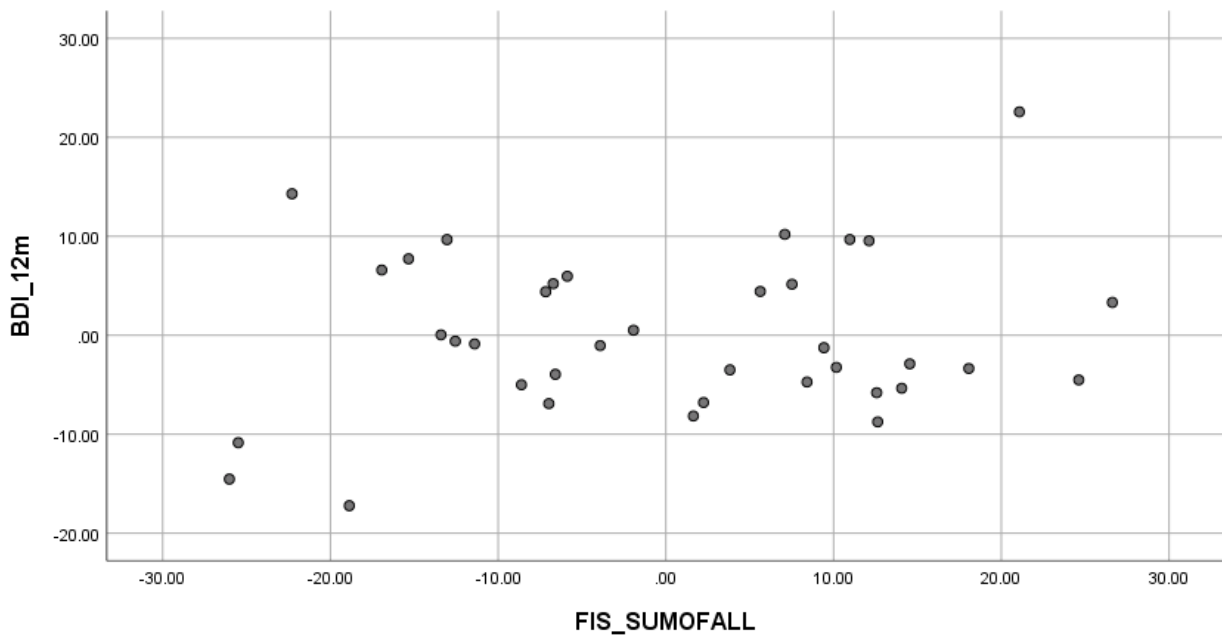
Figure 10

Normal P-P Plot of Regression Standardised Residuals for Analysis Predicting BDI-II at 12 Months
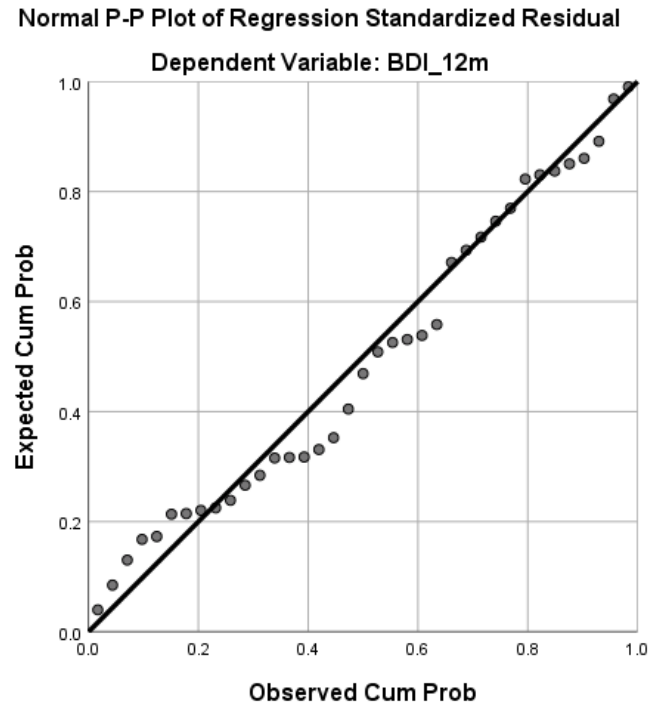
Appendix K: Assessment of Assumptions for ANCOVA Analyses

Visual inspection of scatterplots indicated approximately linear relationships between continuous covariates and dependent variables, at each level of FIS category and depression severity. There was homogeneity of regression slopes between the continuous covariates and the dependent variables at each combination of the predictor groups, as determined by insignificant interaction terms in two-way ANCOVA analyses. There was homoscedasticity within each combination of groups of the two independent variables, as assessed by visual inspection of the studentised residuals plotted against the predicted values for each group (Appendix K). There was homogeneity of variances, as assessed by Levene's test of homogeneity of variance, for both the analyses predicting change in CORE-OM and BDI-II ($p = .743$ and $p = .198$, respectively). Studentised residuals were largely normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), with the exception of the severe depression group with low FIS therapists for the analyses predicting change in CORE-OM ($p = .010$). Given the robustness of ANCOVA to violations of normality, it was decided to proceed with the analysis without transformation.

ANCOVA Predicting Change in CORE-OM

Table 1

Scatterplots of Relationships Between Client Age and Change in CORE-OM Score at Each Level of FIS Category and Depression Severity (With Loess Lines)
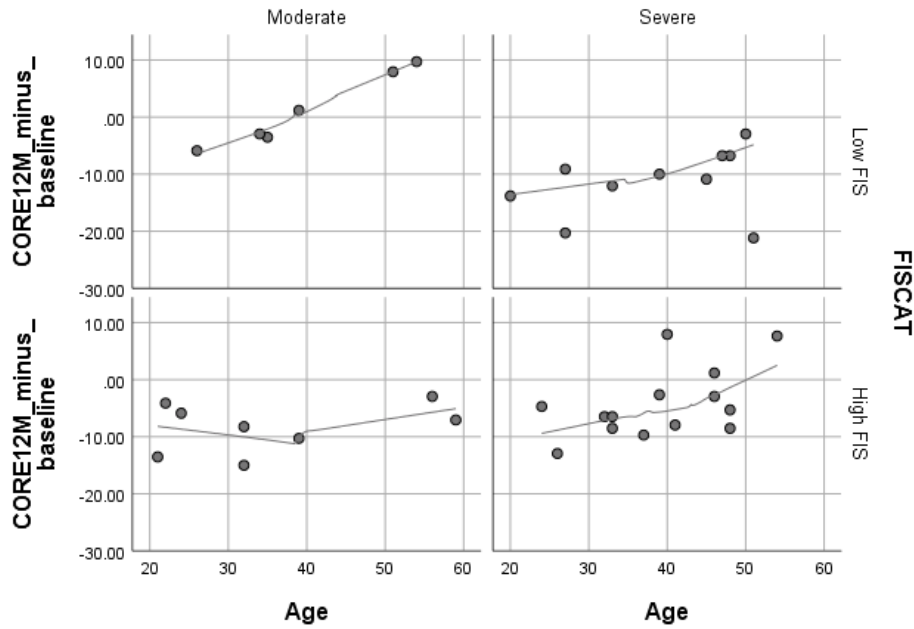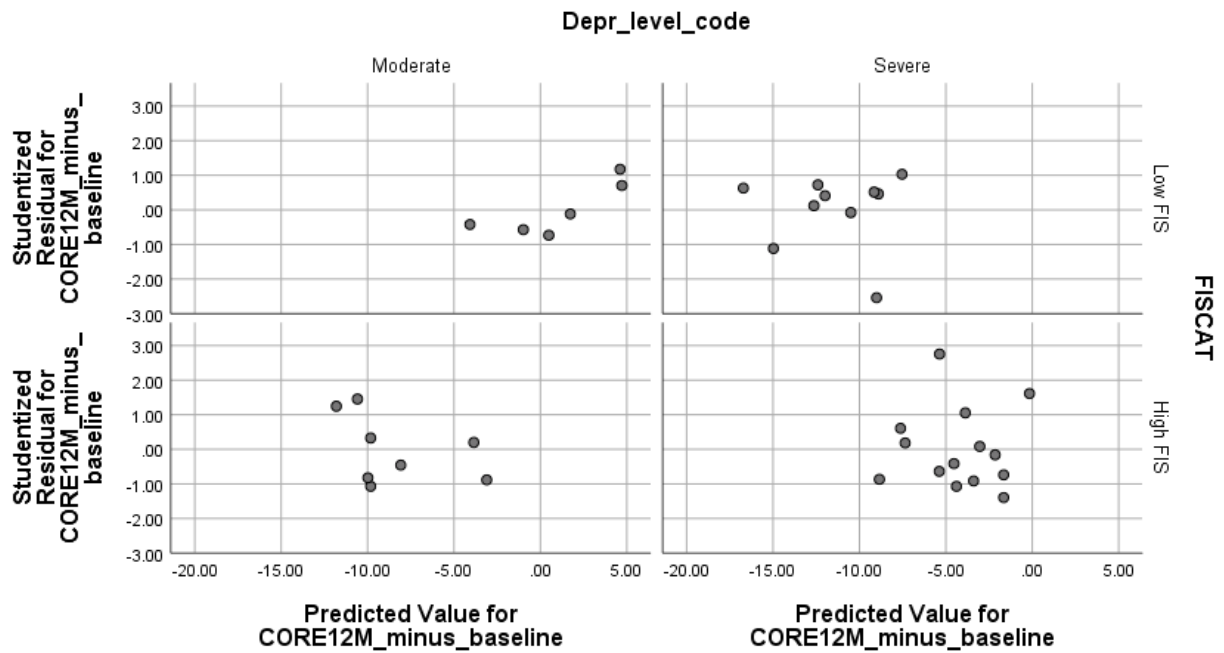

Table 2

Scatterplots of Studentised Residual and Predicted Variable of Change in CORE-OM Score at Each Level of FIS Category and Depression Severity

Table 3

Scatterplots of Relationships Between Client Age and Change in BDI-II Score at Each Level of FIS Category and Depression Severity (With Loess Lines)
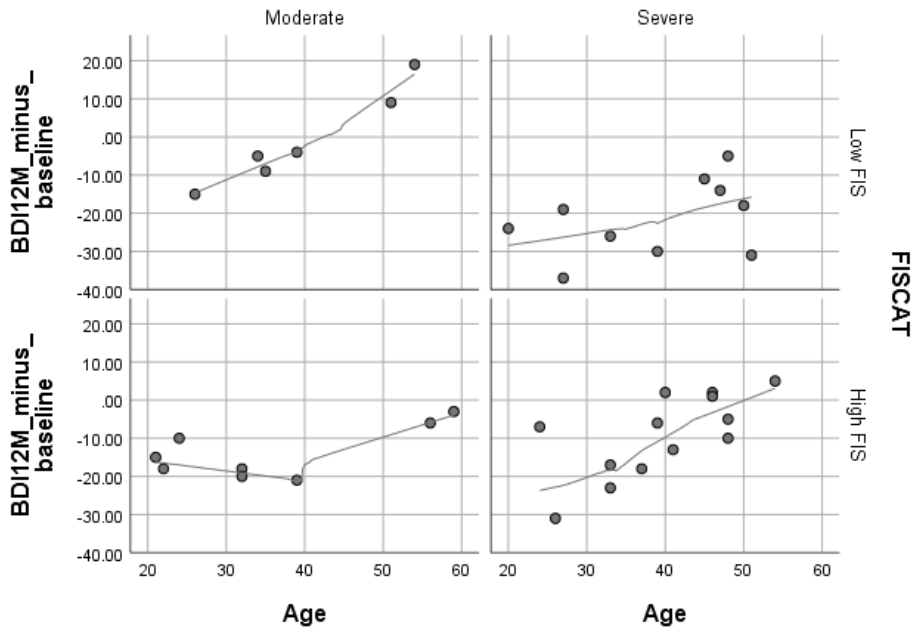

Table 4

Scatterplots of Studentised Residual and Predicted Variable of Change in BDI-II Score at Each Level of FIS Category and Depression Severity