Citation: Delgadillo, J., Branson, A., Kellett, S., Myles-Hooton, P., Hardy, G., & Shafran, R. (*in press*). Therapist personality traits as predictors of psychological treatment outcomes. *Psychotherapy Research*.

Therapist personality traits as predictors of psychological treatment outcomes

Jaime Delgadillo[1], Amanda Branson[2], Stephen Kellett[1],

Pamela Myles-Hooton[2], Gillian E. Hardy[1], & Roz Shafran[3]

1. Clinical Psychology Unit, Department of Psychology, University of Sheffield, Sheffield, UK
2. School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK
3. UCL Great Ormond Street Institute of Child Health, London, UK

Conflicts of interest: None.

Correspondence: jaime.delgadillo@nhs.net

Abstract

Aim: To investigate if therapists' personality influences their patients' treatment outcomes.

Methods: N = 4,052 patients were treated by 69 therapists, including 36 Psychological Wellbeing Practitioners (PWPs) and 33 Cognitive Behavioural Therapists (CBTs). Therapists completed the NEO-PI-R personality inventory, they reported years of clinical experience, and expert assessors rated their clinical competence and reflective abilities. Their patients completed pre and post-treatment measures of depression (PHQ-9) and anxiety (GAD-7). Associations between therapist personality traits and patient treatment outcomes were examined using multilevel modelling, controlling for therapist demographics, clinical experience, technical competence and reflective ability.

Results: Relative to other sources of variability, therapists accounted for 1% to 3% of overall variability in treatment outcomes. However, the magnitude of systematic heterogeneity in performance *between therapists* was around 6%, such that the best-performing therapists outperformed average therapists by a margin of moderate to large effects (g = .57 to 1.10). Clinical experience, technical competence and reflective ability were unrelated to treatment outcomes. Patients treated by PWPs with above-average *agreeableness* scores and CBTs with above-average *openness to experience* scores had poorer treatment outcomes.

Conclusions: Therapist effects may be partly explained by the influence of their personality on their work with anxious and depressed patients.

Keywords: Therapist effects; Cognitive Behavioural Therapy; Depression; Anxiety

Introduction

Differences in treatment outcomes between psychotherapists, referred to as *therapist effects*, have been consistently documented across both randomised controlled trials and practice-based studies (Castonguay & Hill, 2017; Norcross & Lambert, 2019; Wampold & Imel, 2015). Two systematic reviews of this literature indicate that approximately 5% of variability in psychological treatment outcomes is explained by therapist effects (Baldwin & Imel, 2013; Johns, Barkham, Kellett, & Saxon, 2018). Even when the absolute magnitude of the therapist effect is seemingly small, it can have a significant clinical impact. For example, Okiishi, Lambert, Nielsen and Ogles (2003) found that the recovery rates of the most effective therapists (22.45%) were twice as high as those attained by the least effective therapists (10.65%). Similarly, Saxon, Barkham, Foster and Parry (2017) noted that the dropout rates of patients treated by the least effective therapists were four times higher (49%) than patients treated by the most effective therapists (12%). There is some evidence that these performance differences between therapists already become apparent during training, even when therapists are exposed to the same training methods and therapy models (Banham & Schweitzer, 2016).

The reasons why some therapists attain better treatment outcomes are less well understood. There are relatively few studies examining associations between therapists' characteristics and patients' outcomes using sufficiently large samples and applying appropriate statistical methods (e.g., multilevel modelling). Literature reviews by Beutler et al. (2004), and more recently Castonguay and Hill (2017), indicate that "static factors" such as therapist sex, age and race are not associated with treatment outcomes. Wampold, Baldwin, Gross Holtforth and Imel (2002) proposed that there are three core "dynamic" factors that define effective therapists. The first factor is the ability to form

a positive alliance across a range of clients, which is inferred from literature showing that the alliance is consistently associated with treatment outcomes (e.g., see Flückiger, Del Re, Wampold, & Horvath, 2018). Secondly, highly effective therapists have highly developed facilitative interpersonal skills which enable them to effectively work with complex and/or challenging cases (e.g., see Anderson, Ogles, Patterson, Lambert & Vermeersch, 2009). Thirdly, effective therapists are willing to improve their therapeutic skills via deliberate practice, which refers to the identification of specific therapeutic competency deficits followed by targeted practice and feedback (Chow et al., 2015; Rousmaniere, 2016). Consistent with the above, a recent systematic review of the professional and personal characteristics of effective therapists concluded that intrapersonal variables, interpersonal skills and experiencing difficulties with practice, coping mechanisms and attitudes towards therapeutic work were important (Heinonen & Nissen-Lie, 2019). Nevertheless, what is currently known about the characteristics of effective therapists is mostly inferred from studies of process factors (e.g., alliance) and isolated findings from therapist effects studies that have not been replicated.

It is possible that some of the factors that characterise effective therapists could be rooted in trait-like personality features. The most well-established theory of personality (Goldberg, 1990) focuses on the *big five traits*: extraversion, agreeableness, openness to experience, conscientiousness and neuroticism. These personality traits have been associated with indices of job motivation (conscientiousness, openness to experience and extraversion; Hart, Stasson, Mahoney & Story, 2007), job performance (consciousness and emotional stability; Lado & Alono, 2017) and job satisfaction (extraversion, openness, agreeableness and conscientiousness; Judge, Heller, & Mount, 2002) across a range of occupations and in hundreds of studies examined in systematic reviews and meta-analyses (He, Donnellan & Mendoza, 2019).

Much less research is available on therapists' personality and how this might influence their work and clinical outcomes. Peters-Scheffer, Didden, Korzilius and Sturmey (2013) found that therapists that showed higher fidelity to a treatment protocol had a lower level of openness to experience. Casari, Ison, Margarita & Gomez (2019) found that the personal style of therapists was associated with their personality traits; for example, therapists' emotional closeness with their patients was positively correlated with extraversion but negatively correlated with conscientiousness. Chapman et al. (2009) found positive associations between patient-rated therapeutic alliance and trainee therapists' neuroticism, but negative associations with therapists' openness to experience. Furthermore, Rieck and Callahan (2013) found that trainee therapists' neuroticism scores were associated with their patients' post-treatment symptomatic improvement, and that this relationship was moderated by therapists' emotional intelligence. Apart from the latter study, most other studies in the area of psychotherapy do not directly examine relationships between personality traits and treatment outcomes. Furthermore, in order to parse any potential effects of personality traits, therapists' experience and skill-level (i.e., technical competency and reflective ability) should be controlled for. Although some rare studies have examined therapists' personality in the context of carefully controlled features like experience and skill-level (e.g., Antonuccio, Lewinsohn, & Steinmetz, 1982) none have been adequately powered or designed to examine therapist effects.

In summary, the interpersonal skills and attitudes to learning that distinguish effective therapists could be related to more trait-like personality features. Therefore, the present study aimed to contribute to this literature by assessing if therapist effects may be influenced by therapists' personality, after controlling for clinical experience, technical competence and reflective ability. Given the scarcity of prior studies and

theory in this area, we took an exploratory approach to examine the relationship between the big five personality traits and clinical outcomes without specifying any a priori hypotheses. Applying multilevel modelling and meta-analytic methods in a large practice-based dataset, we were able to examine between-therapist and within-therapist variability in treatment outcomes.

Methods

Design, setting and participants

The study dataset contained de-identified information from psychological therapists and all of the patients they treated during a 1-year clinical training period and up to 2-years post-qualification. This study pooled datasets which were previously examined in the studies by Branson et al. (2015, 2018). Ethical approval was granted by the Berkshire Research Ethics Committee (Ref: 09/H0505/128).

Participating therapists were trained to work in services linked to the Improving Access to Psychological Therapies (IAPT) programme in England, which delivers evidence-based treatments for depression and anxiety organised in a stepped care model (Clark, 2018). In this stepped care system, the initial step of treatment is delivered by Psychological Wellbeing Practitioners (PWPs) and involves brief (<8 sessions) guided self-help interventions. Guided self-help is based on principles of cognitive behavioural therapy and has a well-established evidence base for the treatment of mild-to-moderate depression and anxiety problems (e.g., see Cuijpers, Donker, van Straten, & Andersson, 2010; Coull & Morris, 2011). Guided self-help delivered in IAPT services is highly structured and follows five core treatment protocols outlined in clinical guidelines for PWPs (Richards & Whyte, 2009). Patients who do not respond to guided self-help are

stepped-up to therapy delivered by qualified Cognitive Behavioural Therapists (CBTs). CBT in IAPT services involves formal psychological therapy that is supported by evidence from efficacy trials (e.g., see Cuijpers et al., 2013; Hofmann & Smits, 2008) and is highly structured, time-limited (up to 20 sessions) and guided by disorder-specific treatment manuals that are outlined in a clinical competency framework (Roth & Pilling, 2008).

PWPs undertook a year-long, national curriculum-based training programme (National IAPT Team, 2015), combining University-based modular training (1-day per week) with clinical practice (4-days per week) under weekly supervision. Over the duration of training, PWPs received 45 days of teaching, and completed 25 hours of clinical supervision in their service. CBTs also undertook a year-long course based on a national curriculum (Department of Health, 2011), combining University-based training (2-days per week) with clinical practice (3-days per week) under weekly supervision. Over the duration of training, CBTs received at least 300 hours of teaching, 35 hours of clinical supervision in the service, and 35 hours of group-based university supervision. Supervision was delivered by experienced PWPs and CBTs accredited by the British Association for Behavioural and Cognitive Psychotherapies (BABCP).

Measures

Competence and reflective ability

Observed Structured Clinical Examinations (OSCEs), where trainees are observed conducting a therapy session with an actor-patient, were used to assess PWP competence. OSCEs were rated using standardized scales detailed in the PWP national training curriculum (Richards & Whyte, 2009), including domains such as interpersonal skills, information gathering skills, information giving skills and shared

decision-making in the application of self-help strategies, etc. Trainees undertook three OSCEs as training progressed, which were rated by expert clinical trainers using standardized scales, covering competency in assessment (OSCE1, approximately 3 months into training), guided self-help interventions (OSCE2, ~6 months) and case management supervision (OSCE3, ~9 months). OSCEs are rated 0–100, with a pass mark of 50. There were no failed OSCEs at any of the 3 assessments in this group of PWPs. A mean competency score was calculated by averaging all three OSCEs to control for maturation effects (Sackett & Mullen, 1993), and to ensure that the average scores were more representative of the therapist's practice across patients rather than within a single case. Interrater reliability for the OSCE scoring in this sample has been previously reported (Branson, Myles, Mahdi, & Shafran, 2018).

The Cognitive Therapy Scale Revised (CTS-R; Blackburn et al., 2001) was used to assess CBT clinical competence (Reichelt, James, & Blackburn, 2003). The CTS-R is a 12-item scale designed to measure therapist competence; items are measured on a 7-point Likert scale ranging from incompetent (0) to expert (6). The CTS-R covers various domains such as agenda setting, collaboration, interpersonal effectiveness, guided discovery, application of change methods, homework setting, etc. Scores on the CTS-R range from 0-72; the threshold for treatment competence is 36 (i.e., any score falling below 36 is classified as a fail). Trainees submitted audio-recordings of three therapy sessions with different clients over the duration of training (approximately 3, 6 and 9 months into training), which were then rated by expert clinical trainers using the CTS-R manual. Interrater reliability for the CTS-R scoring in this sample has been previously reported (Branson, Shafran, & Myles, 2015). A mean competency index was calculated by averaging all three CTS-R ratings, to ensure average ratings were broadly

representative of each therapist's practice across patients and time. The CTS-R fail rates in the group of CBTs were: Patient 1 = 11.1%, Patient 2 = 2.9%; Patient 3 = 0%.

For both PWPs and CBTs, their understanding of CBT theory and how it relates to clinical practice was examined through the completion of three reflective analyses (1000 words each) of their practice. These assignments were scored on a 0 to 100-point scale by expert trainers and were averaged to obtain a proxy measure of *reflective ability*, which specifically indexes the therapist's understanding of theory and practice links.

Personality and demographics

The big five personality domains (neuroticism, extraversion, openness to experience, conscientiousness, agreeableness) and their lower order facets (32) were measured using the 240-item NEO PI-R (Costa & McCrae, 2006). The measure has good internal consistency, with alphas ranging from .87 for agreeableness to .92 for neuroticism (Costa & McCrae, 2006). The NEO PI-R was completed by PWPs and CBTs during their training programme. Alphas in this sample were: neuroticism = .84, extraversion = .74, openness to experience = .72, conscientiousness = .80, and agreeableness = .76. Information on gender, age, ethnic background and years of clinical experience working in therapeutic roles were also collected.

Patient treatment outcome measures

The PHQ-9 is a 9-item screening tool for depression symptoms, where each item is rated between 0 and 3, yielding a total severity score between 0 and 27 (Kroenke et al., 2001). A score above ≥10 has been proposed as providing the best trade-off between sensitivity (88%) and specificity (88%) for a diagnosis of major depression (Kroenke et al., 2001). The GAD-7 is a 7-item questionnaire used to identify anxiety disorders; each item is also rated between 0 and 3, with a total severity score between 0 and 21

(Spitzer et al., 2006). A cut-off score ≥8 is recommended to identify clinically important anxiety symptoms (Kroenke et al., 2007), with adequate sensitivity (77%) and specificity (82%). Both measures were collected on a session-by-session basis, where the last observed measure was carried forward as the post-treatment score for all cases including completers and those who dropped out. Pre- and post-treatment measures were available for patients who attended more than 1 session of therapy. A *stage of training* variable grouped patients into three consecutive time periods: during the 1-year training period, the initial 6 months following training, >6 months and up to 24-months post-qualification.

Sample selection

The study sample included data for 4,052 patients treated by 69 therapists (PWP = 36; CBT = 33). These records were selected from a wider dataset (N = 4,371; k = 90); only including therapists who treated at least 5 patients that attended more than 1 session. This was in order to meet minimal sample size recommendations to investigate therapist effects in routine care settings (Schiefele et al., 2017). No cases in the dataset overlapped across therapist samples (e.g., no patients were treated by PWPs and later CBTs in this sample). Therapists were excluded if they did not complete the NEO PI-R (n = 20) and if they failed the training programme (n = 1). Sample characteristics for the PWP and CBT groups are summarised in Table 1. CBTs tended to be older and more experienced, while PWPs attained higher average reflective ability scores. There were no significant personality differences between CBTs and PWPs. A transformation of NEO PI-R trait scores into T-scores (following the recommendations of Costa & McCrae, 1992) indicated that sample means for all 5 domains were normally distributed and comparable to general population norms.

Consistent with the stepped care system, patients treated by CBTs were more clinically impaired (mean PHQ-9 = 14.12 vs. 11.72), and had a longer average duration of treatment (mean number of treatment sessions = 9.15 vs. 4.47).

Statistical analyses

The effectiveness of therapy was examined using pre-post treatment effect sizes (Cohen's *d*) using the method described by Minami et al. (2008). Reliable improvement and deterioration rates were calculated based on the reliable change indices for PHQ-9 (≥6) and GAD-7 (≥5), as recommended by Richards and Borglin (2011).

The primary analysis applied multilevel modelling (MLM), with patients (level 1) nested within therapists (level 2). Post-treatment symptoms were entered as the dependent variable, with separate models for therapist groups (PWPs, CBTs) and outcome measures (PHQ-9, GAD-7). Consistent with conventional model-building guidelines, MLM was performed in sequential steps, starting with unconditional models and eventually developing conditional (case-mix adjusted) models that had adequate goodness-of-fit (Raudenbush, 1993). Model comparisons were guided by three goodness-of-fit indices (-2 log likelihood ratio, AIC, BIC), and the loglikelihood ratio test. The intra-cluster correlation coefficient (ICC) was calculated to estimate the amount of variance in outcomes attributable to the therapist-level. Model-building followed three steps. Step 1 was an unconditional model with no predictors other than the random intercept for therapists. Step 2 was a conditional model that added patient-level baseline severity measures (PHQ-9 and GAD-7). This step indicated that a linear function for baseline severity measures had better goodness-of-fit (smaller indices) compared to fitting nonlinear trends (quadratic, cubic). Step 3 was a fully-adjusted model additionally including therapist-level variables: demographics (age, gender), years of experience, mean clinical competence score (OSCE for PWP group, CTS-R

for CBT group), personality traits, reflective ability score, and stage of training. The latter variable enabled the examination of within-therapist variability across time (training phase vs. post-qualification phase).

Once statistically significant personality traits (NEO-PI-R domains) were identified in the preceding analysis, we examined the predictive value of specific personality facets (items) using a supervised machine learning algorithm (LASSO regularization with optimal scaling). Only those facets corresponding to the NEO PI-R personality traits that predicted patients' post-treatment outcomes (PHQ-9, GAD-7) in each of the therapist groups were examined in these secondary analyses. Multicollinearity was expected between NEO PI-R items from the same trait, so regularization (Tibshirani, 1996) was performed to exclude variables that did not improve predictive value and which covaried strongly with other items. LASSO shrinks (penalizes) beta coefficients toward zero, yielding conservative models that minimise overfitting. Variables with coefficients that were shrunk to zero were deemed to have no significant predictive value. In order to determine the model with minimal *expected prediction error*, a 10-fold cross-validation approach was applied (Rodriguez, Perez, & Lozano, 2010). This machine learning algorithm selected and rescaled variables to optimally fit nonlinear relationships between specific NEO PI-R items and post-treatment outcomes. The selected and rescaled items were then entered into MLM predicting post-treatment outcomes for the corresponding group (PWPs, CBTs) and measure (PHQ-9, GAD-7). This facet-based MLM (examining selected items) was compared to a trait-based MLM (examining domain scores) to compare the goodness-of-fit (-2 log likelihood ratio) and the statistical significance of regression coefficients.

The variability in treatment outcomes between therapists was further examined using meta-analytic procedures. A single-level OLS regression (adjusting for baseline

severity) was used to estimate *expected* post-treatment symptom scores for each case, averaging across all therapists in order to model typical (*average*) outcomes. Expected and observed post-treatment scores and pooled standard deviations were then used to calculate a *performance effect size* (expected = control; observed = intervention), using standard procedures to model between-group differences in meta-analysis (Higgins & Green, 2011). The performance effect represents a therapist's relative deviation from the mean, where positive values indicate 'better than expected' performance. These performance effects were modelled using Hedges' *g*, given the differential sample sizes across caseloads. A random effects meta-analysis was run to visually examine performance effects using a forest plot. This random effects model follows the same assumptions as MLM: the included studies (or therapists) are assumed to be a random sample of the relevant distribution of effects in the wider population, and treatment effects are assumed to vary across samples (Borenstein, Hedges, & Rothstein, 2007). However, meta-analysis has some advantages over conventional MLM methods (i.e., caterpillar plots that rank therapists using a residual scale), as it enables a more informative comparison of performance using an effect size scale, and it enables the calculation of systematic differences in outcomes *between therapists* using the I^2 statistic. Following established statistical methods and concepts in healthcare quality control studies (Bradley et al., 2009; Schmidtke et al., 2017; Spiegelhalter, 2005), we used conventional funnel plots to identify *positive deviants* (exceptionally high performing therapists compared to the mean) and precision funnel plots to identify *relative outliers* (therapists who have markedly different outcomes compared to one another), adjusting for sample sizes.

Secondary analyses investigated correlations between therapists' competency, reflective ability and personality traits.

Results

Overall pre-post treatment outcomes

Large pre-post treatment effect sizes were observed in the full sample (PHQ-9 d = 0.87, GAD-7 d = 0.94) and for each group; PWP d = 0.84 to 0.88; CBT d = 0.99 to 1.13. Reliable improvement rates in the full sample were between 45.9% (PHQ-9) and 50.4% (GAD-7); PWP = 43.0% to 48.1%; CBT = 46.1% to 56.9%. Reliable deterioration rates were low in the full sample (2.6 to 3.6) and across groups; PWP = 2.7 to 3.7; CBT = 2.3 to 3.4.

Examining differences in treatment outcomes between therapists

According to MLM analyses, approximately 1% to 3% of variability in treatment outcomes was attributable to the therapist-level (PWP cohort ICC = .015 to .011; CBT cohort ICC = .028 to .033). The forest plot in Figure 1 presents performance effect sizes for all therapists, ranked from least to most effective. Systematic differences (heterogeneity) in performance between therapists were in the range of 6% (I^2 = 5.75%), with some therapists showing moderate (g = .57) to large effects (g = 1.10) above expected outcomes. The group variable was not significant when entered into random effects meta-analysis as a categorical moderator (p = .87), indicating no significant differences in performance between PWPs and CBTs after adjusting for case-mix and sample sizes. The conventional funnel plot in Figure 2A identified four highly performing therapists (*positive deviants*). The precision funnel plot in Figure 2B shows that even when some therapists may not be "better than average", their outcomes are markedly different to one another (*relative outliers*).

Associations between therapist characteristics and treatment outcomes

Fully adjusted linear MLM results are presented in Table 2. Baseline depression and anxiety were significantly associated with post-treatment outcomes in all models, although baseline GAD-7 did not predict depression outcomes in the CBT group. Clinical competence measures, reflective ability and years of experience were not associated with treatment outcomes. Patients treated more than 6-months post-qualification (stage 3) had significantly better outcomes compared to those treated during the training period (stage 1). For PWPs, an above-average level of *agreeableness* was significantly associated with poorer treatment outcomes. For CBTs, an above-average level of *openness to experience* was significantly associated with poorer treatment outcomes.

Figure 1 labels all therapists according to their relative level of agreeableness (PWPs) and openness (CBTs) traits, revealing a systematic trend whereby therapists with above-average (1 standard deviation above sample mean) traits tended to have negative performance effect size estimates. Those with average and below-average traits were randomly distributed across the spectrum of performance effects. This pattern was confirmed statistically in a subgroup meta-analysis which excluded all above-average-trait therapists, resulting in a reduction of the index of heterogeneity in performance ($I^2 = 0.00\%$).

Secondary analysis of personality traits and facets

Following the identification of statistically significant personality traits, machine learning analyses were applied with the 6 facets corresponding to the *agreeableness* trait for PWPs, and the 6 facets corresponding to the *openness* trait for CBTs. For PWPs, only a single facet (*compliance*) was selected in both the PHQ-9 (B = .004, SE = .01) and GAD-7 model (B = .002, SE = .01). For CBTs, the facet *openness to ideas* was selected in the PHQ-9 model (B = .003, SE = .02). Two facets were selected in the

GAD-7 model; *openness to feelings* (B = .003, SE = .02) and *openness to ideas* (B = .005, SE = .02). As shown in Table 3, all facet-based models had better goodness-of-fit when compared to trait-based models. The *openness* trait was not statistically significant in any of the models, and only its facet *openness to ideas* was significant in the PHQ-9 model.

Secondary correlational analyses revealed associations between therapist competency, reflective ability, and specific personality traits. In PWPs, competence (mean OSCE) was moderately correlated with reflective ability (r = .50, p < .001) and weakly correlated with neuroticism (r = .20, p < .001) and extraversion (r = .19, p < .001). In CBTs, competence (mean CTS-R) was strongly correlated with reflective ability (r = .78, p < .001) and weakly correlated with all of the big five personality traits (r = .07 to .41, p < .001) except for conscientiousness (r = - .00, p = .92).

Discussion

This study investigated psychological therapists' personality traits as potential predictors of depression and anxiety treatment outcomes, after controlling for clinical competence, experience and reflective ability. Measures of clinical competence in the delivery of evidence-based psychological interventions were not associated with treatment outcomes. This is consistent with prior research that indicates either weak or non-significant associations between measures of treatment fidelity/adherence and clinical outcomes (Webb, DeRubeis, & Barber, 2010). Therapists' ability to conceptualise theory-practice links (*reflective ability*) was also not statistically significant. We note that reflective ability, as measured in this study, is theoretically different to the notion of *reflective functioning* which indexes therapists' *mentalisation* skills and which has previously been associated with treatment outcomes (Cologon et

al., 2017). Consistent with prior studies reviewed by Beutler et al. (2004), years of experience and age of therapists were unrelated to treatment outcomes. Overall, demographics, experience, technical and theoretical competences did not differentiate between more and less effective therapists.

After controlling for experience, technical and theoretical competences, therapists' personality traits were found to be associated with treatment effects. These findings mirror those of Chapman et al. (2009) who found that the personality factors of *neuroticism*, *agreeableness* and *openness* were associated with ratings of the therapeutic alliance. Although our study did not find a relationship between *neuroticism* and treatment outcomes, we found remarkably similar relationships between *agreeableness* and *openness* with treatment outcomes. Chapman et al. (2009) also carried out exploratory analyses to identify specific facets (items), which revealed that therapists with higher levels of non-antagonistic orientation (*compliance*) had lower alliance ratings. Furthermore, they observed a (non-significant) trend suggesting that therapists with high levels of intellectual interest (*openness to ideas*) tended to have lower alliance ratings. Like the sample in the Chapman et al. (2009) study, the present sample of therapists also had higher-than-average neuroticism, openness and agreeableness scores relative to UK norms (see further details reported by Branson & Shafran, 2015). This indicates that extremely high rather than average or below-average indices of agreeableness and openness were specifically associated with treatment outcomes.

Agreeableness refers to a prosocial, cooperative, considerate, likeable, trustful, and empathic personality. The natural disposition of agreeable people towards empathy and cooperation may at first appear advantageous for therapeutic work, given the well-known associations between empathy (Elliott, Bohart, Watson, & Murphy, 2018) and

the alliance (Flückiger et al., 2018) with positive treatment outcomes. Yet, patients treated by highly agreeable PWPs tended to have poorer treatment outcomes. Our findings indicate that the specific facet of *compliance* may exert an unfavourable influence on the short-term, highly structured and psychoeducational nature of interventions delivered by PWPs. Prior experimental research has shown that highly agreeable people tend to be more accommodating in situations of conflict and emotional arousal, and particularly when under time pressure (Perunovic, & Holmes, 2008). It may be that highly *compliant* PWPs may be overly passive or accommodating, thus colluding with (e.g. failing to challenge and modify) and reinforcing some patients' maladaptive beliefs or behaviours. Although levels of *agreeableness* were similar for PWPs and CBTs, this did not predict patient outcomes in the latter group. This is perhaps because time pressures are less prominent for CBT therapists who tend to see patients for a higher number of sessions, and therefore have more opportunities to overcome barriers to improvement such as patients' avoidance and use of safety seeking behaviours.

Patients with poorer post-treatment outcomes tended to be treated by CBTs with higher *openness to experience*. This personality trait is associated with being creative, curious, imaginative, cultured and perceptive. Previous studies indicate that therapists with high levels of openness tend to be less directive and less adherent to treatment protocols (Casari et al., 2019; Peters-Scheffer et al., 2013). Some studies investigating *openness* have suggested that, at extreme levels, it is associated with nonconformity and adherence to eccentric or unconventional ideas, which can at times strain social understanding and interactions (Piedmont, Sherman, & Sherman, 2012). Such a disposition may possibly undermine a mutual understanding and shared case formulation, whereas therapists with low openness scores may adhere to a more

conventional and accessible way of formulating problems and treatment plans in CBT, thus better promoting an agreement on the goals and tasks of therapy. Chapman et al. (2009) argued that extreme levels in the *openness to ideas* facet may be indicative of therapists who have a propensity to take an overly intellectual stance, which may be intimidating or perplexing for patients with more conventional levels of openness. Furthermore, openness was unrelated to treatment outcomes in the PWP group. Perhaps this is also explained by the differences between guided self-help and more intensive CBT. Guided self-help is highly didactic and accessible, often supported by plain-language reading materials, and does not allow much time for overly complex or intellectual examination of patients' problems.

Our interpretations suggest that therapists' personality traits and interpersonal style may possibly interfere with some aspects of CBT practice, such as forming a collaborative working alliance and facilitating the effective modification of maladaptive beliefs and behaviours. These specific aspects of CBT practice are captured in competency measures such as those applied in this study. However, these competency measures also include other domains which together render a total competency score. Unfortunately, data on specific competencies was unavailable to study more specific aspects of CBT practice and their potential interaction with therapist personality traits. Although a global competency measure (e.g. CTS-R) is evidently unrelated to treatment outcomes, it remains to be seen if specific competencies may be related to treatment outcomes and therapists' personality or interpersonal style.

The present study has a number of strengths, including a large and adequately powered sample to investigate therapist effects, the independent rating of therapists' technical and theoretical competences by expert assessors, and the application of

rigorous multilevel and machine learning analyses to examine therapist-level predictors of patient outcomes. This is also the first study to demonstrate the utility of meta-analytic methods to investigate therapist effects. The use of forest plots enables a more precise understanding of the relative performance of different therapists, expressed in a clinically intuitive effect size metric. This method revealed that, despite the apparently modest variability in outcomes generally attributable to the therapist nesting variable (ICC = 1% to 3%) *relative to other sources of variability* (i.e., patient-level variables and other unmeasured factors), the systematic variability in performance *between therapists* was more substantial (I^2 = 5.75%). The application of precision funnel plots (Figure 2B) also revealed that, while most therapists' performance conformed to a theoretical benchmark (i.e., *the average therapist*), there were evident differences in performance *between therapists* with similar caseload sizes and case-mix. These methods to examine relative outliers (referred to as positive and negative deviant cases) are well established in the field of hospital and healthcare service performance benchmarking (Schmidtke et al., 2017; Spiegelhalter, 2005), but until now have not been used to study therapist effects.

A number of limitations are also relevant to the interpretation of these findings. Despite the availability of depression and anxiety outcomes data for a large clinical sample, no additional patient-level information was available. The availability of multivariable patient-data (e.g., demographic, clinical and characterological features) may have yielded more precise benchmarks and performance effect sizes. Furthermore, we do not know if the assignment of patients to therapists was quasi-random or if systematic patient-allocation took place. The sample was limited to PWP and CBT therapists during training and the early stage of their post-qualification practice, and therefore the study lacked a comparison with more experienced therapists or with other

psychotherapy models. The post-qualification period was short and no repeated measures of personality or post-qualification competency were taken. Further limitations concern the lack of established psychometric properties for the OSCE measure used to rate clinical competence in PWPs. Nevertheless, a psychometrically established measure of competence (CTS-R; Reichelt et al., 2003) applied in the CBT group was unrelated to clinical outcomes. Importantly, we recognise that studies examining the relationships between therapists' personality with treatment processes and outcomes are scarce and mostly inconsistent in their findings. Our interpretations of the present findings are therefore hypothetical and subject to future replication.

In conclusion, current evidence suggests that therapists' personality traits influence their relationships with patients (alliance) and are associated with depression and anxiety treatment outcomes. Future studies could collect the same measures of personality in therapists and patients to examine the potential influence of similar, divergent, or complementary (e.g., compliant matched to dominant) personality styles on outcomes such as treatment dropout and changes in symptoms and interpersonal functioning.

Data sharing policy: In line with the requirements of the ethics review board for this study, requests for access to data are to be made in writing to a.branson@reading.ac.uk.

Funding: AB received funding from the British Association of Behavioural and Cognitive Psychotherapies. All research at Great Ormond Street Hospital NHS Foundation Trust and UCL Great Ormond Street Institute of Child Health is made possible by the NIHR Great Ormond Street Hospital Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

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*, 755-768. doi: 10.1002/jclp.20583

Antonuccio, D. O., Lewinsohn, P. M., & Steinmetz, J. L. (1982). Identification of therapist differences in a group treatment for depression. *Journal of Consulting and Clinical Psychology, 50*, 433-435. doi/10.1037/0022-006X.50.3.433

Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 258–297). New York, NY: Wiley.

Banham, J. A. & Schweitzer, R. D. (2016) Trainee-therapists are not all equal: Examination of therapeutic efficiency, effectiveness and early client dropout after 12 months of clinical training. *Psychology and Psychotherapy: Theory, Research and Practice, 89*, 148-162. doi: 10.1111/papt.12071

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology, 44*(1), 1-26. doi: 10.1111/j.1744-6570.1991.tb00688.x

Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., & Wong, E. (2004). Therapist variables. In M. J. Lambert, A. E. (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*. (Fifth edition). New York: Wiley.

Blackburn, I., James, I., Milne, D., Baker, C., Standart, S., Garland, A., et al. (2001). The Revised Cognitive Therapy Scale (CTS-R): psychometric properties.

Behavioural and Cognitive Psychotherapy, 29, 431–446. doi: 10.1017/S135246580100404

Bradley, E. H., Curry, L. A., Ramanadhan, S., Rowe, L., Nembhard, I. M., & Krumholz, H. M. (2009). Research in action: using positive deviance to improve quality of health care. *Implementation Science, 4*, 25.

Branson, A., Myles, P., Mahdi, M., & Shafran, R. (2018). The relationship between competence and patient outcome with low-intensity cognitive behavioural interventions. *Behavioural and Cognitive Psychotherapy, 46*, 101-114. doi: 10.1017/S1352465817000522

Branson, A., & Shafran, R. (2015). Therapist characteristics and their effect on training outcomes: what counts? *Behavioural and Cognitive Psychotherapy, 43*, 374-380. doi: 10.1017/S1352465814000642

Branson, A., Shafran, R., & Myles, P. (2015). Investigating the relationship between competence and patient outcome with CBT. *Behaviour Research and Therapy, 68*, 19-26. doi: 10.1016/j.brat.2015.03.002

Borenstein, M., Hedges, L., & Rothstein, H. (2007). *Meta-analysis: Fixed effect vs. random effects*. Available at: https://www.meta-analysis.com/

Casari, L. M., Ison, M. S., & Gomez, B. M. M. (2019). Personal style of the therapist and personality dimensions in a sample of Argentinian therapists. *Research in Psychotherapy: Psychopathology, Process and Outcome, 22*, 292-307. doi: 10.4081/rippo.2019.362

Castonguay, L., & Hill, C. E. (Eds.). (2017). How and why are some therapists better than others? Understanding therapist effects. Washington, DC: American Psychological Association. https://doi.org/10.1037/0000034-000

Chapman, B. P., Talbot, N., Tatman, A. W., & Britton, P. C. (2009). Personality Traits and the Working Alliance in Psychotherapy Trainees: An Organizing Role for the Five Factor Model? *Journal of Social and Clinical Psychology, 28*, 577-596. doi: 10.1521/jscp.2009.28.5.577.

Chow, D. L., Miller, S. D., Seidel, J. A., Kane, R. T., Thornton, J. A., & Andrews, W. P. (2015). The role of deliberate practice in the development of highly effective psychotherapists. *Psychotherapy, 52*, 337-345. doi: 10.1037/pst0000015

Clark, D. M. (2018). Realizing the Mass Public Benefit of Evidence-Based Psychological Therapies: The IAPT Program. *Annual Review of Clinical Psychology, 14*, 159-183. doi: 10.1146/annurev-clinpsy-050817-084833

Cologon, J., Schweitzer, R. D., King, R., & Nolte, T. (2017). Therapist reflective functioning, therapist attachment style and therapist effectiveness. *Administration and Policy in Mental Health and Mental Health Services Research, 44*, 614-625. doi: 10.1007/s10488-017-0790-5

Costa, P. T., & McCrae, R. R. (2006). *NEO Revised: NEO Personality Inventory Manual* (UK ed.). Oxford: Hogrefe.

Coull, G., & Morris, P. G. (2011). The clinical effectiveness of CBT-based guided self-help interventions for anxiety and depressive disorders: a systematic review. *Psychological Medicine, 41*(11), 2239-2252. doi: 10.1017/S0033291711000900

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *The Canadian Journal of Psychiatry, 58*(7), 376-385. doi: 10.1177/070674371305800702

Cuijpers, P., Donker, T., van Straten, A., Li, J., & Andersson, G. (2010). Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety

disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychological Medicine, 40*(12), 1943-1957. doi: 10.1017/S0033291710000772

Department of Health. (2011). *National curriculum for Cognitive Behavioural Therapy courses*, Second edition, updated and revised March 2011. London: Kings College London.

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

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

J. W. Hart, M. F. Stasson, J. M. Mahoney and P. Story, "The Big Five and Achievement Motivation: Exploring the Relationship between Personality and a Two-Factor Model of Motivation," Individual Differences Research, Vol. 5, No. 4, 2007, pp. 267-274.

He, Y., Donnellan, B., Mendoza, A. M. (2019). Five-factor personality domains and job performance: A second order meta-analysis. *Journal of Research in Personality, 82*, 2-13. doi.org/10.1016/j.jrp.2019.103848

Heinonen, E. & Nissen-Lie, Helene A (2019). The professional and personal characteristics of effective psychotherapists: a systematic review. *Psychotherapy Research*. doi:10.1080/10503307.2019.1620366

Higgins, J. T., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions version 5.1.0*. London: The Cochrane Collaboration.

Hofmann, S. G., & Smits, J. A. (2008). Cognitive-behavioral therapy for adult anxiety disorders: a meta-analysis of randomized placebo-controlled trials. *The Journal of Clinical Psychiatry, 69,* 621-632. doi: 10.4088/JCP.v69n0415

Johns, R. G., Barkham, M., Kellett, S., & Saxon, D. (2018). A systematic review of therapist effects: A critical narrative update and refinement to review. *Clinical Psychology Review*. doi: 10.1016/j.cpr.2018.08.004.

Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology, 87,* 530-541. doi: 10.1037/0021-9010.87.3.530

Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. Journal of Applied Psychology, 87, 797. doi: 10.1037//0021-9010.87.4.797

Kroenke, K., Spitzer, R. L. and Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine, 16, 606-613. doi: 10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., Spitzer, R. L., Williams, J. B., Monahan, P. O. & Löwe, B. (2007). Anxiety Disorders in Primary Care: Prevalence, Impairment, Comorbidity, and Detection. Anxiety Disorders in Primary Care. *Annals of Internal Medicine, 146,* 317-325. doi: 10.7326/003-4819-146-5-200703060-00004

Lado, M., Alonso, P. (2017). The Five-Factor model and job performance in low complexity jobs: A quantitative synthesis. *Journal of Work and Organizational Psychology, 33,* 175–182. doi.org/10.1016/j.rpto.2017.07.004

Loades, M. E., & Armstrong, P. (2016). The challenge of training supervisors to use direct assessments of clinical competence in CBT consistently: a systematic

review and exploratory training study. *The Cognitive Behaviour Therapist, 9*, e27. doi: 10.1017/S1754470X15000288

Loades, M. E., & Myles, P. J. (2016). Does a therapist's reflective ability predict the accuracy of their self-evaluation of competence in cognitive behavioural therapy? *The Cognitive Behaviour Therapist, 9*, 1-14. doi: 10.1017/S175440X16000027

Minami, T., Serlin, R. C., Wampold, B. E., Kircher, J. C., & Brown, G. S. (2008). Using clinical trials to benchmark effects produced in clinical practice. *Quality and Quantity, 42*, 513-525. doi: 10.1007/s11135-006-9057-z

National IAPT Team. (2015). *National curriculum for the education of Psychological Wellbeing Practitioners*, Third edition. London: University College London.

Nissen-Lie, H. A., Monsen, J. T., Ulleberg, P., & Rønnestad, M. H. (2013). Psychotherapists' self-reports of their interpersonal functioning and difficulties in practice as predictors of patient outcome. *Psychotherapy Research, 23*, 86-104. doi: 10.1080/10503307.2012.735775.

Nissen-Lie, H. A., Rønnestad, M. H., Høglend, P. A., Havik, O. E., Solbakken, O. A., Stiles, T. C., & Monsen, J. T. (2017). Love yourself as a person, doubt yourself as a therapist? *Clinical Psychology & Psychotherapy, 24*, 48-60. doi: 10.1002/cpp.1977

Norcross, J. C., & Lambert, M. J. (Eds.). (2019). *Psychotherapy relationships that work, 3rd Edition. Volume 1: Evidence-based therapist contributions*. New York: Oxford University Press.

Okiishi, J., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An empirical analysis of therapist effects. *Clinical Psychology & Psychotherapy, 10*, 361–373. doi: 10.1002/cpp.383

Perunovic, M., & Holmes, J. G. (2008). Automatic accommodation: The role of personality. *Personal Relationships, 15*, 57-70. Doi: 10.1111/j.1475-6811.2007.00184.x

Peters-Scheffer, N., Didden, R., Korzilius, H., & Sturmey, P. (2013). Therapist characteristics predict discrete trial teaching procedural fidelity. *Intellectual and Developmental Disabilities*, 51, 263-272. doi: 10.1352/1934-9556-51.4.263

Piedmont, R. L., Sherman, M. F., & Sherman, N. C. (2012). Maladaptively high and low openness: The case for experiential permeability. *Journal of Personality, 80*, 1641-1668. doi: 10.1111/j.1467-6494.2012.00777.x

Raudenbush, S. W. (1993). Hierarchical linear models and experimental design. In L. Edwards (Ed.), *Applied analysis of variance in behavioral science* (pp. 459-496). New York: Marcel Dekker.

Reichelt, F. K., James, I. A., & Blackburn, I. M. (2003). Impact of training on rating competence in cognitive therapy. *Journal of Behavior Therapy and Experimental Psychiatry, 34*, 87-99. doi: 10.1016/S0005-7916(03)00022-3

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

Richards, D. A., and Whyte, M. (2009). *Reach Out. National Programme Educator Materials to Support the Delivery of Training for Psychological Wellbeing Practitioners Delivering Low Intensity Interventions* (2 ed.). London: Rethink.

Rieck, T., & Callahan, J. L. (2013). Emotional intelligence and psychotherapy outcomes in the training clinic. *Training and Education in Professional Psychology, 7*, 42–52.

Rodriguez, J. D., Perez, A., and Lozano, J. A. (2010). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*, 569-575. doi: 10.1109/TPAMI.2009.187

Rousmaniere, T.G. (2016). *Deliberate practice for psychotherapists: A guide to improving clinical effectiveness.* Routledge Press (Taylor & Francis).

Roth, A. D., and Pilling, S. (2008). Using an evidence-based methodology to identify the competencies required to deliver effective cognitive and behavioural therapy for depression and anxiety disorders. *Behavioural and Cognitive Psychotherapy, 36*, 129 –147. doi: 10.1017/S1352465808004141

Sackett, P. R., & Mullen, E. J. (1993). Beyond formal experimental design: Towards an expanded view of the training evaluation process. *Personnel Psychology, 46*, 613-627. doi: 10.1111/j.1744-6570.1993.tb00887.x

Saxon, D., Barkham, M., Foster, A., & Parry, G. (2017). The contribution of therapist effects to patient dropout and deterioration in the psychological therapies. *Clinical Psychology & Psychotherapy, 24*, 575-588. doi: 10.1002/cpp.2028

Schiefele, A.K., Lutz, W., Barkham, M., Rubel, J., Böhnke, J., Delgadillo, J., Kopta, M., Schulte, D., Saxon, D., Nielsen, S.L. & Lambert, M.J. (2017). Reliability of therapist effects in practice-based psychotherapy research: A guide for the planning of future studies. *Administration and Policy in Mental Health and Mental Health Services Research, 44*, 598-613. doi: 10.1007/s10488-016- 0736-3

Schmidtke, K. A., Poots, A. J., Carpio, J., Vlaev, I., Kandala, N. B., & Lilford, R. J. (2017). Considering chance in quality and safety performance measures: an analysis of performance reports by boards in English NHS trusts. *BMJ Quality and Safety, 26*, 61-69.

Spiegelhalter, D. J. (2005). Funnel plots for comparing institutional performance. *Statistics in Medicine, 24*, 1185-1202. doi: 10.1002/sim.1970

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

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Methodological, 58*, 267–288. https://www.jstor.org/stable/2346178

Wampold, B. E., Baldwin, S., Martin Grosse Holtforth, M., & Zac E. Imel, Z. E. (2002). Who works for whom and why? Integrating therapist effects analysis into psychotherapy outcome and process research. In L. E. Castonguay & C. A. Hill. Washington; APA

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The research evidence for what works in psychotherapy*, 2nd ed. New York, NY: Routledge.

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*, 200-211. doi: 10.1037/a0018912

Table 1. Sample characteristics

Characteristics	PWP cohort	CBT cohort	Test statistic	*p*
Therapists' demographics	N = 36	N = 33		
Age M (SD; range)	31.58 (8.42; 22 – 54)	39.72 (8.63; 24 – 54)	U (69) = 919.00	< .001
Females (%)	72.2%	69.7%	x^2 (1) = 0.53	.82
Ethnicity				
White British (%)	61.1%	81.8%	x^2 (1) = 3.59	.06
Other (%)	38.9%	18.2%		
Therapists' personality traits (NEO PI-R transformed T-scores)				
Neuroticism M (SD; range)	51.70 (10.94; 33.52 – 73.73)	48.15 (8.65; 29.00 – 72.50)	*t* (65.63) = 1.50	.14
Extraversion M (SD; range)	50.43 (11.31; 28.36 – 74.12)	49.53 (8.49; 27.29 – 62.95)	*t* (67) = 0.37	.71
Openness M (SD; range)	49.52 (10.01; 31.41 – 67.59)	50.53 (10.12; 25.20 – 65.33)	*t* (67) = -0.42	.68
Agreeableness M (SD; range)	49.37 (10.31; 27.67 – 69.49)	50.68 (9.76; 27.67 – 65.30)	*t* (67) = -0.54	.59
Conscientiousness M (SD; range)	50.19 (9.73; 30.63 – 65.12)	49.79 (10.43; 27.64 – 77.11)	*t* (67) = 0.17	.87
Clinical experience and competence				
Years of experience M (SD; range)	2.11 (2.34; 0 – 11)	10.18 (6.93; 3 – 28)	U (69) = 1112.00	< .001
Reflective ability	61.59 (7.82; 45 – 77)	55.68 (7.27; 40 – 71)	*t* (67) = 3.24	.002
OSCE M (SD; range)	74.02 (8.63; 58.00 – 91.33)			
CTS-R M (SD; range)		51.60 (7.35; 39.67 – 69.33)		
Patients' clinical characteristics	N = 2969	N = 1083		
Pre-treatment PHQ-9 mean (SD)	11.72 (5.80; 0 – 27)	14.12 (6.53; 0 – 27)	U (4052) = 1963804.00	< .001
Post-treatment PHQ-9 mean (SD)	6.85 (5.69; 0 – 27)	7.67 (6.64; 0- 27)	U (4052) = 1678539.00	.03
Pre-treatment GAD-7 mean (SD)	10.85 (5.15; 0 – 21)	12.92 (5.34; 0 – 21)	U (4052) = 1973841.50	< .001
Post-treatment GAD-7 mean (SD)	6.31 (5.14; 0 – 27)	6.87 (5.65; 0 – 21)	U (4052) = 1672942.00	.04
Mean number of therapy sessions (SD)	4.47 (2.14; 2 – 17)	9.15 (5.72; 2 – 32)	U (4052) = 2420207.50	< .001

Notes: M = mean; SD = standard deviation; PWP = psychological wellbeing practitioners; CBT = cognitive behavioural therapists; OSCE = observed standardized clinical examination; CTS-R = cognitive therapy scale – revised; PHQ-9 = measure of depressions symptoms; GAD-7 = measure of anxiety symptoms; U = Mann-Whitney U test; x^2 = chi square test; *t* = Student's t-test; all NEO PI-R personality traits are expressed in raw scores (which were used in all analyses)

Table 2. Multilevel models examining the role of therapists' competence, experience and personality features

	PWP group (N = 2969, k = 36)				CBT group (N = 1083, k = 33)			
	PHQ-9 model		GAD-7 model		PHQ-9 model		GAD-7 model	
Fixed effects	**B (SE)**	***p***	**B (SE)**	***p***	**B (SE)**	***p***	**B (SE)**	***p***
Intercept	6.91 (0.31)	< .001	6.41 (0.22)	< .001	7.74 (0.56)	< .001	6.77 (0.51)	< .001
Pre-treatment PHQ-9	0.46 (0.02)	< .001	0.13 (0.02)	< .001	0.48 (0.04)	< .001	0.20 (0.03)	< .001
Pre-treatment GAD-7	0.05 (0.02)	.02	0.36 (0.02)	< .001	0.03 (0.04)	.47	0.25 (0.04)	< .001
Age	-0.04 (0.02)	.06	-0.03 (0.02)	.10	-0.04 (0.04)	.24	-0.01 (0.04)	.67
Gender	0.16 (0.37)	.66	0.27 (0.27)	.31	0.20 (0.74)	.79	0.21 (0.66)	.75
Years of experience	-0.03 (0.07)	.70	0.01 (0.05)	.83	0.03 (0.05)	.48	0.05 (0.04)	.25
Clinical competence*	-0.02 (0.07)	.24	-0.01 (0.01)	.40	-0.07 (0.06)	.23	-0.06 (0.06)	.31
Reflective ability	-0.03 (0.03)	.33	-0.03 (0.02)	.12	0.004 (0.06)	.96	0.01 (0.06)	.84
Neuroticism	0.006 (0.007)	.40	0.004 (0.005)	.45	-0.006 (0.02)	.74	-0.02 (0.02)	.23
Extraversion	-0.002 (0.008)	.82	-0.005 (0.006)	.41	-0.002 (0.02)	.92	-0.002 (0.02)	.93
Openness	-0.01 (0.01)	.24	-0.01 (0.007)	.10	0.04 (0.02)	.02	0.04 (0.02)	.03
Agreeableness	0.02 (0.01)	.03	0.02 (0.008)	.01	-0.04 (0.03)	.22	-0.05 (0.03)	.05
Conscientiousness	0.006 (0.009)	.47	0.001 (0.006)	.82	0.01 (0.02)	.51	0.004 (0.01)	.77
Stage of training 2	0.11 (0.26)	.66	-0.10 (0.23)	.66	-0.44 (0.41)	.28	-0.08 (0.36)	.82
Stage of training 3	-0.60 (0.26)	.02	-0.87 (0.23)	< .001	-1.60 (0.66)	.01	-1.31 (0.58)	.02

Notes: Dependent variables = post-treatment PHQ-9 or GAD-7; PHQ-9 = measure of depressions symptoms; GAD-7 = measure of anxiety symptoms; PWP = psychological wellbeing practitioners; CBT = cognitive behavioural therapists; B = regression coefficient; SE = standard error; Gender reference (0) category = male; clinical competence measure for PWP cohort = Observed Structured Clinical Evaluation (OCSE); clinical competence measure for CBT cohort = Cognitive Therapy Scale – Revised (CTS-R); stage of training reference (0) category = 1-year training period, stage 2 = 6 months post-training, stage 3 = >6 months post-training; NEO PI-R personality traits were entered in their original raw score format

Table 3. Multilevel models comparing personality traits (domain scores) versus facets (selected items)

	PWP group				CBT group			
	Trait-based model predicting PHQ-9		**Trait-based model predicting GAD-7**		**Trait-based model predicting PHQ-9**		**Trait-based model predicting GAD-7**	
Goodness-of-fit	-2LL = 18747.84		-2LL = 18131.69		**Goodness-of-fit**	-2LL = 7163.47		-2LL = 6807.38	
Fixed effects	**B (SE)**	***p***	**B (SE)**	***p***	**Fixed effects**	**B (SE)**	***p***	**B (SE)**	***p***
Intercept	6.91 (0.14)	< .001	6.38 (0.12)	< .001	Intercept	7.41 (0.34)	< .001	6.60 (0.29)	< .001
Agreeableness	0.02 (0.01)	.01	0.02 (0.01)	.01	Openness	0.03 (0.02)	.13	0.02 (0.02)	.24
	Facet-based model predicting PHQ-9		**Facet-based model predicting GAD-7**		**Facet-based model predicting PHQ-9**		**Facet-based model predicting GAD-7**	
Goodness-of-fit	-2LL = 18739.19		-2LL = 16268.33		**Goodness-of-fit**	-2LL = 7155.92		-2LL = 6132.81	
Fixed effects	**B (SE)**	***p***	**B (SE)**	***p***	**Fixed effects**	**B (SE)**	***p***	**B (SE)**	***p***
Intercept	6.90 (0.13)	< .001	7.02 (0.13)	< .001	Intercept	7.42 (0.32)	< .001	7.33 (0.27)	< .001
Compliance	0.42 (0.14)	< .01	0.38 (0.13)	< .01	Ideas	0.71 (0.34)	.04	0.45 (0.30)	.14
					Feelings			0.42 (0.24)	.08

Notes: Dependent variables = post-treatment PHQ-9 or GAD-7; PHQ-9 = measure of depressions symptoms; GAD-7 = measure of anxiety symptoms; PWP = psychological wellbeing practitioners; CBT = cognitive behavioural therapists; B = regression coefficient; SE = standard error; -2LL = -2 log-likelihood (a lower statistic indicates better goodness-of-fit); NEO PI-R personality traits and facets were entered in their original raw score format

Figure 1. Forest plot of therapist effects

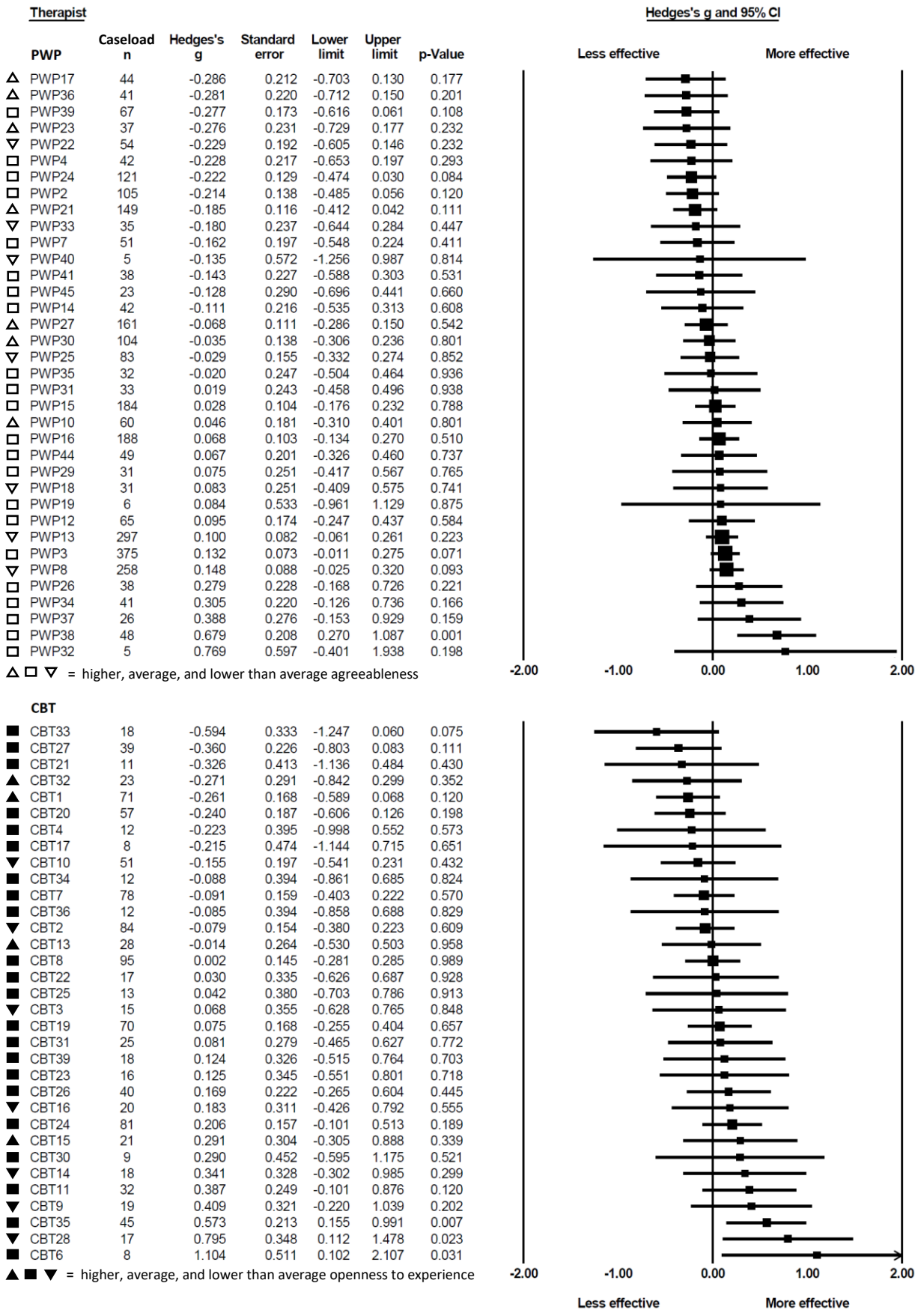

Figure 2. Funnel plots examining systematic heterogeneity in treatment effects