



This is a repository copy of *Therapist effects and IAPT Psychological Wellbeing Practitioners (PWPs): A multilevel modelling and mixed methods analysis*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/103217/>

Version: Accepted Version

Article:

Green, H., Barkham, M., Kellett, S. et al. (1 more author) (2014) Therapist effects and IAPT Psychological Wellbeing Practitioners (PWPs): A multilevel modelling and mixed methods analysis. *Behaviour Research and Therapy*, 63. pp. 43-54. ISSN 0005-7967

<https://doi.org/10.1016/j.brat.2014.08.009>

Article available under the terms of the CC-BY-NC-ND licence
(<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

RUNNING HEAD: THERAPIST EFFECTS & PWPS

Therapist effects and IAPT Psychological Wellbeing Practitioners (PWPs):

A multilevel modelling and mixed methods analysis

Helen Green

Clinical Psychology Unit, University of Sheffield

Michael Barkham

Centre for Psychological Services Research, University of Sheffield

Stephen Kellett

Centre for Psychological Services Research, University of Sheffield

Sheffield Health and Social Care NHS Foundation Trust

David Saxon

Centre for Psychological Services Research

School of Health and Related Research, University of Sheffield

Corresponding author:

Helen Green¹

Clinical Psychology Unit, Department of Psychology

University of Sheffield, Western Bank, Sheffield, S10 2TN, Tel: (+44) 0114 222 6610

¹ Address: Rotherham, Doncaster and South Humber NHS Foundation Trust, East Dene Centre, Lansdowne Road, Doncaster, DN2 6QN. Email: Helen.Green@rdash.nhs.uk. Tel: (+44) 01302 734050, Fax: (+44) 01302 734077

ABSTRACT

The aim of this research was (1) to determine the extent of therapist effects in Psychological Wellbeing Practitioners (PWPs) delivering guided self-help in IAPT services and (2) to identify factors that defined effective PWP clinical practice. Using patient (N=1,122) anxiety and depression outcomes (PHQ-9 and GAD-7), the effectiveness of N = 21 PWPs across 6 service sites was examined using multi-level modelling. PWPs and their clinical supervisors were also interviewed and completed measures of ego strength, intuition and resilience. Therapist effects accounted for around 9 per cent of the variance in patient outcomes. One PWP had significantly better than average outcomes on both PHQ-9 and GAD-7 while 3 PWPs were significantly below average on the PHQ-9 and 2 were below average on the GAD-7. Computed PWP ranks identified quartile clusters of most and least effective PWPs. More effective PWPs generated higher rates of reliable and clinically significant change and displayed greater resilience, organisational abilities, knowledge and confidence. Resilience appears to play a critical role in effective PWP work and warrants further investigation. Study weaknesses are identified and methodological considerations for future studies concerning therapist effects are provided.

Keywords

Therapist effects, Psychological Wellbeing Practitioners, IAPT, resilience

Introduction

In recent years the landscape of psychological therapies in the UK has been transformed with the introduction of the Improving Access to Psychological Therapies (IAPT) programme. IAPT was introduced in response to the Depression Report (Layard et al., 2006), which detailed the personal and societal cost of the lack of access to evidence-based psychological therapies for people experiencing common mental health problems (depression and anxiety disorders). Nascent IAPT organisational models were tested in two demonstration sites from 2006 with effective results (Clark et al., 2009; Parry et al., 2011). Subsequently IAPT was rolled out nationally in the UK from 2008 (CSIP Choice & Access Team, 2008). A central feature of IAPT services is therefore the availability and delivery of treatments consistent with the NICE guidelines for depression and anxiety (Clark, 2011). NICE recommends for mild-moderate depression and anxiety disorders (excluding PTSD and social anxiety disorder), the provision of stepped-care service delivery models. Stepped care entails the delivery of increasing intense psychological treatments that are delivered sequentially and according to patient need (Bower & Gilbody, 2008). A substantial proportion of patients therefore initially receive 'low intensity' treatments (such as guided self-help or psychoeducational classes) delivered at step 2. Those who fail to respond are subsequently stepped up to more traditional face-to-face 'high intensity' psychological therapies, delivered at step 3. Stepping up and down aims to ensure the seamless transition of patients to and from Primary and Secondary Care services (CSIP, 2008). Although the exact configurations of IAPT services differ (Gyani, Shafron, Layard, & Clark, 2013), many have extended the range of therapies to patients with anxiety and depression into both Primary and Secondary Care

The IAPT initiative has therefore depended on recruiting a new mental health workforce (Robinson, Kellett, King, & Keating, 2012). A workforce development goal has

been the creation of the Psychological Wellbeing Practitioner (PWP) role, who deliver assessments and low intensity treatments at step 2 (CSIP, 2008). To qualify, PWPs complete a 1-year Post-Graduate Certificate which is driven by a national curriculum (Richards & Whyte, 2009), to ensure consistency of learning and service delivery. Trainee PWPs are employed by IAPT services and spend one day per week in their academic base and four days per week in service, assessing and treating patients under close clinical and case management supervision. Academic assessment is competency based and consists largely of observed structured clinical exams assessing the delivery of various low intensity assessment and treatment skills (Richards & Whyte, 2009).

On qualifying, PWPs work at step 2 of the IAPT stepped care model, treating patients with mild to moderate anxiety and depression using guided self-help. Work at step 2 is characterised by its 'low contact-high volume' approach (Clark et al., 2009). PWPs receive one hour per week of IT-driven case management supervision using outcomes from IAPT minimum dataset to ensure high volume, and regular clinical supervision to ensure low contact (i.e., fidelity to low intensity methods). The core rationale of low intensity work is explicitly cognitive-behavioural in origin, but with the PWP role being that of a 'coach' as opposed to traditional therapist (Turpin, 2010). Therefore PWPs also use non-traditional methods such a telephone delivery and e-clinics (alongside one-to-one and group psychoeducational sessions) to deliver the seven core self-help treatment protocols that constitute the PWP clinical method (Richards & Whyte, 2009). Due to the relatively recent introduction of PWPs, there is a paucity of research on PWPs in general and no extant research regarding potential therapists effects. This is despite PWPs increasing in numbers and the organisational prominence of the role (IAPT, 2012).

Therapist effects: trials and routine practice

Within the broader psychological therapies literature, there are differing, sometimes opposing, views on the issue of whether therapist effects are a significant factor accounting for patient outcomes. In addition, there is also a range of methodological issues that arise when considering therapist effects. In terms of substantive findings, differences in effectiveness between therapists have been reported in some RCTs (e.g., Huppert, Bufka, Barlow, Gorman, & Shear, 2001), while others have reported small or non-significant effects (e.g., Clark et al., 2006; Wilson, Wilfley, Agras, & Bryson, 2011). These conflicting results are exemplified in two independent analyses of the same data, drawn from the National Institute for Mental Health's Treatment of Depression Collaborative Research Project (NIMH TDCRP; Elkin et al., 1989). Using the same data, but employing differing analytic techniques, one group of researchers found a significant therapist effect (Kim, Wampold, & Bolt, 2006), whilst another group reported no therapist effect (Elkin, Falconnier, Martinovich, & Mahoney, 2006). These conflicting findings have been attributed to differences in the analytical methods employed and the small number of therapists involved (Crits-Christoph & Gallop, 2006; Soldz, 2006). Thompson et al (2012) have provided guidance on investigating therapist effects within trials, with particular emphasis on smaller sized trials.

In contrast to the results from RCTs, naturalistic studies drawing on data from routine practice have tended to indicate consistent therapist effects in the region of 5%-8% of the variance in outcome explained (e.g., Baldwin & Imel, 2013, Brown, Lambert, Jones, & Minami, 2005; Lutz, Scott, Martinovich, Lyons, & Stiles, 2007; Okiishi, Lambert, Eggett, Nielson, Vermeersch & Dayton, 2006; Saxon & Barkham, 2012). In a context where therapy is being delivered in routine practice and therapists may not be implementing protocol-driven interventions, the resulting variability in therapist effects appears understandable. The IAPT initiative provides the setting for the unique integration of both evidenced-based protocol-

driven interventions and the delivery of these protocols during routine practice. It might be expected that the extent of variability in PWPs within IAPT services would be therefore be limited. Accordingly, this sample was the focus of the current study.

Therapist effects: methodological issues

Where a hierarchical structure exists (i.e. the outcomes for patients seen by the same practitioner are likely to be similar in some way and different from the outcomes for patients seen by another practitioner), multi-level modelling (MLM) is advocated as an appropriate method for assessing higher level (in this case, PWP) effects (Goldstein & Spiegelhalter, 1996, Raudenbush & Bryk, 2002, Wampold & Brown, 2005). Other methods include generalised estimating equations which estimate the parameters of a generalised linear model (Hardin & Hilbe, 2003). MLM allows for the partitioning of the total outcome variance between level 1 (patient level) and level 2 (PWP level) with the proportion of total variance at level 2 equating to the therapist effect (Raudenbush & Bryk, 2002; Wampold & Brown, 2005). MLM has been extensively used in other areas of research, most notably in the study of comparative effectiveness of schools, where pupils are nested within teachers or classes, which are in turn nested within schools (Goldstein & Spiegelhalter, 1996). Recent studies of higher intensity treatments in psychological therapy services have also used MLM to estimate therapist effects (e.g. Wampold & Brown, 2005, Saxon & Barkham, 2012).

The recommendation that in MLM studies the N of therapists is greater than 30 (e.g., Soldz, 2006) is a challenge where smaller numbers of therapists are available or sampled. For example, Wiborg, Knoop, Wensing, and Bleijenberg (2012) reported a therapist effect of 21% with a sample of 10 practitioners, but recommended a further study with a larger sample. In contrast, Almlöv et al. (2010) found no practitioner effect during the delivery of internet therapy, but again cited lack of power as a contributing factor. One approach to account for sample size when producing effect estimates, is to provide confidence intervals

around estimates. Therefore, in the present study, Markov Chain Monte Carlo (MCMC) procedures were adopted. This simulation-based procedure produces a very large number of estimates from which the median therapist effect can be derived, along with a 95% ‘probability interval’ (PrI) - analogous to 95% confidence intervals. Because of the limited number of PWPs in the current study, patient pre-treatment severity was also controlled for in the modelling. Pre-treatment severity is strongly associated with outcome (Garfield, 1994) and once this is taken into account other variables have little predictive value (Okiishi et al., 2006, Luborsky, McLellan, Diguier, Woody, & Seligman, 1997).

Components of effective practitioners

Due to the sheer number of PWPs in clinical practice in IAPT services (CSIP Choice & Access Team, 2008; Turpin, 2010), the uniqueness of the PWP clinical method (Richards & Whyte, 2008) and the high throughput of patients at step 2 (Parry et al., 2011), generation of knowledge concerning factors contributing to effective PWP practice is vital. Identifying factors that create or define the work of effective therapists (using higher intensity therapies) has previously proved difficult due to inherent methodological difficulties (Hubble, Duncan, & Miller, 1999). Studies for example have demonstrated that talking therapies can be effective whether the therapists are qualified professional therapists (e.g., Gibbons et al., 2010) or intern/trainees (e.g., Forand, Evans, Haglin, & Fishman, 2011). Research has either focussed on collated patient outcomes and whether variability between therapists exists (i.e., therapist effects) or on the features/characteristics of therapists themselves (e.g., Jennings & Skovholt, 1999; Najavitis & Strupp, 1994). Relatively few studies have simultaneously studied both therapist effects and the features of the practitioners – this was an aim for the current study. Previous research has tended to be focal to traditional psychotherapy roles, characterised as working in a ‘high contact, low volume’ style (Clark et al., 2009).

Those studies that have focussed on therapist characteristics have reported a number of in-session and out-of-therapy factors related to more effective practice. In session aspects include enhanced relational skills (Jennings & Skovholt, 1999), greater empathy (Lafferty, Beutler & Crago, 1989), effective therapeutic alliances (Luborsky, 1985) and showing more warmth, affirmation, understanding, active helping and protecting (Najavitis & Strupp, 1994). Such factors are pan-theoretical and are often referred to as common factors (Weinberger, 1993) - relying to some extent on therapist intuition (Welling, 2005). Whilst research into the mechanisms and use of clinical intuition is growing, whether intuition forms an aspect of effective practice remains somewhat untested (Rea, 2001; Welling, 2005). Given the nature of the coaching role that PWPs take up in relation to their patients when delivering guided self-help (Richards & Whyte, 2008), we hypothesised that higher levels of intuition would not be associated with better patient-reported outcomes.

Out-of-therapy factors associated with effective practitioners include good emotional adjustment (Luborsky, 1985), being self-critical of therapeutic actions (Najavitis & Strupp, 1994), self-reflection (Jennings & Skovholt, 1999), an emphasis on hard work and openness to feedback (Miller, Hubble, & Duncan, 2008). Two factors that may capture this quality of openness to learning in therapists are ego strength and resilience. Ego strength is defined as the ability to maintain a sense of self in the face of challenges without becoming overwhelmed (Markstrom, Sabino, Turner, & Berman, 1997). Resilience is defined as characteristics that enable coping during and bouncing back subsequently from adverse situations (Rutter, 1993). In light of the high clinical volume demands of the PWP role, we hypothesised that ego strength and resilience would be associated with patient outcomes.

Accordingly, the present research aimed to utilise IAPT data with a sample of PWPs in routine practice to investigate (1) the extent of therapist effects when delivering guided self-help after controlling for pre-treatment severity, (2) the role of ego strength, resilience

and intuition in relation to patient outcomes and (3) the factors that contribute to better patient outcomes. To meet the third aim we employed qualitative methods to contextualise the MLM results.

Method

Design

The study adopted a cross-sectional design comprising a volunteer sample of PWPs who completed their training in 2010 and subsequently worked within six IAPT services located across the North of England, UK. Four sources of information were collated and analysed, (1) anonymised electronic download data of patient outcomes routinely collected within their IAPT service, (2) PWP self-rated questionnaires of intuition, ego strength and resilience, (3) interview data with PWPs focusing on style of work engagement and (4) supervisor-rated questionnaire (intuition) and interview. The fourth data source was to gain a supervisor perspective on individual PWP's clinical and organisational practices. The design utilised a triangulated framework of effectiveness (patients, PWPs and supervisors) contributing to a mixed methods approach to addressing the hypotheses.

Participants

Psychological Wellbeing Practitioners (PWPs) and Supervisors

Universities across the North of England who provided PWP training courses (N=3) were approached in order to identify potential PWP participants and their employing services. N=15 services were approached and invited to participate and 9 agreed. Across these 9 services, all eligible PWPs were invited to participate (N=47) and N=31 agreed. Subsequently, 3/9 services were unable to provide client outcome data, due to technical difficulties with data retrieval and N=2 PWPs withdrew. The final research sample therefore comprised N=21 PWPs (5 males and 16 females) employed in 6 IAPT services variously provided by NHS trusts (N=4), 3rd sector (N=1) or voluntary (N=1) organisations. As all

PWPs in the research sample had completed training in 2010, levels of clinical experience and subsequent time in the PWP role were consistent.

Participants had a mean of 3.5 years of previous experience of working in mental health settings, with a range of 0-17 years. Previous settings varied from community, inpatient and forensic settings across the roles of Support Worker, Mental Health Nurse, Assistant Psychologist, Occupational Therapist (OT) and voluntary positions. Almost half of participants had studied undergraduate psychology (9/20 respondents), N=2 had previous core mental health professional training (nurse or OT); N=4 had studied aspects of counselling ranging from short introductory courses to diplomas and N=3 had previous brief (less than one week) training in CBT.

Participants had a mean age of 29.91 years (SD = 7.6 years, range 23 – 52 years) and treated a mean of 53.55 patients in the study period, from when they started in their service to the end of February 2011, ranging from 8-197 patients. The mean age of patients on each PWP's caseload ranged from 36 to 46 years. After two reminders, N = 17 (81%) supervisors participated in an interview about their supervisees approach to work and completed a questionnaire relating to that PWP.

Patients

Routinely collected, anonymised patient data was obtained from electronic downloads from the participating IAPT services, with outcome data comprising closed cases and in-treatment cases. Outcomes were included in the research data set when, (1) patients had attended at least two sessions which included a pre-treatment assessment, (2) patients had completed the IAPT minimum dataset at the first and last session attended and (3) treatment was delivered in a one-to-one format.

Complete datasets were obtained for N = 1,122 patients with a mean age of 41 years (SD = 14.23 years, range = 16-92 years). Females comprised 64.7% of the sample. In terms

of ethnicity, 65.8% identified themselves as Caucasian, 2.8% as Asian, 0.7% as Black Caribbean or African and 1.1% as mixed race. Ethnicity information was not available for 29.3% of the sample. In terms of treatment duration, patients received a mean of 5 sessions (SD = 2.88 sessions; range 2-21). The mean (SD) number of patients per PWP was 53.30 (52.10) with a range of 8-197.

Measures

PWPs, supervisors and patients completed differing batteries of measures. These are outlined below.

PWP Measures and Interview

Ego strength: The Psychosocial Inventory of Ego Strengths measures ego strength (PIES; Markstrom, Sabino, Turner & Bergman, 1997). The PIES comprises 64 items summed to give a total ego strength score. The PIES has been shown to have good internal consistency ($\alpha = 0.94$; Markstrom et al., 1997) and good construct validity (Markstrom & Marshall, 2007). Example PIES item: *“I have strengths that enable me to be effective in certain situations”*.

Intuition: The Rational-Experiential Inventory measured intuition (REI; Pacini & Epstein, 1999). The REI assesses an individual's preference for either rational or experiential cognition (20-item scales each). The two REI scales have good internal consistency (rationality $\alpha = 0.90$, experientiality $\alpha = 0.87$; Pacini & Epstein, 1999) and test-retest reliability (rationality $r = 0.76$, experientiality $r = 0.83$; Handley, Newstead, & Wright, 2000). Example rationality item: *“I have a logical mind”* and example experiential item: *“I believe in trusting my hunches”*.

Resilience: The Connor-Davidson Resilience Scale measured resilience (CD-RISC; Connor & Davidson, 2003). This 25-item measure is summed to provide a total resilience

score. The CD-RSIC has good internal consistency ($\alpha = 0.89$; Connor & Davidson, 2003) and test-retest reliability ($r = 0.87$; Connor & Davidson, 2003). Example item: “Under pressure, I stay focused and think clearly”.

Interview Schedule: The PWP interview schedule was developed based on the Jennings and Skovholt (1999) qualitative study of traditional ‘high contact low volume’ therapists. Four of the questions from the Jennings and Skovholt (1999) schedule were included in the initial schedule for this study, such as “*what is particularly therapeutic about you?*” These were adapted and revised based on feedback from pilot interviews (N=3) and on the specific requirements of the PWP role.

Supervisor Measure and Interview:

Intuition: Supervisors completed the Rational-Experiential Inventory (REI; Pacini & Epstein, 1999). As the focus was on their named PWP supervisee rather than them as supervisors, the questions were re-worded and framed in the 3rd person (i.e., “*the supervisee has a logical mind*”). As the REI was designed for a self-report, reliability and validity data cannot be directly transferred or assumed.

Interview Schedule: The interview schedule for supervisors was again developed on the Jennings and Skovholt (1999) format, but this time adapted for the specifics of a supervisory role. Four of the original interview questions were included in the original supervisor schedule such as “*what distinguishes a good therapist from a great therapist?*” (nb therapist changed to PWP). The questions were adapted and revised based on feedback from a pilot interview, to specifically tap into supervisor perceptions of PWP performance.

Patient Measures:

Patient Health Questionnaire-9 (PHQ-9). The PHQ-9 identifies 'cases' at screening and measures intensity of depression. Caseness is defined as scoring above the cut-off score of 10, which indicates presence of depression (Kroenke, Spitzer, & Williams, 2001). The PHQ-9 has high sensitivity (92%) and specificity (80%) when using the cut-off score (Gilbody, Richards, Brearly, & Hewitt, 2007). The measure also has good construct validity and internal reliability ($\alpha = 0.89$; Kroenke, Spitzer, & Williams, 2001).

Generalised Anxiety Disorder (GAD-7). The GAD-7 identifies 'cases' at screening and measures intensity of anxiety. A cut-off score of 8 indicates clinically relevant anxiety (Spitzer, Kroenke, Williams, & Lowe, 2006). The GAD-7 has good sensitivity (98%) and specificity (82%) (Gilbody et al., 2007) and good construct validity, internal consistency ($\alpha = 0.92$) and test-retest reliability ($r = 0.83$; Spitzer et al., 2006).

Procedures

Quantitative Data and Blinding Procedures

To prevent bias arising from knowledge of the effectiveness of individual PWPs, blinding procedures were employed when requesting the outcome data from services and prior to qualitative interviews. Outcome data was cleaned and stored by a third party (DS, who was not involved in the interviews) and the lead researcher did not access the PWP or supervisor-completed measures prior to conducting interviews. An additional layer of blinding was added by ensuring that PWPs were anonymised in their datasets by data managers, with anonymity checked by a third party (DS). This enabled the multi-level modelling analysis to be completed without identifying any PWPs and so minimised any potential bias in the analysis and ranking of PWP effectiveness.

Data Analyses

Multi-level modelling

Analysis was conducted using multi-level modelling (MLM) software (MLwiN v2.3; Rasbash, Charlton, Browne, Healy, & Cameron, 2009). Two separate models were developed (PHQ-9 and GAD-7) to examine differences in PWP outcomes for depression and anxiety. Pre-treatment scores and interactions between the measures were controlled for by the inclusion of both PHQ-9 and GAD-7 pre-treatment scores, centred around their grand means in both models (Hoffmann & Gavin, 1998; Wampold & Brown, 2005). Variables in the model were considered statistically significant when their coefficients were more than 1.96 times their standard errors. Models were developed in stages using Iterative Generalised Least Squares (IGLS) procedures, beginning with a single level regression model, where the impact of the PWP was 'fixed' and the regression line and outcome intercept was considered to be the same for all PWPs. By progressing to a random intercept, multilevel model with patients at level 1 and PWPs at level 2, the regression lines and intercepts could vary for each PWP but remain parallel.

The final stage was a random slope model, where the relationship between pre-treatment scores and outcome was also allowed to vary between PWPs. At each development stage, improvements in the model were tested for significance by comparing the derived $-2 \times \log$ likelihoods against the chi-squared distribution for the additional degrees of freedom. From the final models, the proportion of the total unexplained outcome variance that was at the PWP level was taken as the therapist effect. Because IGLS procedures tend to slightly underestimate effects and the sample of PWPs was not particularly large, Markov Chain Monte Carlo (MCMC) procedures were utilised. A simulation chain of 25,000 iterations was found to be adequate to stabilise parameter estimates from which the median therapist effect

was derived, along with a 95% 'probability interval' (PrI) taken as the 0.025 and 0.975 percentile values in the chain (Brown , 2009).

The intercept residuals for each PWP, produced by the model represent the error terms by which each PWP's regression line deviates from that of the average PWP, after controlling for client pre-treatment scores. Therefore, residuals may be used to rank and make comparisons between PWPs (Goldstein & Spiegelhalter, 1996; Rasbash, Steele et al., 2009). The MLM analysis was therefore used in three ways, (1) to examine the amount of variance in the outcomes attributable to PWPs, controlling for pre-treatment scores to assess whether the therapist effect was significant (2) to examine the shape of residual plots of PWP variation and (3) to utilise this shape in determining quartiles of PWPs based on the rank of their residual plots.

Patient outcomes

In addition to pre-post effect size, reliable improvement and reliable deterioration on PHQ-9 and GAD-7 were calculated using the Reliable Change Index (Jacobson & Truax, 1991). Patients were considered to have made reliable improvement or deterioration, if their scores had increased or decreased at post-treatment by at least 6 points on the PHQ-9 or by at least 4 points on the GAD-7 (Parry et al., 2011). Therefore patients scoring above the clinical cut-off pre-treatment were deemed to have made a 'reliable and clinically significant' improvement during PWP treatment if, (a) they made reliable improvement and (b) their post-treatment score was below clinical cut-off. Overall sample rates were calculated as well as the rates for individual PWPs.

Analysis of more and less effective PWP clusters

Using the ranked PWP residuals derived from the model, PWPs were grouped into upper and lower effectiveness quartiles enabling quantitative and qualitative analyses of differences. Mann Whitney U tests were used to test for differences on PWP and supervisor

rated measures and uncontrolled effect sizes using Cohen's *d* and recovery rates were calculated for the upper and lower quartile groups. Qualitative analyses of the PWP and supervisor interviews were conducted using Template Analysis (TA; King, 1998). In TA, a priori codes are defined which are expected in the data - the template - but modified and supplemented by additional codes as the analysis progresses. This approach is widely used in health research (e.g., King, Thomas, & Bell, 2003) because it is a flexible approach, easily modified for different and specific areas of study.

A key feature of TA is hierarchical coding, with higher order codes overarching a cluster of lower order codes in a similar theme (King, 2004). A priori, high order themes were based on the interview schedules that were then analysed in two stages, (1) initial exploration of the data of all participants and (2) examination of common lower order themes between upper and lower PWP effectiveness quartiles. High order themes were used as a guide to examine emerging lower order themes across all PWPs and supervisors in the first stage of analysis, consistent with the "listing codes" procedure outlined by King (2004). As a quality check, an independent researcher (a doctoral level student familiar with TA) analysed 6 interviews (15%), $N=2$ taken from each of the PWP effectiveness groups (i.e. upper, lower and middle quartiles). They independently coded lower order themes from the template of high order themes, which were then compared with the lead researcher's codings. There was a 78% agreement rate for coding of the lower order themes.

In the second stage, some higher order themes were deleted or redefined according to the TA procedure (King, 2004). Themes were deleted if less than $N=2$ PWPs or supervisors had described a similar lower order theme of relevance to the high order theme. Lower order themes identified by more than $N=2$ PWPs or supervisors in their respective group (i.e., upper or lower quartile) were included as a final lower order theme. Quality control procedures were then again implemented with another independent rater (doctoral level

student familiar with qualitative analysis). They examined the high and lower order themes for the upper and lower quartile effectiveness groups to determine whether agreement was met regarding their appropriateness and fit. The second stage of quality control resulted in two changes to the labelling of lower order themes, but did not change the content of any of the themes.

Results

Results are presented in three specific phases, (1) analyses of outcomes at patient and PWP level, (2) testing for therapist effects using MLM to yield effectiveness clusters for PWPs and (3) quantitative and qualitative analyses of derived upper and lower PWP effectiveness quartiles.

Outcomes

Patient level outcomes

Table 1 presents the mean patient-level outcomes for the PHQ-9 and GAD-7 (N=1,122 patients). The mean (SD) change scores (i.e. from pre-treatment to final session) on the PHQ-9 and GAD-7 were 3.34 (6.43) and 3.05 (5.82) respectively, with corresponding pre-post effect sizes of 0.52 for depression and 0.55 for anxiety. Reliable change in depression rates were 31.6% of patients made a reliable improvement and 3.6% reliably deteriorated. For anxiety, 41.8% of patients made a reliable improvement and 4.2% had a reliable deterioration. Adopting the more stringent criteria of reliable and clinically significant change, for those patients meeting pre-treatment caseness criterion on PHQ-9 (N=775, 69.1%), 34.1% made a reliable and clinically significant improvement. Of the N=860 (76.6%) who scored above the caseness cut-off at intake on the GAD-7, 36.5% made a reliable and clinically significant improvement.

Insert Table 1 about here

PWP level outcomes

Table 2 presents PWP-level outcome indices. This reports the median and range of effect sizes, reliable improvement and deterioration rates and reliable and clinically significant improvement rates on the PHQ-9 and GAD-7. The pre-treatment caseness rates on PHQ-9 ranged from 25.0% to 93.3%, with a median rate of 68.1%. For GAD-7 caseness rates, the range was the same with a median of 75.0%.

Insert Table 2 about here

Multi-level modelling

The initial stage of multi-level modelling involved using the total dataset of in-treatment and completed treatment cases, to produce two single level regression models for depression (PHQ-9) and anxiety (GAD-7) separately, including and controlling for both pre-treatment scores and the interaction between them. A multi-level model was then developed with PWPs at level-2 and allowing individual PWPs regression lines and their intercepts to vary, whilst keeping a common slope. Using the likelihoods ratio test to estimate the between PWP variation in the intercepts and comparing to the single level model, showed that the difference was significant for both depression and anxiety: PHQ-9, $\chi^2(1) = 142.023$, $p < .001$; GAD-7 $\chi^2(1) = 140.488$, $p < .001$. Results indicate that the random intercept model was a better fit for the data than the single level regression, suggesting significant variability between PWPs, even after adjusting for pre-treatment severity.

The random slopes model considered how the relationship between pre-treatment score and outcome score varied between PWPs. Model 1, for depression outcomes (see below) showed that the intercepts of the individual PWP lines varied, with a mean of 8.664 (SE 0.437) and a variance of 2.779 (SE 1.131). The coefficient of the PHQ-9 average slope was estimated at 0.549 (SE 0.046) and individual PWP slopes varied about this mean with an estimated variance of 0.013 (SE 0.009). The loglikelihood test for Model 1 (see Appendix)

was significant ($\chi^2(2) = 13.725$, $p < .01$), thereby indicating an improvement from the random intercept model.

Model 2 (see Appendix) for anxiety outcomes shows that the intercepts of the individual PWP regression lines were varied, with a mean 7.805 (SE 0.393) and variance of 2.262 (SE 0.917). The coefficient of the average slope was estimated at 0.478 (SE 0.048) and individual PWP slopes varied about this mean, with an estimated variance of 0.015 (SE 0.010). The loglikelihood test indicated the random slope model was significantly better than the random intercept model ($\chi^2(2) = 17.283$, $p < .001$).

Therapist effects

The therapist effect for PHQ-9 outcomes (see Model 1) was 8.7% and for GAD-7 (see Model 2) the effect was 8.8%. This indicates that approaching 9% of the variability in patients' outcomes for both depression and anxiety was due to variability between PWPs - after controlling for pre-treatment severity. MCMC procedures indicated a median therapist effect of 9.7% for PHQ-9 with a 95% PrI of 5.8%–17.4%, while for the GAD-7, the effect was 9.8% (95% PrI: 5.8%–17.6%).

Ranking PWP effectiveness using model residuals

In order to make comparisons between the outcomes of PWPs, residuals of individual practitioners were used. Residuals represent how each PWP departs from the overall outcome mean. These residuals can be used to rank PWPs and plot the shape of the variation in outcome between practitioners. Figure 1 presents the residuals (with 95% CIs) plotted for PHQ-9 and GAD-7 with PWPs (ranked from most to least effective) shown across the x-axis.

Insert Figure 1 here

PWP rankings between patient outcome measures correlated strongly ($r = 0.96$, $p < 0.001$), indicating that practitioners more effective at treating depression were also more

effective at treating anxiety. The plots show that confidence intervals were large, with only four PWPs having 95% CIs that did not cross zero on PHQ-9 and three on GAD-7. The remaining PWPs' residual values all had confidence intervals that crossed zero, indicating that they were not significantly different from the average PWP.

Because of the small number of PWPs that were significantly different from the average PWP and in order to compare clinical outcomes on patient outcomes with data from the PWP measures and interviews, the five PWPs ranked 1-5 and the five ranked 17–21 were used to make comparisons. Those ranked 17–21 were the same for both PHQ-9 and GAD-7. However, for those ranked 1-5 only four were the same. In order to determine the fifth ranked PWP, composite rankings from both measures were used. PWP id 6 who was ranked 5 on PHQ-9 and 8 on GAD-7 had an average rank of 6.5, while PWP id 17 was ranked 7 on PHQ-9 and 4 on GAD-7, giving an average rank of 5.5. Therefore the latter was selected as the fifth, high-ranked PWP. The composite rankings correlated significantly with reliable and clinically significant improvement rate rankings for PHQ-9 ($r = 0.86, p < 0.001$) and GAD-7 ($r = 0.89, p < 0.001$). Service codes indicated that both the high and low-ranked clusters contained PWPs from a range of services.

Comparisons of upper and lower PWP effectiveness quartiles

Quantitative approach

The pre-post effect size for the least effective PWPs was 0.20 on the PHQ-9 and 0.22 on the GAD-7, while for the most effective cluster the effect sizes were 0.92 and 0.95 respectively. Using ANCOVAs to control for pre-treatment scores, the outcomes for patients treated by upper quartile PWPs were significantly better for both depression and anxiety (PHQ-9, $F(1) = 133.0, p < .001$; GAD-7, $F(1) = 125.0, p < .001$). Significantly more patients treated by the most effective PWPs (46.4%) compared to less effective PWPs (19.0%) made a

reliable improvement in depression ($\chi^2_{(1)} = 68.5, p < 0.001$). Significantly more patients treated by the most effective PWPs (58.3%) compared to less effective PWPs (25.7%) made a reliable improvement in anxiety ($\chi^2_{(1)} = 84.1, p < 0.001$) during treatment. The most effective PWPs also had fewer patients who reliably deteriorated – 0.7% on PHQ-9 and 2.5% on GAD-7 – compared with 6.2% on both measures for the least effective PWPs.

Whilst there was no significant difference between the two PWP clusters in the proportion of patients who met caseness prior to treatment, there were significant differences in rates of patients achieving a reliable and clinically significant change. Top ranked PWPs had a rate of 51.9% for depression compared with a rate of 21.1% for the bottom ranked PWPs ($\chi^2_{(1)} = 56.17, p < 0.001$). For anxiety outcomes, the respective rates were 57.8% and 23.0% ($\chi^2_{(1)} = 74.52, p < 0.001$). In terms of the PWP completed measures, resilience was significantly higher in the high ranked PWP cluster ($U = 2.000, N_1 = 5, N_2 = 5, p = .03$). Clinical supervisors for PWPs in the low ranked cluster rated their PWPs as using significantly more experiential intuition during their decision making ($U = .500, N_1 = 5, N_2 = 5, p = .02$). Differences between the clusters on ego strength were not significant ($p > .05$).

Qualitative approach

The qualitative results presented derive from the second stage of analysis of high and lower order themes in the upper and lower PWP effectiveness quartiles and are mapped out in Figure 3. Full templates of lower order themes are available from the first author. High order themes are presented on the left of the figures, with lower order themes of participants in the upper and lower quartiles on the right. Although PWPs in the upper and lower quartiles did report some of the same themes, only the unique lower order themes are presented and discussed further. Two high order themes were deleted due to lack of subthemes emerging amongst practitioners in the two groups: “how previous experience

hindered” (PWP question only) and “how CPD has influenced PWP practice” (supervisor and PWP question). The first three high order themes relate to questions asked to both PWPs and supervisors and the three high order themes in italics relate to questions only asked to PWPs, therefore no lower order themes for supervisors could have emerged. To ensure anonymity the following quotes have all been changed to reflect female respondents, even when the PWP was male.

Insert Figure 2 here

More effective practitioners

Figure 2 shows a total of 11 unique lower order themes reported by PWPs in the upper effectiveness quartile, whilst supervisors reported 5 different unique lower order themes. In relation to the three high order themes relating to both PWPs and supervisors in the upper quartile, this group of PWPs and their supervisors reported that effective PWPs employed proactivity to develop their skills by (1) using online research, (2) using observation of others in clinical practice and (3) being an active participant in supervision. This proactive stance was also reflected in PWPs’ reports of their accessing of supervisors’ skills and knowledge and engaging in supervision at a deeper experiential level:

“With clinical supervision I try to get the most out of it through thinking of different ways we can use it, like by having case discussions, case presentations....role plays and things” (PWP A:5)

Effective PWPs typically ensured that they were prepared and organised for supervision:

“I usually [pull together information on] who I’m currently working with, recently assessed, recently discharged [other information including outcome scores] and

present a copy of that every case management supervision to my supervisor and highlight which *ones I'd like to discuss*" (PWP B:6)

The good organisational skills of effective PWPs were reciprocally reported by supervisors, who noted that such PWPs openly discussed clinical difficulties:

"She's very happy to bring along examples of things that are going well, things that are going less well" (Supervisor A:4)

Some of the lower order themes for effective PWPs indicated skilfulness in use of the PWP clinical method. For example, effective PWPs reported being thorough in their approach with clients, ensuring clarification in their communication:

"I do make efforts to be explicit with clients about exactly why it is that I'm talking to them about doing certain things, the rationale for it and how it's going to help them" (PWP D:7)

This skilfulness was also reflected in the PWPs description of their ability to adapt interventions to fit individual patient needs, whilst not drifting away from treatment protocols. The three high order themes relating to only PWPs and not supervisors indicated that PWPs in the upper quartile had a good understanding of the IAPT model and PWP role, previous knowledge of CBT and felt certain of when to step patients up to high intensity CBT:

"I'm aware that there are things that I'm not sure about, but I think they're more high intensity work that needs doing" (PWP D:1)

Less effective practitioners

Figure 4 shows that five unique lower order themes emerged from PWPs in the lower effectiveness quartile and three unique lower order themes emerged from their supervisors. Lower effectiveness PWPs lower order themes reflected less confidence in the PWP clinical method and the need for further development in specific PWP treatment skills:

“There’s a lot of emphasis on behavioural activation, but I don’t feel like I have very good skills in delivering that” (PWP W:8)

The approach to supervision also reflected less confidence in using supervision to ask specific clinical questions:

“I’ve got loads of guidance on who I shouldn’t be seeing and who needs stepping up, and about things about the disorders we didn’t look at University”

(PWP Y:9)

PWPs in the lower quartile reported that the main way to be effective in the stepped care model was through communication.

“I make my best efforts to introduce myself to the GP so that I’m not just a name, so that they know I’m a presence and that I’m there to support their patients and try and open communication a little bit” (PWP Z:10)

This account differs from PWPs in the upper quartile, as it does not reflect any specific factor associated with IAPT or the PWP role. In quote Z:10 the focus is on communication and not on communication specifically about step two interventions or about the IAPT model generally. Openness was a quality that supervisors of PWPs in the lower quartile reported across two higher order themes, for example:

“If she’s not sure of something she will ask” (PWP V:11)

However, this differs from openness reported by supervisors of upper quartile PWPs, as it does not reflect openness to difficulties.

Discussion

The current research had two aims: (1) to test for the presence of therapist effects using multilevel modelling in a sample of qualified PWPs delivering protocol-driven low intensity interventions in routine practice and (2) to establish the clinical and organisational skills that differentiate more and less effective PWPs. Results indicated that therapist effects

accounted for approximately 9% of the variance in patient outcomes in step 2 IAPT services - when utilising the most appropriate statistical analysis for such nested data and controlling for pre-treatment scores. As the research dataset contained both completed and in-treatment cases this accessed the most contemporary index of PWPs effectiveness and so was considered a better estimate of therapist effects. PWPs do not operate specialist services in which patients are assigned to them, beyond meeting service requirements of mild to moderate anxiety/ depression. The variation evidenced between PWPs is therefore more likely to reflect differences in individual PWP practice, rather than an expression of organisational or systemic influences.

The rate of 9% is comparable to previous studies of therapist effects during routine delivery of the traditional 'low volume, high contact' psychotherapies, where the percentage of variance attributable to therapists ranges from 5–8% (Lutz et al., 2007; Wampold & Brown, 2005). Even the lower bound of the 95% probability interval (5.8%) falls within this range. Clinical outcomes were broadly representative of PWP work (e.g., Glover, Webb, & Evison, 2010). However, in the current study more effective PWPs consistently yielded greater change in their anxious and depressed patients with higher positive change rates and lower patient deterioration rates. Using Cohen's (1990) power primer less effective PWPs produced a small effect size in their patients, whilst more effective PWPs produced a large effect size. More effective PWPs were more resilient, more confident in their skills and had a more proactive/organised/thorough approach to the delivery of the PWP clinical method. The finding of a therapist effect is interesting as the guided self-help interventions at step 2 in IAPT are protocol-driven, with original training and subsequent supervision actively discouraging therapeutic drift (Waller, 2009). The findings challenge the notion that protocol-driven therapy remains uncontaminated and unadulterated by the skills of the practitioner providing the intervention. It is a significant strength of the current report that

MLM analyses were yoked to therapist characteristics (and qualitative results), as that is not the norm in therapist effects research.

Whilst NICE guidelines recommend specific types of treatments for particular diagnoses, the present research prompts the need to reconsider the role of the therapist in the delivery of evidenced-based psychological therapies. The findings suggest that it is not only what self-help intervention patients receive that is important, but also by whom. The skill appears to be melding protocols to fit the needs of the patient, so that evidence drives the treatment whilst acknowledging and respecting each patient's individuality. When traditional therapists are provided with feedback on patient progress, clinical outcomes improve (Lambert, Whipple, Vermeersch, Smart, Hawkins, Nielsen & Goates, 2002). The thorough outcome framework that supports and evaluates IAPT (CSIP Choice & Access Team, 2008) make such feedback possible. PWPs and supervisors should therefore regularly review their recovery rates in order to ensure that the low intensity treatment is actually helping (Kraus et al., 2011). Supervisors may unhelpfully collude with poor clinical performance, if the anxiety of raising the issue of evidence of a consistent lack of patient progress feels too great.

More and less effective PWPs clusters differed in terms of self-rated resilience. Inspection of these means compared with the available (limited) norms from a general population sample (Connor & Davidson, 2003) suggests that the PWPs in the higher cluster scored similarly to the population mean (i.e. they are not 'super resilient'). The resilience scores from the PWPs in the lower effectiveness quartile were however within the bottom quartile of the general population. Therefore PWPs with 'average resilience capacity' facilitate better outcomes for their patients, whilst 'resilience deficits' appears associated with less patient change. The concept of resilience relates to the ability to cope with challenges, adversity or stressors (Rutter, 1993). The 'low contact, high volume' approach at step 2 IAPT services may present a challenge, given that one of the defining features of PWP work

is the safe and effective management of high caseloads (CSIP Choice & Access Team, 2008). It is worth noting that in sampling and then comparing the characteristics of the highest and lowest effectiveness quartiles of PWPS from the caterpillar plots, this meant that the subsequent analysis then compared relatively small Ns of practitioners (i.e. 5 in each PWP cluster). Characteristic comparisons should therefore be viewed with due caution and as a prompt for further investigation. However, any further research that yokes MLM therapist effects to therapist characteristics should aim to increase the sample size of the therapists, in order to meet power considerations in any subsequent characteristic analyses.

Supervisors' accounts of more effective PWPs provided additional evidence of resilience, describing practitioners who remained open to discussing the difficulties of their low intensity work during supervision. This openness to learning was not a factor described by supervisors of the less effective PWPs. In all these instances, resilience would be best understood as a process (Zautra, Hall & Murray, 2010) rather than the expression of the trait of resiliency. Resilience in PWPs is therefore malleable and would be the result of practitioners working in the stepped-care system and developing clinically from case management and clinical supervision support. Training efforts could be directed at developing resilience in PWPs. Future therapist effects research could also include 3 and 4 level models in which practitioners are nested in clinical supervisors, who are nested in different services. Obviously, this inflates the need for much larger samples of PWPs, clinical services and supervisors.

Experiential intuition is defined as information processing that is "preconscious, rapid, automatic, holistic, primarily nonverbal and immediately associated with affect" (Pacini & Epstein, 1999, p. 972) and the use of this processing style was rated significantly higher by supervisors of practitioners in the less effective PWP cluster. This may not be the most useful approach in the PWP role given the typically high caseloads and associated

throughput. This result needs to be treated with some caution however, as the REI is not validated for “other-rated” usage. Use of supervisor ratings represent a strength of the current study adding objectivity and use of supervisor ratings is novel in therapist effects research. The use of a less affect-driven processing style was reflected in the qualitative accounts of the more effective PWPs, and reported as being more proactive, prepared and organised. More effective PWPs also reported an overall sense of being more confident in their competence through being more grounded in the IAPT model, flexibly adapting the clinical method and making more use of supervision. This depicts a virtuous circle of more organised, confident and well-versed PWPs producing better outcomes, no doubt reinforcing being organised, confident and well-versed. These results prompts further PWP therapist effects research to advance the breadth of yoking methodologies, go onto additionally sample both the characteristics and the actual ‘in-session’ behaviour of effective practitioners. This will then identify what it is that effective practitioners are doing differently and how this is actually clinically achieved. For example, the practice of single ‘super-coach’ in the present sample would be perfect to sample.

Less effective PWPs reported gaps in their skills and knowledge of the PWP clinical method (e.g. regarding the behavioural activation treatment protocol). Having a gap in declarative knowledge of behavioural activation would hamstring effectiveness, as this is a cornerstone of the PWP clinical method (Richards & Whyte, 2009). Lack of knowledge may have also been a factor in less effective PWP’s using of supervision to ask specific clinical questions, rather than engaging in the more process-type supervision reported by more effective PWPs. Heppner and Roehlke (1984) found that use of supervision in novice traditional therapists was defined by demanding direction and advice on technique. Stoltenberg and McNeill (1997) defined lower-level supervisees as relying on supervision to seek reassurance and glean specific guidance, whilst upper-level supervisees were open to

developing enhanced self-awareness via reflection. The current research suggests that in practitioners matched in their experience of PWP working, early differences in the use of clinical supervision can still occur.

In terms of methodological limitations, the possibility of study bias was introduced by 9/15 services and 18/47 PWPs declining to participate. Therefore, the research sample may have included only those services and PWPs sufficiently secure in their own practices to feel able to participate. The voluntary nature of research and associated avoidance of participation presents a common problem across practitioner studies. As therapist effects were solely measured by clinical outcome, future research would also benefit from accessing a wider indication of practitioner effectiveness, such as analysis of variation in dropout rates. The current study did not employ an index of fidelity to the PWP clinical method and the results suggest that good knowledge of the clinical method enables effective practitioners to 'flex' low intensity interventions to suit the needs of the individual patient, whilst resisting therapeutic drift (Waller, 2009). It is worth noting, however, that there is currently no published measure of PWP competency for use in routine clinical practice and supervision, in contrast to the regular use in traditional 'high intensity' CBT of the Cognitive Therapy Scale-Revised (CTS-R; Blackburn, James, Milne, Baker, Standart, Garland & Reichelt, 2001). Development of a valid and reliable PWP fidelity measure for routine practice, which is easy to use in supervision, is sorely needed to support PWPs in their clinical work. The final methodological critique is the relatively small sample size of IAPT services accessed in the present research. Gyani et al (2013) evidenced a broad recovery rate range of between 23.9 and 56.5% across IAPT services and did note that service level factors (such as the size of the service and the use of stepped-care) did influence patient outcomes. Indeed the present dataset had a 3-level structure (1: service, 2: PWP, 3: patient) and indicates that further therapist effects studies in IAPT services need to increase the sample size of services at level

1 of any MLM to be able to isolate the effect of service. Any effect of service found would then point to the need for organisational level interventions (e.g. systems reviews) to support practitioners in their work and to improve outcomes for patients.

As this study utilised a small sample of PWPs, future studies could build on the current research by utilising larger PWP samples to potentially replicate findings. Datasets from 6 PWPs could not be included, due to technical retrieval difficulties within the IAPT services in terms of their outcome data systems. Such difficulties are a challenge to the routine retrieval of data and associated feedback that is at the heart of the IAPT philosophy and values (CSIP Choice & Access Team, 2008). Subsequently, the sample utilised was smaller than the 30 practitioners recommended for use in MLM (Soldz, 2006). However, by using MCMC procedures, confidence intervals were derived to assess and report the reliability of estimates.

In conclusion, this study has found evidence of therapist effects in a practitioner group using guided self-help interventions in routine IAPT practice and adds to the growing body of evidence indicating that therapists can and do provide significant contributions to outcome. The results in combination with the resilience findings have wide implications for the selection, education and training of PWPs. The treatment of PWPs as a random factor in the MLM means that findings can be generalised, despite the reservations that have been made concerning avoidance of participation. More research is needed using larger PWP samples, use of in-session measures of fidelity and further longitudinal exploration of the concept of clinical and organisational resilience as a malleable and on-going process.

References

- 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 edition. Wiley and Sons.
- Barkham, M., Stiles, W.B., Connell, J., Twigg, E., Leach, C., Lucock, M., Mellor-Clark, J., Bower, P., King, M., Shapiro, D.A., Hardy, G.E., Greenberg, L., & Angus, L. (2008). Effects of psychological therapies in randomized trials and practice-based studies. *British Journal of Clinical Psychology*, 47, 397-415.
- Blackburn, I-M., James, I. A., Milne, D. L., Baker, C., Standart, S. H., Garland, A. & Reichelt, F. K. (2001). The Revised Cognitive Therapy Scale (CTS-R): psychometric properties. *Behavioural and Cognitive Psychotherapy*, 29, 431–446.
- Browne, W. J. (2009). MCMC estimation in MLwiN Version 2.13. Centre for Multilevel Modelling, University of Bristol.
- Brown, G.S., Lambert, M.J., Jones, E.R. & Minami, T. (2005). Identifying highly effective psychotherapists in a managed care environment. *American Journal of Managed Care*, 11, 513-520.
- Clark, D.M., Layard, R., Smithies, R., Richards, D.A., Suckling, R. & Wright, B. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy*, 47, 910-920.
- Connor, K.M. & Davidson, J.R.T. (2003). Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). *Depression and Anxiety*, 18, 6-82.
- Crits-Christoph, P., & Gallop, R. (2006). Therapist effects in the National Institute of Mental Health Treatment of Depression Collaborative Research Program and other psychotherapy studies. *Psychotherapy Research*, 16, 178-181.

- Care Services and Improvement Partnership Choice and Access Team (2008) Improving Access to Psychological Therapies (IAPT) Commissioning Toolkit. London: Department of Health.
- Cohen, J. (1990). A power primer. *Psychological Bulletin*, 112, 155-159.
- Elkin, I., Falconnier, L., Martinovich, Z., & Mahoney, C. (2006a). Therapist effects in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Psychotherapy Research*, 16, 144-160.
- Elkin, I., Shea, M.T., Watkins, J.T., Imber, S.D., Stotsky, S.M., Collins, J.F. & Parloff, M.B. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archive of General Psychiatry*, 46, 971-982.
- Forand, N.R., Evans, S., Haglin, D., & Fishman, B. (2011). Cognitive Behavioral Therapy in practice: Treatment delivered by trainees at an outpatient clinic is clinically effective.
- Gibbons, C.J., Fournier, J.C., Stirman, S.W., DeRubeis, R.J., Crits-Christoph, P., & Beck, A.T. (2010). The clinical effectiveness of cognitive therapy for depression an outpatient clinic. *Journal of Affective Disorders*, 125, 169-176
- Gilbody, S., Richards, D., Brearley, 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, 1596-1602.
- Glover, G., Webb, M., & Evison, F. (2010). Improving Access to Psychological Therapies: A review of the progress made by sites in the first roll-out year. North East Public Health Observatory.

- Goldstein, H. & Spiegelhalter, D. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance-with discussion. *Journal of the Royal Statistical Society*, 159, 385-443.
- Gyani, A., Shafron, R., Layard, R., & Clark, D.M. (2013). Enhancing recovery rates: Lessons from year one of IAPT. *Behaviour Research and Therapy*, 51, 597-606.
- Handley, S.J., Newstead, S.E., & Wright, H. (2000). Rational and experiential thinking: A study of the REI in *International Perspectives on Individual Differences Volume 1 Cognitive styles*. R.J. Riding & S.G. Rayner. (Eds) Ablex Publishing Corporation, USA Stamford.
- Hardin, J. & Hilbe, J. (2003). *Generalised estimating equations*. London: Chapman & Hall.
- Heppner, P.P., & Roehike, H.J. (1984). Differences among supervisees at different levels of training: implications for the development of a model for supervision. *Journal of Counselling Psychology*, 31, 76-90.
- Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, 23, 723-744.
doi:10.1177/014920639802400504
- Hubble, M.A., Duncan, B.L., & Miller, S.D. (1999). *The heart and soul of change: Delivering what works in therapy*. Washington DC: American Psychological Association.
- Huppert, J.D., Bufka, L.F., Barlow, D.H., Gorman, J.M., Shear, K.M, & Woods, S.W. (2001). Therapists, therapist variables, and Cognitive-Behavioral Therapy Outcome in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology*, 69, 747-755.
- Improving Access to Psychological Therapies (2008). *Implementation plan: Curriculum for low-intensity therapies workers*. London: Department of Health.

- Improving Access to Psychological Therapies (2012). Psychological well-being practitioners: Best practice guide. London: Department of Health.
- Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 2-19.
- Jennings, L. & Skovholt, T.M. (1999). The cognitive, emotional and relational characteristics of master therapists. *Journal of Counseling Psychology*, 46, 3-11
- Kim, D., Wampold, B.E., & Bolt, D.M. (2006). Therapist effects in psychotherapy: A random-effects modelling of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *Psychotherapy Research*, 16, 161-172.
- King, N. (1998). Template analysis. In G.Symon & C. Cassel, C. (Eds). *Qualitative methods and analysis in organisational research: A practical guide*. London: Sage Publications.
- King, N. (2004). Using templates in the thematic analysis of text. In C. Cassell & G. Symon (Eds). *Essential guide to qualitative methods in organizational research*. London: Sage Publications.
- King, N., Thomas, K., Bell, D., & Bowes, N. (2003). Evaluation of the Calderdale and Kirklees out of hours protocol for palliative care: Final report.
- Kraus, D.R., Castonguay, L., Boswell, J.F., Nordberg, S.S., & Hayes, J.A. (2011). Therapist effectiveness: Implications for accountability and patient care. *Psychotherapy Research*, 21, 267-276.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16, 606-613.
- Lambert, M.J., Whipple, J.L., Vermeersch, D.A., Smart, D.W., Hawkins, E.J., Nielsen, S.L., & Goates, M. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. *Clinical Psychology and Psychotherapy*, 9, 91-103.

- Layard, R., Bell, S., Clark, D., Knapp, M., Meacher, M., Priebe, S., Thornicroft, G., Turnbull, A. & Wright, B. (2006). The depression report: A new deal for depression and anxiety disorders. Centre for Economic Performance's Mental Health Policy Group.
- Luborsky, L.L., McLellan, A.T., Woody, G.E., O'Brien, C.P. & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry*, 42, 602-611.
- Lutz, W., Scott, L., Martinovich, Z., Lyons, J.S., & Stiles, W. B. (2007). Therapist effects in outpatient psychotherapy: A three-level growth curve approach. *Journal of Counseling Psychology*, 54, 32-39.
- Markstrom, C.A., & Marshall, S.K. (2007). The psychosocial inventory of ego strengths: Examination of theory and psychometric properties. *Journal of Adolescence*, 30, 63-79.
- Markstrom, C.A., Sabino, V.M., Turner, B.J. & Berman, R.C. (1997) The Psychosocial Inventory of Ego Strengths: Development and validation of a new Eriksonian measure. *Journal of Youth and Adolescence*, 26, 705-732.
- Najavits, L.M., & Strupp, H. H. (1994). Differences in the effectiveness of psychodynamic therapists: A process-outcome study. *Psychotherapy*, 31, 114-123.
- Okiishi, J.C., Lambert, M.J., Nielsen, S.L., & Ogles, B.M. (2003). Waiting for supershrink: Empirical analysis of therapist effects. *Clinical Psychology and Psychotherapy*, 10, 361-373.
- Okiishi, J. C., Lambert, M. J., Eggett, D., Nielson, S. L., Vermeersch, D. A., & Dayton, D. D. (2006). An analysis of therapist treatment effects: Toward providing feedback to individual therapists on their patients' psychotherapy outcome. *Journal of Clinical Psychology*, 62, 1157-1172.
- Pacini, R. & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality basic beliefs and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, 76, 972-987.

- Parry, G., Barkham, M., Brazier, J., Dent-Brown, K., Hardy, G., Kendrick, T., Rick, J., Chambers, E., Chan, T., Connell, J., Hutten, R., de Lusignan, S., Mukuria, C., Saxon, D., Bower, P. & Lovell, K. (2011). An evaluation of a new service model: Improving Access to Psychological Therapies demonstration sites 2006-2009. Final report. NIHR Service Delivery and Organisation programme.
- Rabash, J., Charlton, C., Browne, W.J., Healy, M., & Cameron, B. (2009). MLwiN Version 2.11. Centre for Multilevel Modelling. University of Bristol.
- Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2009). A User's Guide to MLwiN, v2.10. Centre for Multilevel Modelling, University of Bristol.
- Raudenbush, S.W., & Bryk, A.S. (2002). Hierarchical linear models: Applications and data analysis methods. 2nd Edition. Newbury Park, CA: Sage.
- Richards, D. & Whyte, M. (2009). Reach Out: National programme student materials to support the delivery of training for Psychological Wellbeing Practitioners delivering low intensity interventions. 2nd Edition. Rethink, UK.
- Robinson, S., Kellett, S., King, I. & Keating, V. (in press). Role transition from mental health nurse to IAPT high intensity psychological therapist. *Cognitive and Behavioural Psychotherapy*.
- Rutter, M. (1993). Resilience: Some conceptual considerations. *Journal of Adolescent Health*, 14, 626-631.
- Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: Therapist effects and the contribution of patient severity and risk. *Journal of Consulting and Clinical Psychology*, 80, 535-546.
- Soldz, S. (2006). Models and meanings: Therapist effects and the stories we tell. *Psychotherapy Research*, 16, 173-177.

- Spitzer, R. L., Kroenke, K., Williams, J. B., & Lowe, B. (2006). A brief measure for assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine*, 166, 1092-1097.
- Stoltenberg, C.D., & McNeill, B.W. (1997). Clinical supervision from a developmental perspective: Research and practice. In C.E. Watkins Jr. (Ed.) *Handbook of psychotherapy supervision*. Pp. 184-202. New York: Wiley.
- Thompson, D., Cachelin, F., Striegel-Moore, R.H., Barton, B., Shea, M., & Wilson, G.T. (2012). How many therapists? Practical guidance on investigating therapist effects in randomized controlled trials for eating disorders. *International Journal of Eating Disorders*, 45, 670-676.
- Turpin, G. (Ed.) (2010): *IAPT Good Practice Guide to using Self-help Materials*. NMH DU/IAPT, 1-40.
- Waller, G. (2009). Evidence-based treatment and therapist drift. *Behaviour Research and Therapy*, 47, 119-127.
- Wampold, B.E., & Brown, G.S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, 73, 914-923.
- 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.
- Zautra, A.J., Hall, J.S., & Murray, K.E. (2010). Resilience: A new definition of health for people and communities. In J.W. Reich, A.J. Zautra, & J.S. Hall (eds.). *Handbook of adult resilience*. New York: Guilford

Table 1: Patient level outcomes scores at pre- and post-treatment, change scores, and pre-post treatment effect sizes

Outcome measure	Mean score (SD)			Pre-post therapy effect size
	Pre-treatment	Post-treatment	Change score	
PHQ-9	13.17 (6.43)	9.83 (7.15)	3.34 (6.43)	0.52
GAD-7	12.04 (5.57)	8.99 (6.32)	3.05 (5.82)	0.55

Table 2: PWP level outcomes for PHQ-9 and GAD-7

Outcome criterion	PHQ-9	GAD-7
	Median (Min - Max)	Median (Min - Max)
Pre-post effect size	0.74 (-0.13–1.27)	0.76 (-0.16 – 1.26)
Reliable improvement (%)	40.5 (0–52.9)	53.3 (0 – 66.7)
Reliable deterioration (%) ¹	0 (0–14.7)	1.70 (0 – 12.2)
Reliable and clinically significant improvement (recovery) (%)	35.7 (0 – 55.6)	43.2 (0 – 85.7)

¹13 PWPs had no reliable deteriorations on PHQ-9, while 8 had no deteriorations on GAD-7

Table 3: PWP upper and lower cluster means (SD) for Resilience, Ego Strength, and Intuition

Quartile	CD-RISC Resilience	PIES		REI Intuition		REI Intuition	
		M (SD)	M (SD)	PWP rated	Supervisor rated	Rational	Experiential
Upper	90	82.00 (8.86)	274.80 (9.99)	64.00 (2.35)	58.40 (5.18)	62.40 (2.88)	50.80 (4.82)
Lower	73	70.20 (4.44)	268.00 (9.62)	66.00 (2.00)	57.20 (3.11)	61.25 (2.22)	59.75 (4.99)

¹ Published norms for US adult general population (N=577) are M = 80.40, (SD = 12.8);

Median = 82 (Connor & Davidson, 2003).

Figure 1: PWP residuals (with 95% CIs) for PHQ-9 and GAD-7

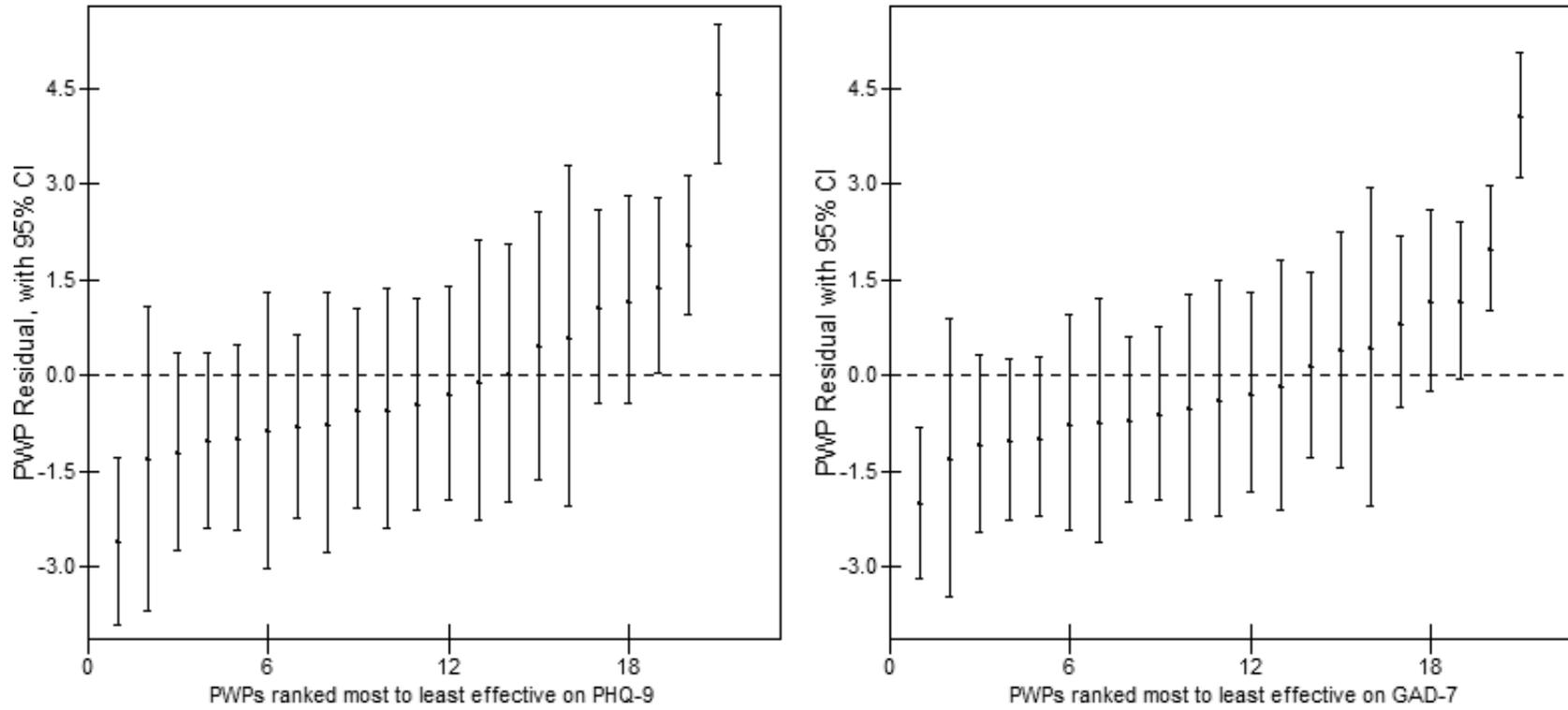


Figure 2: High order and lower order sub-themes for PWP top and bottom effectiveness quartiles

High order themes	Lower order theme: Upper quartile		Lower order theme: Lower quartile	
	PWP	Supervisor	PWP	Supervisor
Engagement with supervision to improve skills	Being prepared and organised Utilising supervisor's knowledge Process supervision	Open to discussing difficulties Active supervision participant	Specific clinical questions	Openness
Hallmarks of clinical practice	Communication skills Adapting interventions to the individual	Organisational skills		Interpersonal skills Openness
Methods of improving practice	Observing others	Proactive in improving practice Online research		
<i>How previous experience helped</i>	Knowledge of CBT principles		Developed interpersonal skills	
<i>Gaps in skills or knowledge</i>	Knowledge of medication Gaps go beyond low intensity		Specific skills Knowledge of specific interventions	
<i>Working in the stepped care model</i>	Understanding IAPT model and PWP role Knowledge of when to step up		Communication	

Appendix

Model 1:

$$\text{PHQlast}_{ij} = \beta_{0j} + \beta_{1j}(\text{PHQfirst-gm})_{ij} + 0.117(0.041)(\text{GADfirst-gm})_{ij} + 0.012(0.005)(\text{PHQfirst-gm})_{ij} \cdot (\text{GADfirst-gm})_{ij} + e_{ij}$$

$$\beta_{0j} = 8.664(0.437) + u_{0j}$$

$$\beta_{1j} = 0.549(0.046) + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 2.779(1.131) & \\ 0.074(0.074) & 0.013(0.009) \end{bmatrix}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \sigma_e^2 = 29.122(1.250)$$

$$-2 * \loglikelihood = 7010.868(1122 \text{ of } 1122 \text{ cases in use})$$

Model 2:

$$\text{GADlast}_{ij} = \beta_{0j} + \beta_{1j}(\text{GADfirst-gm})_{ij} + 0.131(0.032)(\text{PHQfirst-gm})_{ij} + 0.014(0.004)(\text{GADfirst-gm})_{ij} \cdot (\text{PHQfirst-gm})_{ij} + e_{ij}$$

$$\beta_{0j} = 7.805(0.393) + u_{0j}$$

$$\beta_{1j} = 0.478(0.048) + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 2.262(0.917) & \\ 0.112(0.073) & 0.015(0.010) \end{bmatrix}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \sigma_e^2 = 23.436(1.006)$$

$$-2 * \loglikelihood = 6765.856(1122 \text{ of } 1122 \text{ cases in use})$$