Intervening to reduce therapist variability and improve patient outcomes in routine NHS psychological therapies services

K M James

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

2022

Intervening to reduce therapist variability and improve patient outcomes in routine NHS psychological therapies services

Katy Margaret James

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield

Faculty of Science

Department of Psychology

April 2022

List of Contents

I, Katy M James, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

Acknowledgements

A project of this sort can only be done collectively. I have been coaxed and pushed, challenging and nurtured by my endlessly patient supervisors, Michael Barkham and Dave Saxon. Dave has succeeded where all others have failed, talking me through complex statistics and leading me to an understanding of multilevel modelling. Who thought I could enjoy maths?! Michael has picked me up when I stumbled and collapsed, deftly steered my thinking to bring structure to chaos, and encouraged me to have confidence about this whole crazy project. His ability to encourage and challenge clinicians to embrace and own the science of our craft is unparalleled. Steve Kellett was an inspiration at the beginning of this journey, making what felt like far too intimidating an undertaking feel like something that even little ol' me could venture into.

My family has been an enduring support throughout my life and has been cheerleading throughout. You don't choose your family, so I feel thankful everyday to have lucked out in the way I have, being able to be part of such an incredible clan. From viva tips and punctuation consultation, to tea, cake and untiring belief – all these contributions have built this manuscript. My friends have also joined the cheerleading crew, patiently listening to thesis-writing woes, and bemusedly cheering at the exciting news of "1.8". I have had little to give back to these contributors over this time, so I am thankful for their patience.

Writing this thesis brought closure to one of the most exhilarating periods of my career so far. I could not have taken the professional risks I did without my wing-woman, and operational counterpart, Hayley Matthews. Her competence and willingness to put herself on the line for our joint vision, made the impossible, possible. Without the bravery of the entire leadership team, and our amazing therapists, none of the things described in these pages would have happened, and the benefits to patients would not have been realised.

And finally, I just simply could not have survived the last 6 years, physically or emotionally, without my incredible wife, Adele. There have been countless moments when it was only her unwavering faith in me that kept me believing I could finish this. She has walked alongside me throughout, leading me into the sunlight when I had been locked away with my laptop for days, bringing me sustenance when I had failed to eat, and generally keeping our life going when I had to check out of it. It seems I did finish it after all, Adele - we made it!

Abstract

The psychological therapies literature is dominated by comparisons between psychological treatments. However, therapist effects can account for more outcome variance. The attention of researchers to the phenomenon of therapist effects is increasing but there has been no research considering whether therapist variability can be reduced whilst at the same time also improving patient outcomes. Accordingly, the current thesis is framed as a proof of concept study testing the idea that interventions to support practitioners can be used to both reduce therapist variability in an Improving Access to Psychological Therapies (IAPT) service and also improve outcomes. Outcome data from 126 therapists and 6476 patients, spanning a 4.5-year period, were utilised to compare therapist effects in three different Phases: Baseline (Phase 1), Directive Intervention (Phase 2), and Collaborative Intervention (Phase 3). A core sample of 35 therapists who were constant across the three Phases were the primary focus, with a secondary analysis utilising the full dataset. Significant variables impacting outcomes were controlled for within multilevel models to allow the identification of therapist effects. A therapist effect of 4.9% was present at Phase 1 in the primary dataset, reducing slightly to 4.7% in Phase 2, and dropping to 1.8%, a minimal effect, in Phase 3. There was a significant improvement in patient outcomes on a key clinical service measure (i.e., PHQ-9 reliable improvement rate) across Phases. Results suggest it may be possible to reduce therapist effects and improve or maintain overall clinical outcomes in a service context. If implemented on a larger scale this would provide more consistent patient outcomes in services. It is not possible to determine cause and effect in relation to the moderators of therapist effects (e.g., deliberate practice). However, the results provide encouragement for efforts to reduce therapist variability and achieve more consistent patient outcomes in psychological therapies services.

Table of Contents

Chapter	Page

List of Tables

List of Figures

Glossary and Abbreviations

ADSM	Anxiety disorder specific measure
ANCOVA	Analysis of Co-variance
ANOVA	Analysis of Variance
CBT	Cognitive behavioural therapy
CCfD	Couples counselling for depression
CCG	Clinical commissioning group
CI	Confidence interval
CPD	Continuing professional development
DH	UK Department of Health
DIT	Dynamic interpersonal therapy
DP	Deliberate practice
DP-lite	Deliberate practice-lite – a less intensive form of deliberate practice
EMDR	Eye movement desensitisation and reprocessing
GAD-7	Generalised Anxiety Disorder-7: a patient reported measure of generalised anxiety disorder
GP	General practitioner
HI	Healing involvement
HLM	Hierarchical linear model
IAPT	Improving Access to Psychological Therapies
ICC	Intraclass correlation co-efficient
IGLS	Iterative Generalised Least Squares
IMD	English index of multiple deprivation
IPT	Interpersonal psychotherapy
IST	Intensive Support Team (part of NHS Improvement)
KPI	Key performance indicator
LI-CBT	Low intensity cognitive behavioural therapy

MCMC   Markov Chain Monte Carlo

MDS   Minimum data set

MLM   Multi-level model

NHSE   National Health Service England

NICE   National Institute for Health and Care Excellence

NPR   Negative personal reaction

PBE   Practice-based evidence

PCE-CfD   Person-centred experiential counselling for depression

PHQ-9   Patient Health Questionnaire-9: a patient reported measure of

                  depression

PrI   Probability interval

PROM   Patient-reported outcome measure

PSD   Professional self-doubt

PWP   Psychological wellbeing practitioner (also known as low intensity

                  therapists)

RCT   Randomised control trial

ROM   Routine outcome measure

SI   Stressful involvement

TAU   Treatment as usual

TE   Therapist effect

WSAS   Work and social adjustment scale: a patient reported measure of

                  functional impairment

wte   Whole time equivalent

Chapter 1

Introduction

Overview

This chapter provides the context for the development of the central question that this thesis sets out to address. First, the chapter will introduce the main subject area, with the practical and clinical questions that first prompted the primary researcher to investigate this area. Second, a definition of the key phenomenon is provided, with a summarised history of the early identification of its existence within the field of psychological therapies research. Finally, an overview of the central research question, key hallmarks of the study design, and an overview of the thesis structure are provided.

Identification of the Research Questions

Therapists at the beginning of their careers generally start with the simple intention of wanting to provide help. Wanting to help their patients to solve the problems they come to therapy with, wanting to help them to feel better and live richer, more fulfilled lives. The question is: how is this best done? Generally, to answer this question an aspiring therapist will chose to train in a specific therapeutic approach, choosing this approach based on their own information gathering, interests, background and other more intangible preferences. Whilst in training, generally the first tendency of a trainee therapist is to emulate others, often those viewed as an expert in the field. An eloquent lecturer, a beloved supervisor, or a 'big name' in the therapy world. The trainee tries to model their approach on any glimpses they get of the way this person practices. But here is the first stumbling block. Opportunities to observe therapy are rare. Often the only experience of seeing what therapy looks like in reality is the trainee's own experience of being in the patient's chair – a quite different experience altogether. Increasingly there are videos clips of sessions, used for training purposes, with trainee therapists furiously noting down the phraseology, the facial expressions, the quality of the non-verbal communication of the therapist, to try out in the therapy they provide to their own patients. After qualifying, the therapist is free, their skills validated, and, so we are told, now the 'art' of therapy begins.

Therapists are even less exposed to the work of others following training, as supervision tends to be one-to-one, and opportunities to role play or see the work of peers becomes minimal, if ever. There may be opportunities to hear a case study being described by a fellow therapist, but this is therapy through the eyes of another, and rarely provides the objective detail of the processes in the therapy room to compare against one's own. So, and as we see from the literature, therapists' efficacy over time generally remains static (Kraus et al., 2016) or may even reduce over time (Erekson et al., 2017; Goldberg, Rousmaniere et al., 2016), though their confidence may increase (Nissen-Lie, Monsen et al., 2013; Nissen-Lie et al., 2017). The myth of the 'experienced therapist' tells them that they will naturally improve over time due to the simple exposure to working with different patients. And yet how do therapists know they are doing a good job relative to others? How do they know if they are working in the therapy room in a different way to their peers? And how do they know if the way they are working is more or less effective than the work of others. Certainly, the understanding of the mechanisms of therapy have established that therapies, and many of them, are effective, and generally equally so (Barkham & Lambert, 2021). But this has not dispelled the long-standing suspicion that there are those therapists whose conduct of therapy yield better patient outcomes than others (the legacy of 'supershrink'; Ricks, 1974). And the evidence has become sufficiently strong to support the view that some therapists *are* more effective than others – that there is, indeed, a therapist effect (Wampold & Owen, 2021).

But how can this be the case when therapies and therapy services are becoming increasingly manualised and standardised in their approach? When the components of therapy have been clearly spelled out in competency frameworks (e.g., Roth & Pilling, 2008), and therapy textbooks/training manuals (e.g., Hawton et al., 1989; Murphy, 2019). When ratings of clinical improvement and response to therapy are taken at every contact with a patient to check the progress and efficacy of the therapy work. A similar phenomenon can be seen in mainstream education settings even where a standardised teaching curriculum is applied. There is evidence that, despite this stringent adherence to a curriculum, by teachers

with similar training experiences, there are differences observed in student's examination outcomes. This is the case when student level variables, such as baseline educational achievement level and other significant variables are controlled for (Goldstein & Healy, 1995; Goldstein & Speigelhalter, 1996).

In the case of therapy, we see the same phenomenon. Despite the standardisation and manualisation of therapy in modern psychological therapies services, the variability in the outcomes observed by different therapists – even when patient variables such as age, gender, and symptom severity are controlled for – is seen across studies (Wampold & Owen, 2021). Given the context of the impact of psychological therapies on mental health outcomes, this variability has great significance. As a patient, it means that there is an element of chance associated with the extent of the improvement in their mental health they may experience as a result of their therapy. A therapist may look around and wonder, "*Am I doing this as well as I could?", "Why did that therapy not go as well, when my colleague providing the same therapy got such good outcomes with a patient with a similar presentation?"* Until recently a therapist might not have even known that a colleague was getting different outcomes to themselves. Even today, feedback based on the benchmarking of therapists' outcomes is only beginning to be routinely shared across therapist cohorts, and yet the ability to do such analysis of outcome data has been possible for decades (Wampold & Owen, 2021). So, an individual therapist is still left with the question: *"How can I help my patients with their difficulties?"*, and perhaps, *"How can I help more of my patients, and help my patients more effectively?"* At a service level, managers and commissioners are asking, *"How can we ensure that patients experience similar results no matter which of our therapists they see?".* Clinical leaders and supervisors are asking *"How can we better support all of our therapists to develop and learn to be the best they can be for their patients?".* It is from the starting point of these questions that provides the basis for the development of this thesis.

Summary of the History of Therapist Effects Research

In light of the view that therapist effects may be present, it is timely to provide an overview of the development of therapist effects research, establishing a necessary foundation upon which the work reported in this thesis is built. A basic summary of what therapist effects are and the way they are calculated is provided. In addition, the methodological and statistical implications for therapist effect research are briefly presented. Finally, an overview of the general understanding of the moderators of therapist effects that predate the initiation of this research project are summarised.

Understanding What Makes Therapists Effective

The debate about the elements of psychotherapy that affect change has been ongoing since both the widespread development of psychological therapies (for a summary, see Crits-Christoph & Gibbons, 2021; Miller et al., 2013) and the confirmation of the primary efficacy of the psychological therapies – namely that it does improve the symptoms of those experiencing mental health difficulties in comparison with alternative treatments or waiting list controls (Barkham & Lambert, 2021). However, despite there being recognition of the relative difference in impact of *therapists* on patients since as early as Rosenzweig's (1936) article, the psychological therapy literature has instead been dominated by comparisons between different *therapy treatments*, seeking to establish their relative efficacy (Wampold & Owen, 2021), together with the investigation of common factors in therapy comprising most of the remainder of the research field. This is despite the demonstration that therapist effects are greater than therapy treatment effects when therapist variability is accounted for (for a summary, see Barkham et al. 2017; Wampold, 2001; Wampold & Owen, 2021).

The first study of therapist effects (Ricks, 1974) resulted in a small flurry of interest within the psychotherapy research field. However, the first substantial meta-analysis of therapist effects, analysing data taken from 10 clinical trials, was not published until 1991 (Crits-Christoph & Mintz). Since then, reviews have been undertaken of therapist variables, such as a recurring chapter in earlier editions of both *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (Beutler et al., 1994, 2003), and *What Works for*

Whom (Roth & Fonagy, 2005; Roth, 1996). However, in the 6th edition of *Bergin & Garfield's Handbook of Psychotherapy and Behavior Change* (2013), chapters from previous editions of the *Handbook* on therapist *variables* were replaced by a new chapter, and focus, on therapist *effects* (Baldwin & Imel, 2013). This marked a watershed in the research field investigating therapy outcomes. The change marked a recognition and acceptance that the therapist is not merely an inconvenient variable or noise to be factored out of studies, but rather the effect of the therapist on therapy outcomes is a robust and legitimate phenomenon worthy of further exploration. Crucially, this focus on the therapist effect galvanised the use of random effects studies as the most favourable way to study the phenomenon, with the key result in such studies being the estimate of the amount of variance in outcome that can be associated with the therapist (Baldwin & Imel, 2013).

What is the Therapist Effect?

The therapist effect can most easily be described as the variability in outcomes between therapists. Despite people having the same therapy type for the same condition, there is variability in patient outcomes. This is, in large part, due to the differences between patients, such as their level of intake severity, but it is also due to differences between therapists. The extent to which the differences, the variability, between therapists impacts on patient outcomes is the therapist effect. This is depicted in Figure 1.1 (taken from Wampold & Owen, 2021), where the shade of grey represents the different objective clinical outcomes experienced by the patient (the darker the shade the better the clinical outcome). Patients may experience different outcomes compared to each other, based on each patient's particular features (e.g., symptoms severity, age, gender, presenting problem etc.).

However, what we find in practice is that patient A will get different outcomes depending on which therapist they are treated by, even more so if they have particular features themselves compared to other patients (such as higher symptom severity; Saxon & Barkham, 2012).

Figure 1.1

Illustration of Outcomes with no Therapist Effect
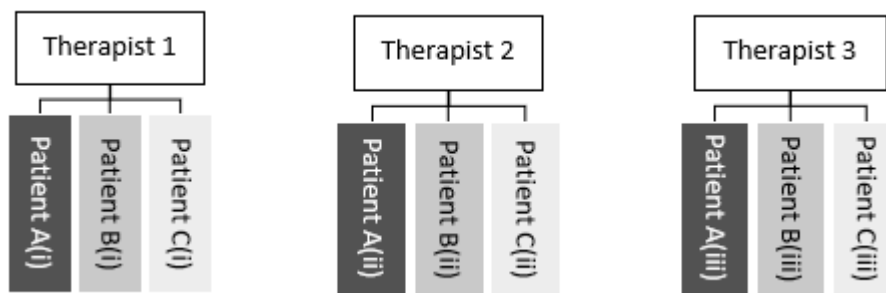
There is something about the therapist that is resulting in them achieving different outcomes with the same (or in reality, similar) patients. This phenomenon is called the therapist effect (see Figure 1.2 after Wampold & Owen, 2021).

Figure 1.2

Illustration of Outcomes with a Therapist Effect
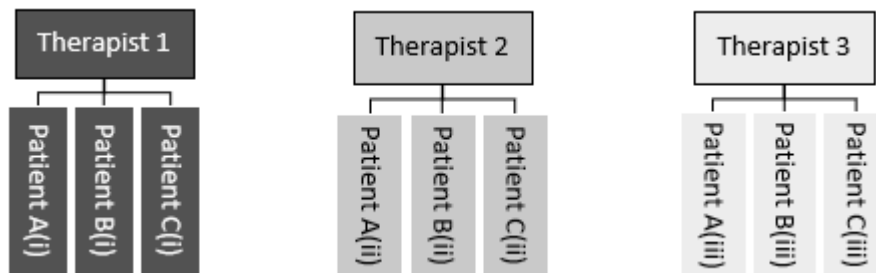
While psychological therapy services have limited control over the variability between patients entering a universal primary service, they are likely to have greater influence over the selection and appointment of therapists employed by a service. If the service is aware of the variability between therapists, it may be able to address this issue by training, supervision or support. In identifying what support might reduce this variability, we need to hypothesise what might be different about individual therapists. What is producing the therapist effect? And by implication, how can we make sure that patients are not exposed to a therapist lottery? How can we ensure that patient A gets the same, or similar, outcomes with therapist 2 and 3 as they do with therapist 1? Whilst at the same time ensuring that average outcomes overall are improved or at least remain the same. To answer this,

arguably, we must use the therapist effect as a flag or pointer that alerts us to the presence of such differences. By utilising the therapist effect as an outcome variable, this allows the possibility of manipulating potential influencing factors and measuring the impact this manipulation has on the therapist effect and thereby beginning to illuminate the elusive aspects of therapists' practice that explain this variability.

The implications of this point are considerable, if we consider the impact at a service, or indeed national level. Instead of simply comparing group means in isolation, these methods allow a meaningful understanding of the outcomes of a service. For example, two services with comparable patient populations may have the same mean outcome on a measure, but have very different outcomes for individual patients, as illustrated in Figure 1.3. In the example, both services have a mean patient outcome of 10 points of clinical change. However, in Service B it is clear that there is considerable variability in the outcomes achieved by the different therapists. In a scenario where a score of 8 represents clinically significant change, we can see that all the therapists in Service A are achieving clinically significant change, compared to only 58% of the therapists in Service B. As a patient or service manager, Service A would be the better service, despite the average outcome being the same.

In terms of Service B, one way of improving the experiences for patients would be to discontinue the employment of the least effective therapists – a quick, if brutal, way of reducing the variability and improving the mean patient outcome. Alternatively, if there was a way to improve the outcomes of the least effective therapists, or reduce the variability in the group whilst maintaining the average outcome, such as in Service A, this would provide a more consistent patient outcome for the population served by Service B. Through this example, the interrelation of variability and outcome is illustrated: more consistent outcomes across therapists to reduce the outcome lottery is desirable, but not at the expense of the overall outcome mean. Thus, isolating the therapist level factors that influence this variability in a positive way and seeking ways to equalise these factors across the therapist population may result in outcomes more akin to those seen in Service A.

Figure 1.3

Graph Showing Services with the Same Group Mean with Low and High Variability
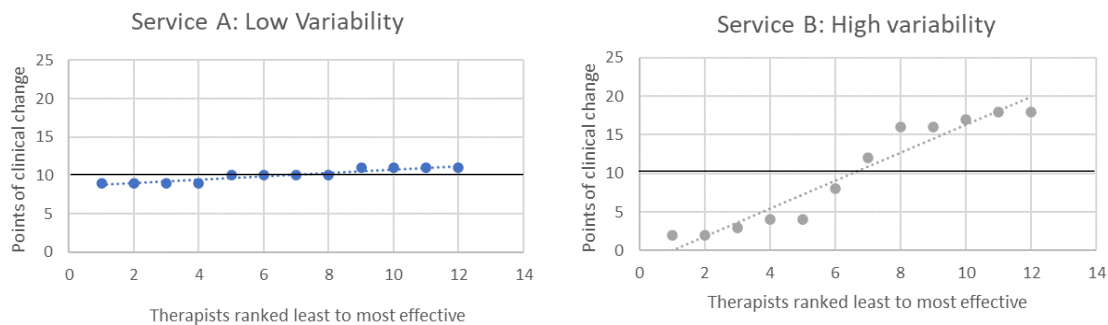
In summary, in order to have an influence on the therapist effect, we need to: 1) know the size of the therapist effect and whether it is having a significant effect on outcomes, 2) provide a therapist team-wide intervention, and 3) measure the size of the therapist effect afterwards to see if it has reduced. By so doing this could show the feasibility of the idea of reducing therapist variability while increasing/maintaining patient outcomes at an absolute level, thereby laying the foundations for future large-scale implementation.

The statistical method for detecting therapist variability and estimating the size and significance of therapist effects, multi-level modelling (MLM) – splits the patient outcome variance between the therapist level effects and the patient level effects after controlling for patient variables and allows the calculation of the population intraclass correlation co-efficient (ICC). The ICC is calculated as follows:

Equation 1

ICC equation

$$ICC = \frac{variance\ due\ to\ therapists}{variance\ due\ to\ therapists + variance\ due\ to\ patients}$$

The ICC is therefore the proportion of the outcome variance that is associated with therapists, and is a measure of how similar the patient outcomes are for a therapist and how different they are to outcomes of another therapist. The ICC is often multiplied by 100 to express the therapist effect as a percentage. This thesis aims to test whether an intervention can reduce the proportion of variance that is at the therapist level, as measured by a reduction in the ICC/therapist effect.

Establishing the Size of the Therapist Effect

The two most recent meta-analyses of therapist effects studies, Baldwin and Imel (2013) and Johns et al. (2019), both found a similar therapist effect value in naturalistic settings (ICC of .05, compared to .07 reported by Baldwin & Imel). However, in randomised control trials (RCTs), the additional meta-analysis of Johns et al., of 3 RCTs undertaken in the same period resulted in a therapist effect of .174 (compared to .03 reported by Baldwin & Imel) based on a weighted average of number of patients. Wampold and Owen (2021) suggest Baldwin and Imel's value, due to the larger number of studies included (29 studies), may be the more reliable. However, as highlighted by Baldwin and Imel (2013) at the time, all the RCTs included in the meta-analyses were underpowered for the number of therapists in the individual studies, highlighting the design and methodological problems in these studies.

Baldwin and Imel (2013) characterised the problems that have resulted in the neglect of therapist effects, as falling into three areas:

1. Studies are often designed to minimise therapist differences (as they generally want to study treatment effects)

2. Studies comprise numbers of therapists and patients too small to detect therapist effects

3. Where bigger datasets are used, heterogeneity of patients can be so large as to obscure therapist effects

There have been several major recommendations made on how best to study therapist effects (Baldwin & Imel, 2013; Crits-Christoph & Gallup, 2006; Elkin et al., 2006; Schiefele et al., 2017; Wampold & Bolt, 2006, 2007; Wampold & Owen, 2021). Due to the relatively small (though often significant) effect size in therapist variability studies, large number of therapists and patients per therapist are generally required to allow the variability to be reliably detected (Maas & Hox, 2004; Schiefele et al. 2017), with a general rule of a minimum of 1200 patients being required to detect therapist effects (Schiefele et al., 2017). A refresh of the 2013 Baldwin and Imel review (Johns et al., 2019), a recent review of

effective therapist characteristics (Heinonen & Nissen-Lie, 2020), and the Wampold and Owen (2021) review of therapist effects, concluded that, although patient and therapist numbers had increased, studies specifically designed to measure therapist effects continue to be lacking in the field. In bringing together the history of therapist effects, Wampold and Owen (2021) argue that the lack of therapist effect studies and reporting of the therapist effect in the psychotherapy research field, has become difficult to defend, given the presence of statistical methods for calculating the impact of random factors, and thereby allowing the investigation of the impact of the therapist. However, they note that studying this important variable (i.e., the provider of the therapy) has had little attention, with less than one quarter of one percent of clinical trials between 1965 and 2013 reporting a calculation of a therapist effect.

Taking a critical position in response to the lack of reporting and consideration of the therapist effect in modern psychotherapy research, Wampold and Owen detail the significant consequences of ignoring the therapist effect. They illustrate how, by so doing, this continues a flawed methodological assumption: that all therapists provide therapy treatments uniformly. In addition, omitting to factor in therapist effects into research design and analysis leads to results inflating the error rate of the treatment fixed effects, thereby also inflating the comparative treatment effects estimate (for details, see Wampold & Owen, 2021). In practical terms, this leaves a question mark over the relative efficacy of treatments, impacting broader decisions regarding the funding for, and availability of, a range of psychological therapy treatments for patients.

Baldwin and Imel's (2013) review of therapist effects research provided the most comprehensive and recent review of this research field at the time of the current research. The follow up to the Baldwin and Imel review, Wampold and Owen (2021), was published *after* the research study and was therefore utilised within a validation review rather than the main literature review. Chapter 2 presents the systematic literature review of publications between the Baldwin and Imel (2013) review paper and the development of the research Phase of the study (2018). The validation literature review from the period 2018 - 2020,

including the findings from Wampold and Owen (2021), is presented in Appendix A.

Accordingly, the following section provides a summary of the literature establishing

explanatory factors for the therapist effect prior to the literature review in Chapter 2, as a

context for the development of the research intervention being studied in this thesis.

Explanatory Variables for the Therapist Effect

The focus of the current study was to assess the potential for reducing therapist

variability and improving patient outcomes. In better understanding the phenomenon of the

therapist effect, the preparation for the work reported in this thesis concentrated on literature

that proposed or measured factors associated with the therapist that might result in or

explain the variability of therapist outcomes. Isolating therapist factors that may explain

variability between therapists would provide the focus for an intervention targeting these

factors in order to reduce therapist variability but also positively impact on outcomes (i.e.,

would not reduce overall outcomes for patients). This section therefore focuses on therapist

factors that improve patient outcomes, referred to as therapist effect variables or explanatory

variables.

The influence of the therapist on patient outcome, differentiated from therapy type

and patient factors, started to be measured more robustly with the advent of large practice-

based datasets both in the US (e.g., Wampold & Brown, 2007), and increasingly in the UK

(e.g., Saxon & Barkham, 2012). Studies that link measures of therapist effects with variables

that may influence these effects have begun to be published, though this remains a

developing design for studies measuring therapist effects. This can perhaps be explained by

the lack of consensus on the factors influencing successful therapy and the complexity of the

factors involved (i.e., a combination of patient effects, therapist effects, and intervention

effects).

Baldwin and Imel (2013) did not include a review of explanatory variables for

therapist effects in their comprehensive review of therapist effects as it was reported that no

major findings had been published in the intervening years between the 5th and 6th editions

of the *Handbook* in 2003 and 2013, respectively.

Summary of Findings from Beutler et al. (1994, 2003) Reviews

A summary of the key findings from the Beutler et al. (1994, 2003) reviews is structured by the state/trait groupings presented in the 2003 review, which in turn is based on the taxonomy of therapist variables presenting by Beutler et al. (1994) (see Figure 1.4).

Figure 1.4

Classification of Therapist Characteristics (Beutler et al., 1994, 2003)
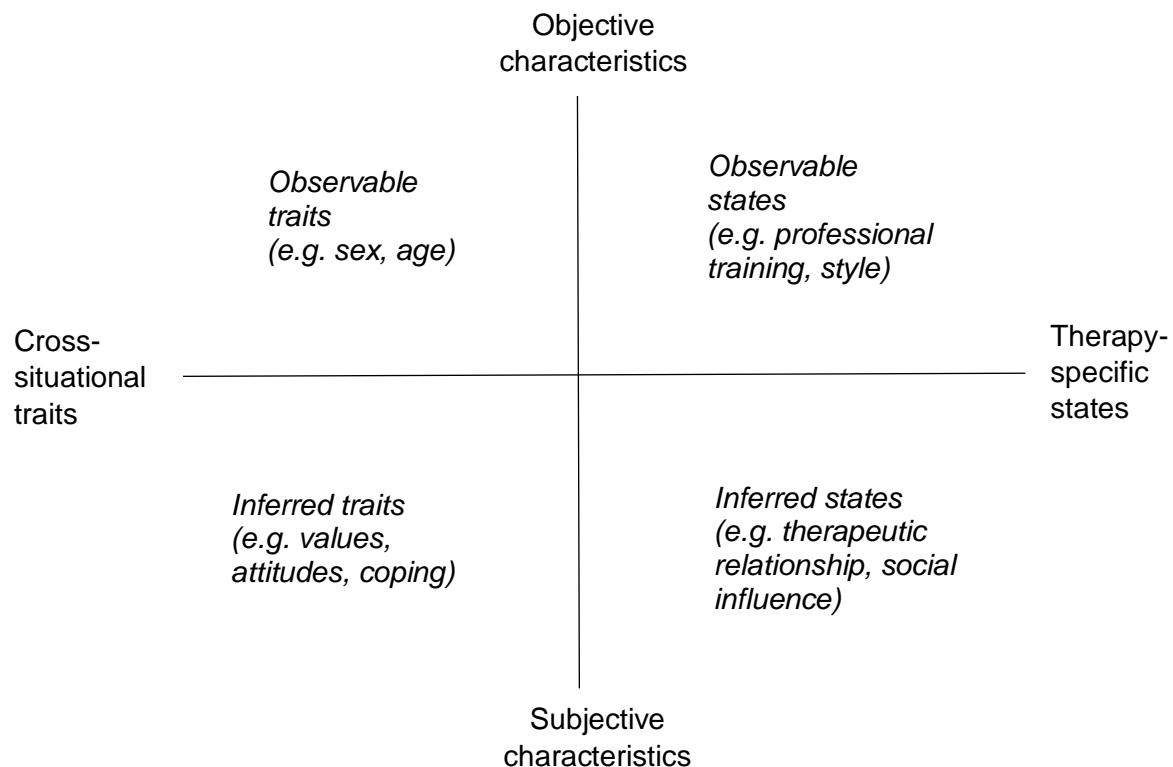
The therapist effect variables from the 4th (1994) and 5th (2003) editions are summarised in Table 1.1.

Observable Traits. Therapist sex, age and ethnicity were generally found to be poor indicators of patient outcome within the reviews. It was noted that age is not a robust variable in therapist effect research, mainly due to the confounding impact of experience and therapeutic orientation. Only one study was found between 1990 and 2000 that systematically studied age as a predictor variable (Barber & Muenz, 1996) and it reported there was no association between therapist age and patient outcomes. Out of ten studies investigating sex included in the 2003 review, only one found a significant effect of sex on outcome (Krippner & Hutchinson, 1990).

Table 1.1

Summary of Potential Explanatory Variables Included in Literature to 2012

Explanatory Variables Studied Beutler et al., **1994**	**Additional Explanatory Variables** Beutler et al., **2003**
Observable traits	
Age	No additional variables studied
Sex	
Ethnicity	
Observable states	
Professional background	No additional variables studied
- Level of professional training	
- Amount of experience	
- Professional discipline	
Therapeutic styles	No additional variables studied
- Interpersonal styles	
- Verbal styles	
- Nonverbal styles	
- Combined verbal & non-verbal patterns	
Therapist interventions	Therapist interventions
- Use of therapy manuals	- Supervision
- Therapist skill	- Adherence/Compliance
- Specific therapeutic procedures	- Therapist directiveness vs. patient self-
- Therapist directiveness	direction incl. use of homework assignments
- Therapist self-disclosure	- Insight-oriented vs. symptoms-oriented
- Therapist interpretation	interventions
	- Emotive vs. supportive interventions
	- Treatment intensity

Inferred traits	
Personality & coping patterns	No additional variables studied
- Dominance & related constructs	
- Locus of perceived control	
- Conceptual level	
Emotional well-being	
- Disturbance/distress	
- Personal therapy	
Values, attitudes, and beliefs	
- Religious beliefs	
- General values & attitudes	
Cultural attitudes	Cultural attitudes
- Gender & lifestyle	- Gender & ageism
- Socioeconomic background	

Inferred states	
Therapeutic relationships	No additional variables studied
Social influence attributes	
Therapist expectations	
Therapeutic philosophy/orientation	

The argument was made by the authors that more research should be conducted on this subject due to the socio-political concerns regarding the power of men and how this might impact therapy with women. Research investigating the impact of therapists' ethnicity on patient outcomes tends to focus on the ethnic similarity/difference between patient and therapist, based on the theory that a shared cultural background may have a positive impact on outcomes. Of ten studies (1990-2000) investigating outcome (clinical improvement), five studies found evidence of a significant positive effect of patient-therapist ethnic similarity on

outcomes (Hosch et al., 1995; Ricker et al., 1999; Snowden et al., 1995; Sue et al., 1991; Yeh et al., 1994). Drop-out from therapy also appeared to be impacted in naturalistic studies (e.g., Sterling et al., 1998), with patients from ethnic minorities more likely to drop out if their therapist was from the dominant ethnicity.

Observable States. Therapist training, experience, style, and skill are generally poor indicators of patient outcome. There is evidence that specific interpersonal styles encourage reciprocal patient responses, such as complementary and positive styles, and have a positive relationship to patient outcomes (e.g., Andrews, 1990; Beyebach & Carranza, 1997). There is some evidence that encouraging a reciprocal verbal interaction (e.g., initiating and agreeing new topics) with patients will produce better patient outcomes (e.g., Tracey, 1986). More recent research within the reviews focused on the area of treatment methods and outcomes, with some evidence for the benefits of manualised approaches. However, generally these would be likely to be non-significant if patient factors were included in the studies. The reviews concluded that there were inconsistent findings in relation to adherence/compliance and therapist skills and a lack of research investigating a relationship between supervision and patient outcomes.

Some evidence of the benefits of the use of homework were observed, but again the authors concluded that these were likely to be non-significant if patient factors were included in the analyses. The reviews found limited evidence of any significant difference in patient outcomes between insight vs. symptom-oriented interventions. No clear pattern was found within the evidence for the benefit of emotional arousal in therapy. However, studies including patient factors did show benefit for some patients with specific characteristics (e.g., Beutler et al., 1996).

There was an inconsistent picture when considering the impact of treatment intensity on outcomes, with some studies finding increasing intensity for more impaired patients to improve outcomes (e.g., Beutler et al., 2000), while others found that it was the least impaired patients that benefitted from an increase in intensity (e.g., Schulberg et al., 1998). The authors concluded there was a lack of research into the relationship between therapist

disclosure and patient outcomes. However, the three studies that were reviewed all found a positive impact of disclosure on patient distress/symptoms (Barrett & Berman, 2001; Piper et al., 1998,1999), though the studies had methodological limitations.

Inferred Traits. The reviews found inconsistent evidence for personality related variables on patient outcomes and concluded that this was generally an area of little interest in the contemporary research field, with studies on this topic reducing over time. Most literature in this area focussed on the matching of patient-therapist dyads in relation to outcome, but again this research showed no clear pattern of results. Similarly, the impact of therapist conceptual level has been concentrated on therapist-patient dyads and research has not progressed since early findings suggesting a match of cognitive level between patient and therapist yields lower drop out and quicker early session improvement (e.g., Hunt et al., 1985).

There is some evidence that therapist dominance produces a negative impact on patient outcomes (e.g., Henry et al., 1990). The Beutler et al. (1994) *Handbook* review found promising results reported in the relationship between therapist locus of control and patient outcomes (e.g., Antonuccio et al., 1987), but little research on this topic has been conducted in the time since. Beutler et al. (2003) concluded there was evidence of a positive effect of therapist wellbeing and cultural attitudes on outcomes, but a lack of research on the impact of therapist personal therapy on patient outcomes. It was noted that therapist wellbeing cannot be assumed and studies on this topic demonstrated that there is variability in the wellbeing of therapists. There has been little development in the understanding of the relationship of therapists' values on patient outcomes, and literature on the effect of matching patient-therapist values is contradictory up to 2000 (Beutler et al., 2003).

Inferred States. The quality of the therapeutic relationship has been established in the literature over many decades and found to be consistently correlated with patient outcome. However, based on the reviews over the 1994-2000 period, Beutler et al., (2003) concluded that the strength of this relationship may be less than previously thought, proposing that it accounts for between 7% and 17% of outcome variance. Research

investigating the therapist as a 'social influencer' for patients was present in the 1994 review but with no progression in the later 2003 review. This was despite promising findings in relation to therapists' perceived expertise, attractiveness and trustworthiness by their patients having a positive relationship with therapy outcomes (e.g., Heppner & Heesacker, 1983; McNeil et al., 1987; Zamosttny et al., 1981). However, findings were mixed overall and methodological problems remain, perhaps explaining the lack of further research in this area. Therapist expectations of therapy and patients have been investigated historically, but no patterns determined, and this line of enquiry has largely been abandoned in the literature. Theoretical orientation of the therapist as a predictor of patient outcome has limited support in the research literature, though Beutler et al. (2003) noted that interest in the research field appeared to be shifting to specific therapeutic techniques as a predictor for patient outcomes.

Overall, the reviews highlight a lack of clear conclusions when it comes to the factors influencing therapist effects. These are largely due to the methodological problems explored by Baldwin and Imel (2013). However, the reviews do provide potential avenues for exploration when methodological challenges have been addressed and certainly discount areas for further research, such as more simplistic explanations of observable traits (e.g., age, sex etc.). The overarching view presented by these reviews up to 2013 is of a complex, multi-factorial configuration of, as yet undefined, therapist level explanatory variables that influence patient outcomes.

Forming the Research Area

Set against the above literature, as a therapist, supervisor and clinical leader working in the context of a large psychological therapies service the primary researcher sought to address the following question: Can therapist variability be reduced and, at the same time, patient outcomes be improved in a practical and economical way across a whole service? To respond to this question, the work reported in the current thesis set out to develop a research-informed package of support for therapists (a therapist-level intervention) that could be realistically implemented across a whole therapist cohort. The support package was

developed collaboratively with therapists within the service, based on a systematic review of the key therapist effect literature (see Chapter 2). This intervention was deliberately and consciously developed and implemented within the context of a collaborative and compassionate approach to change across the leadership and research team.

Key Features of Overall Research Study

The overall research study comprised four key features: (1) a large psychological therapies service; (2) the embedded nature of the primary researcher; (3) a longitudinal research design; and (4) two sequential intervention conditions.

Feature 1: Set in a Large Psychological Therapies Service. The research took place in a large psychological therapies service, under the Improving Access to Psychological Therapies (IAPT) initiative in England, one of the most standardised and highly measured nationalised psychological therapies initiatives in the world (described in more detail in Chapter 4). This provided a large cohort of therapists and patients, in a natural clinical setting, but with high levels of standardisation and control over key variables (such as training standards, therapy protocols, outcome monitoring, and supervision levels).

Feature 2: Embedded Nature of the Primary Researcher. As a senior member of the clinical leadership team, the primary researcher provided access to information about processes and procedures within the service. This allowed a fuller description of the implementation and service context during all study Phases, than would ordinarily be possible when undertaking such research. Although this was a strength of the research, it was also important to control the potential for bias inherent in such a design. This was done by the robust oversight of the study within the context of a research PhD supervision team, and by the inclusion of a therapist staff member into the onsite research team throughout the study, who was independent of the overall research project.

Feature 3: Longitudinal Design. Due to the primary researcher undertaking their PhD part-time, this allowed for a longitudinal study design, with a data collection period spanning 4.5 years. This was a vital component of the study, allowing the design to incorporate additional features naturally occurring in the service at that time and the ability to

observe therapists' effects over time. It provided sufficient time periods for each Phase of the study to enable service-wide changes to be implemented and stabilise, reflecting the realistic change cycles of such services in practice.

   Feature 4: Two Sequential Intervention Conditions. The study design compared a research-informed, collaboratively developed therapist level intervention (Phase 3) with an immediately preceding naturally occurring, directive service intervention, or period of service change (Phase 2). These two sequential intervention Phases were themselves preceded by a baseline period of service as usual (Phase 1). In this way, the two intervention Phases (2 & 3) were the primary comparators, with the baseline providing a secondary comparison for each intervention Phase.

   Although the study comprised a robust design for a large practice-based study, due to the nature of such research being in the context of a live service with multiple variables impacting on it, establishing causal relationships was not possible. A concurrent control group would have been required which would only be possible by randomising by services and introducing variability between services. This was not feasible on a practical level.

   The number of therapists and patients needed to meet sample size recommendations would require multiple services implementing the intervention. Without evidence that such as intervention might have the potential to impact therapist effects without a negative effect on patient outcomes, it would be unrealistic to expect multiple active services to commit to such an endeavour. Accordingly, the current study aimed to assess the proof of concept, of introducing a therapist level intervention to reduce therapist variability and improve outcomes, within a single service.

   Although this may limit the generalisability of statistical findings, the focus on a single service has advantages. Individual services are complex, ever-changing environments with multiple drivers and influences at play, the focus on a single service allowed insight into possible influencing factors on the service and its staff during the study. These are included in Chapters 4-6. However, the possibility that there may have been other factors, not

available in the data or outside the awareness of the researcher that may have impacted any findings should be noted.

To fully appreciate the sequential nature of the overall study, the above Phases are represented here chronologically and in historical context. Prior to the initiation of the research, the service was underperforming in its clinical outcomes (see Chapter 4) and was preparing to undertake a programme of change. The period prior to the implementation of this programme of change provided the initial Baseline Phase, to measure the impact of these service changes against. The second Phase comprised the programme of service change itself, which was externally driven by the senior management team and not under the control of the research team or primary researcher. Importantly, this Phase served to determine the effects of a focused intervention on the baseline measures. This served as a more sophisticated comparator condition for the third Phase, allowing a comparison of the effects of an additional research-informed intervention following a period of focused service changes/improvements. Given the underperformance of the service initially, improvements in outcomes following any reasonable service enhancements may have been inevitable. Therefore, the design of the study aimed to establish the service baseline performance, then measure the impact of a naturally occurring, externally driven series of service changes, with sufficient time for these to embed and stabilise, prior to the implementation of the research intervention. This design thereby provided a comparator for the research intervention with increased external validity than simply a baseline period of service underperformance. Expressed in another way, the second Phase, following the baseline, acted as a quasi-control for the impact of a system wide organisational change on therapist outcomes. This in turn provided the context for the final Phase, the implementation of a research-informed therapist support package, with any effect of this specific intervention to be compared primarily to the second Phase of service improvements. In all, this unique design provided the ability to compare a group of therapists' outcomes during two different intervention Phases, with any differences in these two intervention Phases being more reliably attributed to some aspect of the intervention itself, rather than simply the impact of any organisational

change during that period. To provide this assurance, comparisons could also be made between the intervention Phases and the Baseline Phase.

In summary, the work reported in the thesis comprised a naturalistic longitudinal design with two sequential comparators (baseline, and organisational change) giving a baseline for a subsequent research-informed intervention. The overall aim was to determine if therapist variability could be reduced and patient outcomes improved, through the implementation of a service-wide, collaboratively developed but research-based, low-cost package of therapist support in the context of a large, NHS psychological therapies service.

Due to the longitudinal design, and the naturally occurring service wide change taking place prior to the research intervention, the structure of the initial chapters of this thesis do not reflect the chronology of the study, therefore a diagram is provided in Figure 1.5 as a map to the study and thesis structure.

Figure 1.5

Research and Thesis Structure Map

	Phase-specific Chapters	Overall Chapters
Oct 2015 — **Phase 1:** Baseline (Control Phase) Baseline data collection of therapist effects and outcomes	Chapter 3	Chapter 1: Introduction
Mar 2017 — **Phase 2:** Directive Intervention (Control Phase) Period of externally-driven service changes	Chapter 4	Chapter 6: Methods
Development of collaborative intervention	Chapter 2	Chapter 7 & 8: Results
Oct 2018 — **Phase 3:** Collaborative Intervention (Research Phase) Implementation of collaboratively developed, research informed therapist support package	Chapter 5	Chapter 9: Discussion
Mar 2020		

Summary

This chapter has presented the early development of the understanding of the therapist effect, as a backdrop to the foundational questions that prompted the three-Phase study presented in this thesis. The chapter has provided a basic explanation of what the

therapist effect is, why it is important in psychological therapies research and practice, and the early investigations of what may explain or moderate this effect. This provides a starting point for the systematic review of more contemporary therapist effects research presented in the following chapter. A map to the thesis structure has been provided, illustrating the longitudinal design of the research study Phases, whilst providing the context for the point in the study where the results of the systematic review were applied. This chapter has also highlighted the key hallmarks of this thesis, which will be demonstrated in more depth within the main body of the thesis. Overall, this chapter has aimed to give a sense of the practical, as well as the research-based, motivations for investigating the central question: Can therapist variability be reduced and patient outcomes be improved, in a practical and economical way across a whole service?

Chapter 2

Therapist Effects Systematic Literature Review

Overview

As discussed in Chapter 1, although therapist effect literature has grown over the last two decades, there has been limited exploration of the explanatory or moderating factors associated with the variability of therapists' outcomes. The purpose of the current review was to identify any factors that may have been hypothesised as influencing therapist effects. Using this literature and themes highlighted in previous key review papers on therapist effects (e.g., Baldwin & Imel, 2013; Beutler et al., 2003,1994), areas that may be relevant to any increased between-therapist consistency of outcomes, could then be identified to contribute to the development of a package of therapist support with the aim of reducing variability and improving patient outcomes overall.

The purpose of this review was to support the development of a feasible intervention to reduce variability between therapists and improve outcomes for patients in practice. With this in mind, a more focussed and pragmatic literature search was undertaken, but adopting the search terms used in the Baldwin and Imel (2013) review (see Chapter 1). The current review focused on studies that went beyond identifying the therapist effect, in addition to looking at explanatory factors for this effect. Research that was primarily focused on the investigation or critique of different research or statistical methods of studying therapist effects or psychotherapy outcomes was excluded. Comprehensive critiques of, and recommendations for, research methods in this area have been set out in other literature: Baldwin and Imel (2013), Crits-Christoph and Gallup (2006), Elkin et al., (2006), Schiefele et al., (2017), and Wampold and Bolt (2006, 2007). A summary of the review of therapist effects included in the recently published 7th edition of *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (2021) is included in the validation literature review in Appendix A. It is not included here as it was not published at the time of the review and was therefore not an available source for the development of the research intervention in 2018.

Summary of Baldwin and Imel (2013)

Baldwin and Imel's (2013) chapter marked a significant milestone in therapist effect research, bringing together a clear definition of the therapist effect, and recommending ways of measuring and researching this phenomenon. They defined the therapist effect as *"the effect of a given therapist on patient outcomes as compared to another therapist"* (p. 259). Their review of the literature on therapist effects sought to confirm the existence and prevalence of this effect. For the purposes of this literature review, however, Baldwin and Imel concluded that no further progress had been made in relation to identifying the explanatory factors for the therapist effect since the preceding chapter (Beutler et al., 2003, summarised in Chapter 1). They focused on the conclusions from the Beutler et al. (2003) review that inferred states, meaning therapy specific but subjective aspects, such as the therapeutic relationship, may hold the most promise currently when accounting for the differences among therapists. However, they also recognised the limitations of existing research in being able to isolate the therapist contribution to these factors, highlighting an ongoing challenge within this research field.

When accounting for differences between therapists there is an assumption made that certain factors will be more under the influence of therapists than others (e.g., therapeutic alliance, adherence to treatment protocols). However, this relationship has not been established through current research methodology. A small number of studies have determined the association for both between- and within-therapist alliance with outcome using multi-level modelling, as summarised by Baldwn and Imel. However, this remained a limited area of study.

In their concluding recommendations, Baldwin and Imel suggest that a better understanding of the factors influencing the therapist effect – or stated another way, what differentiates a less or more effective therapist? – would then lead to ways of improving the outcomes of less effective therapists. Based on studies with implications for influencing the therapist effect to date, they suggest that training or support in specific treatment protocols, development of therapy micro-skills, and therapeutic relationship building are all areas with

some potential. An alternative would be providing therapists with real-time patient outcomes to assist with the treatment of live cases.

Methods

The aim of this review was to provide evidence for the development of an intervention that could be implemented in practice with the goal of reducing therapist variability (the therapist effect) whilst improving patient outcomes. The primary search was the identification of literature published after the Baldwin and Imel (2013) review up to May 2018, that relates to therapist effects and investigates potential factors that may improve outcome and that appear to operate at the therapist level.

Search Criteria and Strategy

Using 2012 as a cut off based on the Baldwin and Imel (2013) review, the search aimed to capture studies that would not have been included in this comprehensive review of the therapist effects literature. As the intervention was to be implemented in October 2018, the review was carried out in July 2018 to provide sufficient time for an intervention to be developed by the implementation date. Therefore, this review includes studies published between January 2012 and May 2018. An additional search was conducted following research analysis completion to capture studies from June 2018 – Jan 2021 and these are included in the validation review chapter but are not included in this primary systematic review. Inclusion and exclusion criteria are shown in Table 2.1. As the purpose of this review was to begin to generate ideas for the development of a feasible in-practice intervention package with the aim of reducing therapist variability and improving outcomes, only studies that included investigation or discussion of possible explanatory variables for therapist effects, or ways of moderating therapist effects, were included.

Four online electronic literature databases were searched as follows: PsycINFO, Medline, Web of Science, & SCOPUS. Search terms were developed based on the Baldwin review; "*therapist effects*", "*differential effects of therapists*", "*therapist outcome*", "*therapist variance*", "*effective therapist*", "*ineffective therapist*" AND "*intraclass correlation*"; "*multilevel model*"; "*hierarchical linear model*"; "*mixed models*". Titles and abstracts were screened by

the first author, followed by a full-text review. Of the eligible studies following full text review,

reverse citation searches were conducted manually to identify any further eligible studies.

Table 2.1

Inclusion and Exclusion Criteria

	Inclusion Criteria	Exclusion Criteria
Patient Population	Patients aged 18yrs+ accessing psychological therapy for depression and/or anxiety disorders.	Studies exclusively with children and/or adolescents. Studies exclusively with non-clinical samples. Studies exclusively with drug & alcohol misuse populations
Intervention	Any form of psychological therapy aiming to treat depression and/or anxiety disorders, delivered individually in person/via telephone.	Studies that do not use psychological therapies to treat depression and/or anxiety. Studies that used interventions provided exclusively via group or internet (e.g. cCBT).
Outcomes	Studies that report the therapist effect. Studies that explicitly investigated, or discussed possible explanatory variables for therapist effects, or ways of moderating therapist effects.	Studies where the therapist effect is not measured/reported. Studies where explanatory variables for therapist effects are not mentioned.
Setting	Any outpatient settings	Studies exclusively based on in-patient settings.
Study design	Randomized controlled trials and observational studies. Studies published in peer-reviewed journals, in the English language.	Grey literature such as dissertation abstracts not published in peer-reviewed journals. Literature not published in the English language

Figure 2.1 presents a PRISMA diagram of the selection and screening process.

Figure 2.1

PRISMA Diagram of Study Selection
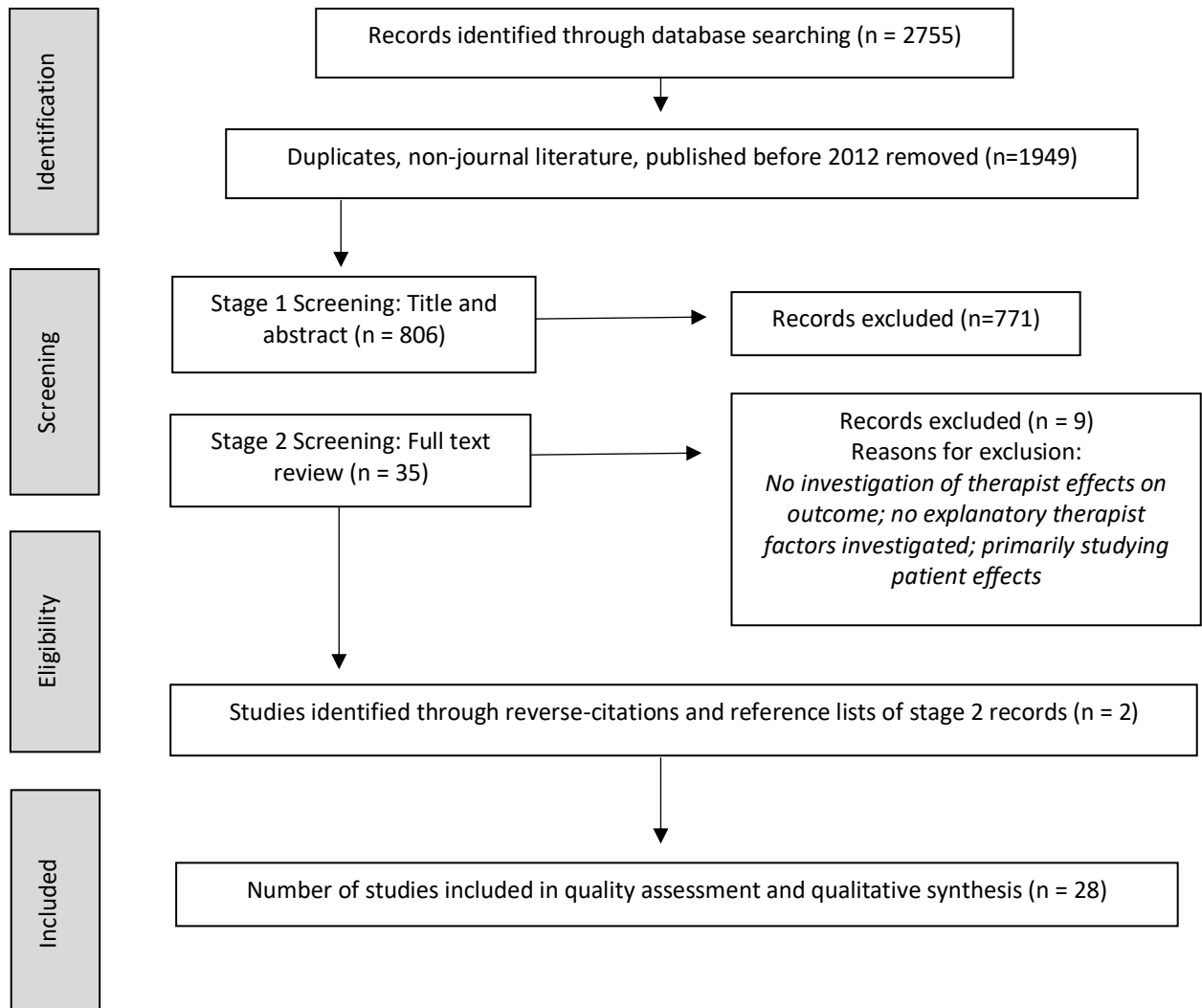
Each eligible study was rated by the primary researcher using the *Quality*

Assessment Tool for Observational Cohort & Cross-sectional Studies or the *Cochrane*

Handbook for Systematic Reviews of Interventions risk of bias tool (Higgins et al., 2011). An

additional rater, a counsellor from within the service, determined reliability of the quality

checklist scores by examining a set of 20% of all studies (i.e., 6 studies). One study rated as

'fair' quality and five studies rated as 'good', as determined by the primary researcher, were

examined by the additional rater.

Results

  The results of the systematic review are presented below. The features of the studies included are presented first, followed by a summary of the studies under key headings, consistent with areas identified earlier in this chapter.

Study Characteristics

  Twenty-eight papers were included in the final review and are described in Table 2.2. Two controlled trials were included in the review with the remaining 26 being observational studies. Both RCT's were rated as having a moderate risk of bias. Of the 26 observational studies, 8 studies were rated as fair quality and 18 studies were rated as good quality (see Appendix B). Samples sizes in the studies ranged from 22 – 586 therapists, with a mean of 87.9 (SD = 109.2) and 98 – 13664 patients, with a mean of 3583.6 (SD = 3733.3). Of the studies that reported patients per therapist ($k = 26$), there was a mean patients per therapist range of 5 to 184, with a mean of 47.5 (SD = 47.5). Treatment setting included out-patient clinics ($k = 10$), primary care ($k = 9$), university counselling clinics ($k = 8$), secondary care services ($k = 4$), voluntary/workplace/private counselling centres ($k = 4$). A range of psychological interventions were provided, including low intensity cognitive behavioural therapy (LI-CBT, $k = 4$), Cognitive Behavioural Therapy (CBT, $k = 5$), counselling ($k = 3$), psychodynamic psychotherapy *($k = 1$)*, cognitive processing therapy ($k = 1$) and settings offering mixed or non-specified psychotherapies ($k = 18$). Patient presentations tended to be depression and/or anxiety disorders ($k = 8$), post-traumatic stress disorder (PTSD, $k = 1$) or mixed presentations ($k = 19$).

Table 2.2

Study Characteristics

Study	No. of patients	No. of therapists	Mean patients per therapist	Diagnosis	Outcome measure(s)	Intervention(s)	Treatment Setting(s)	Therapist effects analysis	Significant therapist effects found	Quality assessment
RCT										
Errázuriz & Zilcha-Mano 2018	547	28	Not reported	Mixed	OQ-30	Mixed psychotherapy	Outpatient mental health centre	MLM	No	Moderate risk of bias
Zilcha-Mano & Errázuriz 2015	547	28	20	Mixed	OQ-30	Mixed psychotherapy	Outpatient mental health centre	MLM	No	Moderate risk of bias
OBSERVATIONAL STUDIES										
Ali et al. 2014	1376	38	36	Depression/Anxiety	PHQ-9; GAD-7	Brief low intensity CBT	Primary care IAPT	HLM	Yes	Good
Chow et al. 2015	4580	69	66	Depression/Anxiety	CORE-10	Mixed psychotherapy	Mixed (voluntary (42%; independent practice (39.1%); primary care (8.7%); secondary care (4.3%)	MLM	Yes	Fair
Cologon et al. 2017	1001	25	39	Depression/anxiety/mixed	OQ-45	Mixed psychotherapy	University counselling centres	HLM	Yes	Good

Erekson et al. 2017	4047	22	184	Depression/anxiety/mixed	OQ-45	Mixed psychotherapy	University counselling centres	HLM	Yes	Good
Firth et al. 2015	6111	56	109	Depression/Anxiety	PHQ-9; GAD-7; WSAS	Brief low intensity CBT	Primary care IAPT	MLM	Yes	Good
Goldberg, Hoyt et al. 2016	5828	158	37	Mixed	OQ-45	Mixed psychotherapy	University counselling centres	MLM	Yes	Good
Goldberg, Rousmaniere et al. 2016	6591	170	39	Mixed	OQ-45	Mixed psychotherapy	University counselling centres	MLM	Yes	Good
Green et al. 2014	1122	21	53	Depression/Anxiety	PHQ-9; GAD-7	Guided self-help (low intensity CBT)	Primary care IAPT	MLM	Yes	Good
Hayes et al. 2015	228	36	6	Depression/Anxiety/relationship problems/academic distress	OQ-45	Mixed psychotherapy	University counselling centres	MLM	Yes	Fair
Kraus et al. 2016	3540	59	60	Mixed	TOP	Mixed psychotherapy	Mixed (outpatient; independent practice; hospital; residential; day treatment program)	HLM	Yes	Good

Laska et al. 2013	192	25	8	PTSD	PCL	Cognitive processing therapy	Veterans hospital – outpatient and community	MLM	Yes	Good
Nissen-Lie, Havik et al. 2013	227 / 70	70	5	Depression/anxiety/mixed	IIP-64	Mixed psychotherapy	Out patient clinic	MLM	Yes	Fair
Nissen-Lie, Monsen et al. 2013	255 / 70	70	5	Depression/anxiety/mixed	GAF; SCL-90; IIP-64	Mixed psychotherapy	Out patient clinic	MLM	Yes	Fair
Nissen-Lie et al. 2016	6348	189	16	Mixed	OQ-45	Mixed psychotherapy	Mixed (University counselling centres, primary care units; psychiatric clinics)	MLM	Yes	Good
Nissen-Lie et al. 2017	370	70	5	Depression/anxiety/mixed	SCL-90; IIP-64	Mixed psychotherapy	Out patient clinic	MLM	Yes	Fair
Odyniec et al. 2017	621	40	15	Depression/anxiety/mixed	BSI; IIP-64	CBT	Out patient clinic	MLM	Yes	Good
Owen et al. 2016	13664	586	23	Mixed	BHM-20	Mixed psychotherapy	University counselling centres	MLM	Yes	Good
Pereira et al. 2017	4980	37	135	Depression	PHQ-9; GAD-7; IMD	CBT, counselling, low intensity CBT	Primary care IAPT	MLM	Yes	Fair
Rousmaniere et al. 2016	6521	175	37	Mixed	OQ-45	Mixed psychotherapy	Counselling centre	HLM	Yes	Good

Saxon & Barkham, 2012	10786	119	91	Depression/anxiety	CORE-OM	CBT, counselling	Primary care psychotherapy service	MLM	Yes	Good
Saxon, Barkham et al. 2017	10521	85	124	Depression/anxiety	PHQ-9; drop out	Mixed	Primary care IAPT	MLM	Yes	Good
Saxon, Firth et al. 2017	4034	61	Not reported (20+)	Depression/anxiety	PHQ-9	CBT, counselling	Primary care IAPT	MLM	Yes	Good
Slone & Owen 2015	247	49	5	Mixed	PEI	Mixed psychotherapy	University counselling centre	MLM	Yes	Fair
Xiao et al. 2017	5253	83	59	Mixed	CCAPS	Mixed psychotherapy	Counselling centres	MLM	Yes	Good
Zeeck et al. 2012	98	26	Not reported	Mixed	TWIS	Psychodynamic psychotherapy	Outpatient clinics; day hospitals	Mixed models	Yes	Fair
Zimmerman et al. 2017	707	66	11	Mixed	Drop out	CBT	Outpatient centre	MLM	Yes	Good

Note: BHM-20 = Behavioral Health Measure; BSI = patient reported assessment of symptom severity; CCAPS = Counselling Centre Assessment of Psychological Symptoms; CORE-10 = 10-item version of Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM); GAD-7 = Generalised Anxiety Disorder-7; GAF = Global Assessment of Functioning Scale; IAPT = Improving Access to Psychological Therapies; IMD = Index of Multiple Depravation; IIP-64 = Inventory of Interpersonal Problems-64; OQ-45 = Outcome Questionnaire-45; PCL = PTSD Checklist; PEI = Patient Estimation of Improvement; PHQ-9 = Patient Health Questionnaire-9; PTSD = Post-traumatic stress disorder; PDS = Posttraumatic Diagnostic Scale; RCT = Randomised Control Trial; SCL-90 = Symptom Checklist-90; TOP = Treatment Outcome Package; TWIS = Therapist Work Involvement Scales; WSAS = Work and Social Adjustment Scale.

Therapist Effects: Explanatory Factors That May Moderate Therapist Variability

The key areas identified within the literature as potentially having an influence on the phenomenon of therapist effects have been separated into three, as shown below: Intrapersonal, interpersonal, and contextual factors. Some studies investigated several factors which may fall under more than one category, and these are included under each heading.

1. Intrapersonal Characteristics of the therapist – aspects of the therapist that relate to their own emotional experiences, coping, and personal characteristics (e.g., resilience, mindfulness)

2. Interpersonal Factors between therapist and patient – aspects of the therapist that directly relate to the development and maintenance of relationships with others (e.g., therapeutic alliance, empathy)

3. Context of the therapist – the context that the therapist experiences when carrying out their work, including the development of professional skills (e.g., supervision framework, knowledge/skills acquisition, support structures)

Two studies investigated therapist effects in relation to drop out, with a therapist effect range from 5.7% (Zimmerman et al., 2017) to 12.6% (Saxon, Firth et al., 2017); or non-attendance, with a therapist effect range from 1.1% - 1.4% (Xiao et al., 2017); and deterioration, with a therapist effect of 10.2% (Saxon, Barkham et al., 2017). Explicit explanatory variables were not explored for these effects. However, the impact of therapist variability in relation to these outcomes are important when considering possible interventions to improve patient outcomes, so they were included in the review scope.

Intrapersonal Characteristics. As with previous studies and reviews, measures of characteristics such as age and gender were not an explanatory variable for therapist effects on outcomes within the review of studies that included this as a potential factor (Chow et al., 2015; Cologon et al., 2017). Ten studies included exploration of other intrapersonal therapist characteristics as potentially explaining some of the therapist variability in outcomes and are detailed here.

Therapist Attachment & Reflective Functioning. Cologon et al. (2017) included therapist attachment style as a variable in a study of 25 therapists and 1001 patients. Therapists were either post-graduate students or therapists working at a university counselling centre, with a range of theoretical orientations. Patients comprised a range of ages and presented primarily with depression and anxiety. Hierarchical Linear Modelling (HLM) was used to analyse the data which revealed no interaction between therapist attachment style and therapist effectiveness. The study also measured therapist reflective functioning (ability to reflect from different perspectives), which was found to predict therapist effectiveness. In addition, therapist attachment was reported to have an interaction with reflective functioning. Within the study, the effect of reflective functioning on outcomes was greater for those therapists with higher attachment anxiety, with reflective functioning ameliorating to some extent the impact of this anxiety. For those therapists with lower reflective capacity, higher attachment anxiety had a more negative impact on their clinical outcomes, with a more secure attachment style ameliorating the effects of the lower reflective capacity. In conclusion, the authors suggested that encouraging and developing the reflective capacity of therapists could have a positive impact on patient outcomes and may be a common factor applicable across therapies. Considering ways of improving therapist reflective capacity to a consistent standard across the service as part of the Phase 3 intervention package could therefore be a consideration. This naturalistic study design had lower numbers of therapists than recommended for this type of data analysis (below the 1200 patients suggested by Schiefele et al., 2017), and large variability in numbers of patients per therapist (from 4 to 219).

Resilience. Two studies measured therapist resilience as a possible factor explaining therapist variability, and one study implicated therapist 'burn-out' within its conclusions. Green et al. (2014) utilised a dataset of 21 psychological wellbeing practitioners delivering cognitive behavioural therapy-based guided self-help to 1122 patients in a naturalistic setting. Multilevel modelling was used to analyse the data, and self-reported measures of ego strength, intuition and resilience were captured, as well as a supervisor

rated intuition measure. Therapist effects accounted for 8.7% of outcome variability on a depression measure and 8.8% on an anxiety measure. Therapists were ranked according to their outcomes, which yielded therapist groups of higher and lower effectiveness which were used to compare the therapist characteristics. However, these groups had small therapist numbers (n=5 in each group). This comparison indicated that the more effective therapists had higher resilience scores than less effective therapists. In the lower effectiveness therapist group, supervisors rated these therapists as being more intuitive when making clinical decisions. This was in contrast with the more effective therapists who were rated as having a less affect-driven processing style. Measures of ego strength were not significantly different between the groups.

In a naturalistic study of 37 therapists treating 4980 patients, similar in clinical context to Green et al. (2014), Pereira et al. (2017) used benchmarking and MLM to identify more and less effective therapists. They then used these groups to compare therapist factors in each. In this study, two subgroups were created of more and less effective therapists working with moderate and then moderately severe patients. This grouping was based on a previous finding that revealed greater therapist variability with high severity patients (Saxon & Barkham, 2012). Findings demonstrated that for the higher severity patient group, the more effective practitioners had significantly higher levels of mindfulness alone, as well as resilience and mindfulness combined, compared to less effective therapists. This finding was non-significant for the lower severity patient group. The therapist effect across the entire dataset was calculated at 6.7%. Similar to Green et al. (2014), this study is limited by using relatively small numbers of therapists overall, meaning that the subgroup comparisons were limited; though the patient numbers were well above the suggested minimum level, and there was a minimum patient per therapist of 24.

Linked to resilience, and a precursor to the Green et al. (2014) and Pereira et al. (2017) studies, Saxon and Barkham (2012) suggested the impact of therapist 'burn-out' as a possible reason for the therapist effect, within their discussion, following analysis of 119 therapists and 10786 patients within a naturalistic setting, over an 8-year period. This study

investigated therapist effects across different levels of initial patient severity and risk and found therapist effects of 6.6% for patients with an average severity of symptoms, rising to 10% for higher severity patients. Overall, the outcomes for higher risk caseloads were poorer. In discussing these findings, the authors speculated that burn out, which has been linked to poorer patient outcomes (McCarthy & Freize, 1999), could be a factor in accounting for this, though this was not investigated within the study remit. If this was indeed a factor, then it is possible that therapist resilience could ameliorate the impact, linking to the findings of Green et al. (2014) and Pereira et al. (2017). Taken together, these studies may indicate that finding ways of creating less variability in the resilience levels, and/or mindfulness skills of therapists may lead to more consistent, and better, outcomes for patients across a therapist cohort. This study benefitted from large therapist and patient numbers within the sample studied.

 Therapist Distress. In a series of studies, Nissen-Lie, Havik et al. (2013), Nissen-Lie, Monsen et al. (2013) and Nissen-Lie et al. (2017) investigated a range of therapist factors that may be related to the therapeutic alliance. The therapeutic alliance has consistently been associated with good therapy outcomes, to the extent to which it has been suggested as the strongest predictor of patient outcomes (Beutler et al., 2003). Indeed, recommendations have been made that further studies into this association are only required where they focus on therapist variance in alliance (Beutler et al.,1994) and where this can be effectively partitioned accordingly (Baldwin & Imel, 2013). Nissen-Lie, Havik et al. (2013) used data from 70 therapists treating 227 patients to investigate whether therapist's stressors and satisfactions contributed to the variability in alliance as reported by patient and therapist. The study reported 18.2% of variance in longitudinal patient -rated alliance being ascribed to therapist differences and 24.2% of variance in therapist-rated alliance. Therapist-reported stress in their personal lives impacted patient -reported alliance levels negatively but had little effect on therapists' ratings of alliance. Therapists' reporting satisfying personal lives were self-rated as higher on alliance, but there was no significant effect on patient -rated alliance. The study's conclusion that therapists' personal distress leads to patients

experiencing a weaker therapy alliance, may have implications for possible ways to moderate this effect, such as increased therapist resilience. The outcomes reported suggest that therapists do not appear to recognise the impact of their personal distress on the alliance with their patients whilst patients are more able to identify alliance problems. This could suggest that improving reflective capacities, as discussed within the Cologon et al. (2017) study reviewed previously, may reduce the impact of these therapist factors on the alliance and, in turn, minimise differences in skill levels in this area, thus reducing the variability between therapists and improving clinical outcomes. Alternatively, developing ways of gaining and using feedback from patients on their experience of the alliance may prove useful, such as in the Zilcha-Mano and Errázuriz (2015) study (see below). Small numbers of patients per therapist was a limitation of the Nissen-Lie, Havik et al. (2013) study (mean 5 patients per therapist), as well as a low overall patient number. In addition, the dataset was taken from a multi-site (16 sites in total) outcome study which had a wide range of patient presentations and severities, a wide variety of therapist theoretical orientations and therapies offered, and a mixture of treatment durations, all of which may have impacted the identification of therapist effects and may make conclusions regarding use of the information gleaned from the study limited when applied to specific settings. The study did not directly investigate the impact of these factors on patient outcomes, so the relationship between alliance and patient outcomes is therefore assumed.

Professional Self-Doubt and Negative Personal Reactions. Additional explanatory therapist variables were explored by Nissen-Lie, Monsen et al. (2013) in a study generated from the same multisite, naturalistic dataset as the previous studies reviewed. This study utilised data from 70 therapists and 255 patients. Therapists completed a self-report questionnaire which included reliable subscales incorporating difficulties in practice (including professional self-doubt (PSD); warm interpersonal style; and advanced relational skills. Overall therapist effects on a broad outcome measure of patient functioning was 28%. On a measure of symptom distress the therapist effect was 4.2%, and on a measure of patient interpersonal problems it was 20.8%. The small number of patients overall, and few

patients per therapist, may have distorted the therapist effects found in this study. In relation to the therapist variables accounting for these effects, having a warm interpersonal style did not significantly impact patient outcomes on a measure of functioning, but advanced relational skills had a negative effect. This negative effect was not moderated by warm interpersonal style when this was modelled. Patients' interpersonal distress levels did not influence the association between advanced relational skills and patient functioning. No association was found between professional self-doubt and negative personal reaction scores, and patient outcomes on symptoms or functioning, though these were associated with patients' interpersonal problems outcomes. Professional self-doubt had a positive effect on patients' interpersonal problems, whereas a negative personal reaction had a negative effect. More distressed patients benefitted much less from therapists rating themselves as having advanced relational skills. These somewhat counterintuitive findings will be explored further in the interpersonal factors section below. However, it is interesting to note that therapists' professional self-doubt explained some of the positive outcomes for patients' interpersonal problems.

In a further development of this study, utilising a dataset comprising 70 therapists and 255 patients, Nissen-Lie et al. (2017) investigated the potential interaction between therapist self-reports of difficulties in practice in the form of: professional self-doubt (PSD); coping strategies when faced with difficulties; and more global, personal self-concepts to patient outcomes. Building on previous studies identifying the positive impact of therapist PSD on patient outcomes, this study sought to determine any relationship between PSD, therapist self-affiliative introject, and outcomes. The study reported a significant interaction between therapists' PSD and self-affiliation on patient outcomes of distress. Therapists with higher PSD seemed to evoke better outcomes if they also had a self-affiliative introject. However, those with lower PSD combined with high self-affiliative introject had the poorest outcomes. This led the authors to speculate that therapists with a healthy critical approach to their professional work, were most likely to produce positive clinical outcomes; but therapists who are self-compassionate without professional self-reflection and critique, would not yield

positive clinical outcomes. In addition to this finding, the study explored ways that therapists employed to cope with difficulties (via self-report), and how this impacted outcomes. Therapists who reported dealing actively with a clinical problem (e.g., seeking consultation, problem-solving with patient) appeared to have a positive impact on patients' levels of interpersonal distress. Therapists who avoided problems withdrew from therapeutic engagement, or acted out frustration in the therapy relationship had poorer outcomes on patients' symptom measures. The study faced the same limitations as previously stated for this multi-site dataset.

A follow-up and extension to the Nissen-Lie, Monsen et al. (2013), and Nissen-Lie et al. (2017) studies was undertaken by Odyniec et al. (2017) with a dataset comprising 40 qualified or trainee therapists and 621 patients. This study utilised state and trait-based measures of PSD and negative personal reaction (NPR) to explore any relationship between these factors and patient outcomes. The overall therapist effect was reported as 4.5%. Neither trait PSD nor trait NPR influenced patients' symptom change. However, both were relevant for an outcome measure of patients' interpersonal problems. Contrary to the Nissen-Lie, Monsen et al. (2013) and Nissen-Lie et al. (2017) studies, NPR was found to have a positive effect and PSD a negative effect. The authors hypothesised that this contradictory finding may be due to the therapists' stage in training (all either at an early stage or still in training), suggesting therapists may have different development needs at different career stages. Again, a small sample size was a limitation of this study.

Zeeck et al. (2012) explored the correlates of therapists' experience of 'stressful involvement' (SI) in therapy. The dataset comprised 26 therapists and 98 patients offering intensive psychodynamic treatment across two settings: an outpatient setting, and a day hospital setting. Several therapist scales and measures were used to assess experiences of SI and 'healing involvement' (HI). Overall, low levels of SI and high levels of HI were reported by therapists, though there was high between- and within-therapist variance on the SI measure. Working in the day hospital context was experienced as more stressful than working in an outpatient context. Therapists' emotional processing during the time between

sessions was a strong predictor of SI – therapists with high stressful involvement were more likely to be thinking negatively about cases during this time. Interestingly, no association was found between therapists' experience of SI and the severity level of the patients' symptoms, indicating that this may be more to do with the setting/context itself. Therapists that described themselves as submissive and socially avoidant in relationships were more prone to experience SI clinically. Although this study did not directly investigate therapist effects on patient outcomes, the findings do have relevance within the context of the previous studies investigating the impact of therapist anxiety, distress, and negative personal reactions on patient outcomes.

This line of research may have implications for the Phase 3 intervention package with indications that some level of professional reflection and doubt, coupled with a positive view of self and compassionate self-management of stressful involvement with patients, could all be aspects that vary across therapists, and that may impact patient outcomes. An intervention seeking to support therapists' ability to manage these internal experiences may therefore have positive results in relation to reduced therapist outcome variability.

Interpersonal Characteristics. These included aspects of the therapist that directly relate to the development and maintenance of relationships with others (e.g., therapeutic alliance, empathy).

Interpersonal/Relational Skills/Style. Laska et al. (2013) used a historical dataset of 192 veteran patients with posttraumatic stress disorder (PTSD) who were treated using Cognitive Processing Therapy by 25 therapists. Overall, using MLM, 12% of patient outcome variability was due to therapist effects. One supervisor provided supervision for the therapists treating patients within the dataset. The supervisor provided a retrospective rating of each therapist, based on their experience of them within supervision and these characteristics were compared between more and less effective therapists. The supervisor's ratings were statistically significant, with their ratings predicting therapists' effectiveness based on patient outcomes. Supervisor-identified characteristics and actions of effective therapists were: effectively addressing patient avoidance; language used in supervision;

flexible interpersonal style; and an ability to develop a strong therapeutic alliance. Small numbers of patients and therapists, as well as the subjectivity of using only one supervisor were limitations of this study.

As previously presented, Nissen-Lie, Monsen et al. (2013) reported results indicating that having a warm interpersonal style did not significantly impact patient outcomes on a measure of functioning, but advanced relational skills had a negative effect. This negative effect was not moderated by warm interpersonal style when this was modelled. Patients interpersonal distress levels did not influence the association between advanced relational skills and patient functioning. More distressed patients benefitted much less from therapy from therapists rating themselves as having advanced relational skills, and this was not moderated by having a 'warm interpersonal style'.

Therapist Negative Personal Reactions to / Stressful Involvement with Patients. As previously presented, Zeeck et al. (2012) explored the correlates of therapists' experience of 'stressful involvement' (SI) with patients in therapy. Therapists' emotional processing during the time between sessions was a strong predictor of SI. Those with high stressful involvement were more likely to be thinking negatively about their patients during this time. Therapists that described themselves as submissive and socially avoidant in relationships were more prone to experience SI clinically. In a potentially related dimension, and as previously discussed, Odyniec et al. (2017) explored any relationship between therapists' state and trait negative personal reaction (NPR) and patient outcomes. They found that although NPR did not influence patients' symptom change, it was a factor positively related to improvements on a measure of patients' interpersonal problems.

Development of the Alliance. In a study based on data from a University Counselling Centre, Slone and Owen (2015) utilised data from 247 patients treated by 49 therapists. They investigated therapist alliance activity, the quality of the systemic alliance, therapist comfort and any relationship between these factors and patient outcomes. Alliance activity was rated by the patient using descriptions of behaviours theorised as being involved in the development of therapeutic alliances, (e.g., the therapist asking for feedback about the

therapy relationship). The quality of the alliance was also patient-rated, and in addition patients completed a measure of how comfortable they felt their therapist to be within sessions. Improvement in a patient's life functioning was measured by self-report, alongside a retrospective measure of pre-therapy functioning. Overall, using MLM, therapist effects accounted for 7% of variance in outcomes within the study, though the small sample size should be noted when interpreting therapist effect size in this study. Therapist effects on patient ratings of alliance activity was 19%, with positive alliance activities and therapists who appeared more at ease, predicting better patient outcomes. An interesting nonsignificant interaction was found between alliance activity and therapist activity. The authors tentatively hypothesised that this may indicate the positive impact of alliance activity may be present despite less perceived therapist comfort. Generalisability is relatively low for this study which was based in a university counselling centre.

Although relevant to the context of the therapist, but discussed later, a study by Zilcha-Mano and Errázuriz (2015) also focused on the alliance, outcomes, and therapist variability. This study had a sample size of 547 patients and 28 therapists and focused on patient and therapist factors that may moderate the strength of the therapy alliance and outcome association. The study also involved the comparison of five feedback conditions, involving the provision of feedback to therapists on the strength of the alliance, as rated by the patient. Using MLM this study was unusual in the review in that it found no overall significant therapist effect on outcomes (patient effects of 65% were found). However, of interest for this review, was the effect of giving therapists' their patient alliance ratings on outcome. This indicated that there was a greater relationship between alliance and outcome, in conditions where therapists were given feedback on the alliance.

From the studies summarised within the category of 'Interpersonal Characteristics' there are indications that variability in therapists' micro-skills as they relate to the development of the therapeutic alliance and interpersonal style may be contributing to patient outcomes. Given the nature of the therapeutic alliance being common across therapies, and this well-established association between alliance and outcome, including a

focus on honing these micro-skills across a service as part of the Phase 3 package could potentially yield an impact on therapist level variability in outcomes.

Context of the Therapist. This category included studies relating to the context that the therapist experiences when carrying out their work including the development of professional skills (e.g., supervision framework, knowledge/skills acquisition, support structures).

Supervision. Four studies investigated or included supervision as a factor that may have a bearing on therapist variance and outcomes. As previously discussed, Laska et al. (2013) focused on supervisor ratings to explain variance in therapist outcomes. One of the areas identified as being associated with effective therapists, was the language used within supervision. For more effective therapists, their language tended to indicate a non-defensive attitude and an openness to talking about struggles and difficulties. They also discussed attempts to use suggestions provided to them by others. Limitations of this study have been explored previously.

In a naturalistic study, Ali et al. (2014) investigated variability of outcomes across 38 therapists treating 1376 patients. Three-level MLM analysis found just 0-1.3% of variability was accounted for by therapist effects, with between patient variability accounting for most of the variance. The fact that this study used a three-level model (sessions, patients, therapists) may account for the finding of such a low therapist effect. However, the authors also hypothesised that the low therapist effect may have been due to the nature and structure of the context of these therapists. The therapists were 'low intensity' therapists, who provided brief, manualised, guided self-help; had the same training, and high levels of clinical case supervision with a specific and focused structure. However, it should be noted that a larger therapist effect was found by both Green et al. (2014) and Firth et al. (2015) when studying low intensity therapists in a similar context. In addition, there were low levels of variability in therapist experience within the therapist participant cohort. In relation to training and number of sessions attended, this study may also indicate that frequent, highly

structured supervision may have an impact on reducing the variance of outcomes across therapists (i.e., reducing therapist effects).

The results from the Ali et al. (2014) study were in contrast to a study by Green et al. (2014) which also investigated therapist effects across a cohort of low intensity therapists (also known as psychological wellbeing practitioners) but found a significant therapist effect on outcome variance of 9%. (Nb. This study was presented previously within Intrapersonal Factors). In the group of more effective therapists several specific, supervision-related factors were found to be present, in contrast to the less effective therapist group. Supervisors reported that these more effective therapists tended to be more proactive in their approach to learning and supervision, and also tended to be more prepared and organised in supervision. Less effective therapists tended to be less confident in using supervision. It is perhaps, therefore, not the frequency or structure of the supervision, but rather how supervision is used and approached, that may make a difference in accounting for some of the therapist variance though this has not been studied directly.

In a study investigating variance between supervisors and linking this with patient outcomes, Rousmaniere et al. (2016) utilised a historical naturalistic dataset of 6521 patients treated by 175 trainee therapists, supervised by 23 supervisors. Utilising HLM, less than 1% of the variance in outcomes could be explained by supervisor effects, but again, it is important to note the three-levels to this MLM. Variation at the supervisor level would be harder to measure due to the impact of patient and therapist factors that may obscure it. For example, this study found that therapist factors were significant in this sample, so it is possible that a highly effective therapist could compensate for poor supervision, and a less effective therapist could skew outcomes for a good supervisor. However, although therapist effects were significant, they were small (0.84% of variance accounted for by therapist effects). It is possible that this study did not use a large enough sample to be sensitive enough to identify supervisor effects at this third level. The study did not have a control group who were not receiving supervision, so it was not possible to conclude that supervision did not influence outcomes. Indeed, it was equally possible that the impact of

supervision may have been similarly positive across supervisors in this sample. However, the authors explored other variables which may account for the lack of supervisor effect found, and these are discussed below.

Use of Feedback. Three studies included direct study or discussion of the use of feedback for clinicians. Zilcha-Mano and Errázuriz (2015) included the use of five feedback conditions involving the provision of feedback to therapists on the strength of the alliance, as rated by the patient within their study. This study has been presented previously, and of note was the evidence that in conditions where therapists were given feedback on the alliance, there was a greater relationship between alliance and outcome for the patient. Also, as previously discussed, Rousmaniere at al. (2016) explained the lack of significant supervisor effect in their study by suggesting that, as all therapists were utilising session-by-session feedback data within their work, the positive effect of this may have confounded the impact of any potential supervisor effect.

Errázuriz and Zilcha-Mano (2018) investigated several different feedback conditions to identify any relationships with outcomes. This randomised control trial comprised a sample size of 547 patients and 28 therapists. Patients were randomly assigned to one of five therapist feedback conditions: no feedback; feedback on symptoms; feedback on alliance; feedback on both; or use of a standardised feedback system. Analysis used MLM to identify effects of the different conditions on several specific variables. They found that feedback condition had no significant effect on outcomes, session attendance, or alliance. However, there were moderators such as the patient's mental health history and initial severity, combined with lack of progress. For the different patient groups, "positive feedback" to the therapist (i.e., providing feedback on the low symptom severity of the patient), had a positive impact on outcomes; whilst "negative feedback" (i.e., providing feedback on the patients' high symptom severity) had the reverse effect. Overall, the study did not find significant therapist effects on patient outcomes. However, there was a therapist effect of 5.49% accounting for the variance in the percentage of sessions attended by patients. There

were also therapist effects in the variance of therapeutic alliance, with 4.14% of the variance in the alliance being attributable to therapist effects.

Activities to Deliberately Improve Practice. In a study previously presented, Green et al. (2014) suggested that regular use of feedback to clinicians of their patients' outcome measures, was a means of potentially improving those outcomes. Within the group of more effective therapists studied, there was evidence that this group was more likely to actively set out to learn from observing the practice of others, and to seek out more CPD opportunities than their less effective colleagues.

Chow et al. (2015) utilised a sub-set of patients from a full study comprising 4580 patients and 69 therapists. Their subset had a sample size of 1632 patients and 17 therapists. Therapists were practising as part of the Human Givens Institute Practice Research Network in the UK, offering therapy to adults with anxiety and depression. Using MLM, the full study found therapist effects accounted for 5.1% of the variance in outcomes, after adjusting for initial severity. This smaller study focused on the impact of deliberate activities to improve therapists' practice on their outcomes. Therapist characteristics (i.e., years of experience, gender, age, profession, highest qualification, caseload) did not significantly predict outcomes. However, the amount of time therapists spent on activities specifically focused on improving therapeutic skills was a significant predictor of outcomes. Interestingly, no specific activity was related to outcomes, suggesting perhaps that the specific activity needed to improve outcomes may differ between therapists, according to individual skills and deficits. Alternatively, it may be that the amount of time spent on activities to improve their practice reflects the attitude that the therapist has towards their therapy work, and that this itself has an impact on their outcomes. Highly effective therapists reported that reviewing therapy recordings required more cognitive effort than did their poorer performing colleagues. It was suggested by the authors that activities requiring higher mental effort may result in a higher likelihood of learning taking place. In discussion, the authors suggested that therapists should target the following areas to improve outcomes: focusing on at-risk cases; creating their own social experiments in naturalistic settings to test

and improve their empathic accuracy; targeting learning of fundamental therapeutic skills such as rehearsing difficult conversations; using standardised case vignettes to improve interactions with patients; and setting aside time to reflect on their sessions and to plan forthcoming sessions. Limitations of this study include the small number of therapists within the subset, and the use of retrospective self-report of the specific activities undertaken to improve practice. The use of deliberate practice as a means of improving therapist outcomes was also recommended within the discussions of two studies of therapist experience, outlined below (Erekson et al., 2017; Goldberg, Rousmaniere et al., 2016).

Training Level and Experience. Generally, studies that included years of experience (normally number of years since qualifying), found that this was not a significant explanatory variable for therapist effects (Ali et al., 2015; Chow et al., 2015). However, there were two studies that found a negative association between therapist effectiveness and either training level or years of experience (Goldberg, Rousmaniere et al., 2016; Erekson et al., 2017). There was also one study that investigated the stability of therapist effectiveness over time, focusing on domains of skill, which found no significant association with years of experience within the analysis (Kraus et al., 2016).

Kraus et al. (2016) included 3540 patients and 59 therapists in a naturalistic setting within their dataset. Random forest models were used to risk-adjust outcomes based on case-mix variables, and HLM was used for the main analysis, utilising the classification of confidence intervals for therapists. Treatment outcomes were measured covering 12 outcome domains. The outcomes of the first 30 patients of each therapist were used to classify therapists' effectiveness on these domains, with the next 30 patients used to compare the predictive validity of the outcome domains. The study findings demonstrated that therapist effectiveness was relatively stable, but somewhat domain specific. However, there was no specific domain in which more effective therapists appeared to be stronger. The results appeared to demonstrate that more effective therapists may be effective across more domains than their less effective colleagues. Therapists identified as 'exceptional' were more likely to remain above average with subsequent patients, suggesting that past therapist

performance was a good indicator of their future performance. There was no significant difference between years of experience identified within the analysis. The authors suggested identifying individual therapists' areas of deficit and focusing on improving these as a way of improving patient outcomes.

Goldberg, Rousmaniere et al. (2016), using MLM, analysed 6591 patients across 170 therapists in a large scale, longitudinal study. The study findings suggested that outcomes were slightly worse as therapist experience increased. However, early termination was lower as therapist experience increased, and this was the case when controlling for patient -level, case-load and therapist level effects. There was significant variability in therapists' trajectories of improvement over time, with some therapists obtaining better outcomes over time and some getting worse. Although significant, the therapist variability in rate of change over time accounted for relatively little patient outcome variance. In discussing these findings, the authors suggested that it was the content of experience that may be more relevant than the amount of experience. They suggested therapists should gain focused experience based on specific skills with direct feedback on this experience, similar to the principles of deliberate practice.

Erekson et al. (2017) investigated the impact of therapist training on patient outcomes. With a sample size of 4047 patients and 22 therapists, HLM was used to analyse therapist effects on rate of change and overall patient outcomes, adding therapist training level as a potential explanatory variable. The results showed that 0.75% of variance in outcomes was attributable to therapist effects. There was a significant effect for initial severity and total number of sessions on the rate of change patients demonstrated. A higher level of severity or greater number of total sessions were both associated with a slower rate of change on patient outcome measures. Therapists had faster rates of change at graduate level and slower rates at licensed level (licensed level being a later stage than graduate level). In terms of therapists' improvements in practice over time, therapists with licenced training level either showed no, or worse, improvement in patient outcomes compared to graduate level therapists. The authors suggested that there is increased variability in

therapists as they progress through their training and recommended the use of specific methods of improving therapists' practice, including the use of deliberate practice.

      Length of Therapy. Several studies included the length of therapy as a variable related to therapist effects and patient outcomes. Goldberg, Hoyt et al. (2016) investigated the treatment length for high and low performing therapists with a naturalistic historical dataset of 5828 patients and 158 therapists. Using MLM, a therapist effect of slightly less than 1% was identified. When accounting for treatment length overall, patients who remained in treatment longer had poorer outcomes. But interestingly, there was significantly more variance between therapists as the treatment length increased, with patients of more effective therapists experiencing significantly better outcomes than patients of less effective therapists, when at the 16-session duration. The outcomes of patients of more effective therapists increased with treatment length. These differences were negligible at the three or four session duration. Thus, the authors argued, a seemingly small therapist effect overall does not reflect the clinical significance of therapist variance for patients' outcomes when engaged in a full course of therapy. This study utilised data from a single counselling centre, so may lack generalisability, and a single self-report measure of patient outcome was used which is a further limitation of this study.

      With similar findings to Goldberg, Hoyt et al. (2016), Saxon, Firth et al. (2017) found that the greater the number of attended sessions, the greater the variability across therapists in their patients' outcomes. In this naturalistic study, a sample of 4034 patients treated by 61 therapists was analysed using MLM, identifying more and less effective therapists whilst controlling for case mix. The study reported that 5.8% of the variance in patient outcomes was due to therapist effects. Overall, higher numbers of sessions improved outcomes, whilst non-completion of therapy reduced the amount of pre-post change significantly. Greater therapist variability was apparent with higher number of sessions, but there was little variability when patients attended fewer sessions overall. For patients who attended 8 sessions, the above average therapists were over twice as effective as the below average therapists. For patients who attended 12 sessions, above average therapists increased to

being three times more effective than below average therapists. Recommendations made by the authors included a focus on reducing patient early non-completion of therapy through common alliance factors and addressing therapy ruptures. The study did not include any therapist variables, such as alliance measures, so it was not possible to hypothesise the reason or variables related to early non-completion within the study. Erekson et al. (2017), discussed previously, demonstrated that a patient's rate of change slowed as the number of sessions increased. Firth et al. (2015) found that there was greater therapist variability in therapy up to six sessions.

Nature of Caseload. Saxon and Barkham (2012) investigated factors impacting therapist variability in a study with a large sample comprising 119 therapists and 10,786 patients. MLM was used to analyse the dataset, with a therapist effect size of 6.6% reported for patients whose symptom severity (as measured on a self-report outcome measure) was average. The impact of therapist effects was larger for patients whose initial severity was greater, and patients treated by therapists with a higher overall risk caseload (i.e., a caseload with more high-risk patients) had worse outcomes than those with a non-risk caseload. This may be linked to the findings previously described in relation to therapist resilience or distress. It would certainly have implications for services in monitoring therapists' caseload, to balance the risk-level of therapist caseloads overall. However, therapist effects were not accounted for by risk or patient severity, causing the authors to call for more research into the additional factors that may be responsible for these therapist effects.

Firth et al. (2015) utilised a sample of 6111 patients treated by 56 low intensity therapists to explore therapist effects on outcome and potential moderators of the therapist effect. MLM was used to analyse the naturally occurring historical dataset, and a therapist effect of 6-7% was observed. This effect was found to be moderated by patients' initial symptom severity, the duration of treatment (greater variation in therapy up to 6 sessions), and patient drop-out. More effective therapists tended to have a bigger overall caseload, and

it was hypothesised by the authors therefore that more practice may result in more effective therapists. However, this was not empirically investigated within the study.

Patient Ethnicity/Domain Specific Skills. Hayes et al. (2015) identified therapist effects of 8.7% in a sample of 36 therapists and 228 patients. Therapists were trainees in a University training clinic providing a range of non-specified counselling interventions. The patient's race/ethnicity accounted for 19.1% of this variance, when included as a patient level variable within MLM analysis. This study was limited by small numbers of patients and therapists, and the dataset also demonstrated poor clinical outcomes overall, with approximately two thirds of the patients in the study failing to demonstrate symptom reduction. In terms of the purpose of the literature review, the study suggests that some therapists may have particular strengths in working with particular patient groups, in this case patients from ethnic minorities, providing the potential for skills sharing to improve the outcomes of therapists who do not have these strengths, thus reducing the gaps in effectiveness between therapists. However, such potential would depend on these therapists being identified.

The idea of some therapists being more effective across specific patient outcome domains was explored by Nissen-Lie et al. (2016) in two datasets: one with a sample size of 5828 patients and 158 therapists; the second a smaller replication study comprising 520 patients and 31 therapists. In the primary study, a therapist effect of just under 2% was observed, while in the replication study this effect dropped to approximately 1%. The authors went on to investigate the variability in outcomes across the domains of functional improvement measured in both datasets. These domains were identified as: wellbeing, anxiety, depression, close relationships, general functioning, and social relationships. The analyses in both studies concluded that effectiveness was reliable across domains; therefore, therapists that were effective, or ineffective, in one domain were likewise so across all domains. This would suggest that effective therapists are not necessarily strong at producing outcomes for a particular domain, but instead are able to effect change uniformly for their patients. The authors hypothesised that this finding may be due to effective

therapists being more flexible in the adaptation of their approach to suit the individual patient's needs. They suggest this may also, therefore, relate to a sensitivity and responsiveness to the patient's needs, hypothesising that these common factors may be more relevant than specific techniques, when it comes to effectiveness. However, these suggestions were not empirically tested within the scope of these studies.

In terms of their relevance for this literature review, Nissen-Lie et al. (2016) suggest that their findings may support previous literature indicating that improvements in the effectiveness of therapists could be realised by: the use of sessional feedback (allowing adjustments to treatment to be made in response to patient improvement or deterioration); improved skills in developing alliances with a wide range of patients; focusing on interpersonal skills that facilitate patient engagement; and encouraging therapists to acknowledge and discuss their professional doubt and challenges. The authors recognised that studies to investigate the effect of implementing these development areas would be required to provide robust empirical support, or indeed to refute such suggestions.

Contrary to Nissen-Lie et al.'s (2016) findings, Owen et al. (2016) also explored the therapist effect on different patient outcome dimensions, within a study of rates of change across different outcomes. They found varying therapist effects across the dimensions. It is worth noting, however, that the measures used were different to those included in the Nissen-Lie et al. (2016) study, which all related to life functioning. Owen et al. (2016) utilised a large, naturalistic sample of 13664 patients and 586 therapists. Data was gathered from a range of therapy sites, and most settings were University Counselling Centres. MLM was used to identify therapist effects on the different outcome dimensions. This analysis identified therapist effects of 0.4% for a measure of overall wellbeing, 4.6% for a measure of symptom change, and 7.5% for a measure of general life functioning. The authors concluded that variation among therapists is likely best detected when the patient outcome is relatively more complex. They suggested that the ability to effect deeper life functioning of patients may be a more advanced ability, leading to the higher variability in this outcome measure compared to the other outcome measures across therapists.

Summary and Conclusions

In exploring the variables that may explain therapist effects, three main areas emerged within the literature: intrapersonal characteristics, interpersonal characteristics, and the context of the therapist. These are summarised in Table 2.3 chronologically within sub-categories. The methods of studying the potential explanatory variables for therapist effects appears to be in its infancy, with only four studies directly using therapist effects analysis to investigate explanatory variables of these effects (Chow et al., 2015; Green et al., 2014; Laska et al., 2013; Pereira et al., 2017).

Table 2.3

Summary of Potential Explanatory Variables from Current Review

Explanatory Variable	Study
INTRAPERSONAL CHARACTERISTICS	
Therapist Attachment & Reflective Functioning	Cologon et al. (2017)
Therapists Resilience (& mindfulness)	Saxon & Barkham (2012)
	Green et al. (2014)
	Pereira et al. (2017)
Therapist Distress	Nissen-Lie, Havik et al. (2013)
	Nisen-Lie et al. (2017)
Professional self-doubt & negative personal reactions	Zeeck et al. (2012)
	Nissen-Lie, Monsen et al. (2013)
	Nissen-Lie et al. (2017)
	Odyniec et al. (2017)
INTERPERSONAL CHARACTERISTICS	
Interpersonal/Relational Skills/Style	Laska et al. (2013)
	Nissen-Lie, Monsen et al. (2013)
Therapist negative personal reactions to / stressful involvement with patients	Zeeck et al. (2012)
	Odyniec et al. (2017)
Development of the alliance	Slone & Owen (2015)
	Zilcha-Mano & Errázuriz (2015)

CONTEXT OF THE THERAPIST

Supervision	Laska et al. (2013)
	Ali et al. (2014)
	Green et al. (2014)
	Rousmaniere et al. (2016)
Use of feedback	Zilcha-Mano & Errázuriz (2015)
	Rousmaniere et al. (2016)
	Errázuriz & Zilcha-Mano (2018)
Activities to deliberately improve practice	Green et al, (2014)
	Chow et al. (2015)
	(Goldberg, Rousmaniere et al., 2016)
	(Erekson et al., 2017)
Training Level and experience	Ali et al. (2014) non-significant
	Chow et al. (2015) non-significant
	Goldberg, Rousmaniere et al. (2016)
	Kraus et al. (2016)
	Erekson et al. (2017)
Length of Therapy	Goldberg, Hoyt et al. (2016)
	Erekson et al. (2017)
	Saxon, Firth et al. (2017)
Nature of caseload	Saxon & Barkham (2012)
	Firth et al. (2015)
Patient ethnicity/domain specific skills	Hayes et al. (2015)
	Nissen-Lie et al. (2016)
	Owen et al. (2016)

Many areas included evidence in the form of recommendations or suggestions from authors rather than robust empirical support for a specific variable in practice. The number of contradictory findings was of interest, perhaps not surprising for a field in its infancy in terms of both areas of focus, and methods of study. For example, studies found greater therapist variability in outcomes both for more sessions (Goldberg, Rousmaniere et al., 2016, Saxon, Firth et al., 2017) and fewer sessions (Firth et al., 2015). However, variation in the types of

therapies, therapists and service contexts examined within the literature is relevant here. For example, the Firth et al. (2015) study utilised low intensity therapists, who are trained to work briefly and would not be expected to offer more than 6-8 sessions. The therapists/therapies in the Goldberg, Rousmaniere et al. (2016) and Saxon, Firth et al. (2017) studies were longer term, up to 16 sessions.

Studies that investigated intrapersonal characteristics of the therapist were both broad both in scope and on the method of studying therapist effects. Two studies investigated resilience, and one of these also studied mindfulness as an additional explanatory factor. More effective therapists appeared to have higher resilience (Green et al., 2014), and higher mindfulness and resilience combined (Pereira et al., 2017). However, there were contradictory findings in that resilience alone did not account for therapist effects in the Pereira et al. (2017) study. Potentially linked to resilience, studies of stress and anxiety in therapists suggested that reduced stress is related to more effective therapists, and that reducing anxiety may be more helpful in terms of outcomes than increasing positive experiences for therapists (Nissen-Lie, Monsen et al., 2013). It is also important to note, however, that stressful involvement with patients appeared to be a function of the therapist and not correlated with the severity level of the patient (Zeeck et al., 2012). This again may suggest that therapists with higher resilience may be more likely to be effective and less impacted by high severity/distressed patients. In other words, more resilient therapists may be effective with a range of severity of patients, making them more effective overall. Less effective (and less resilient) therapists may become increasingly less effective with the higher severity patients. Consistent with this, it has been observed that there is greater therapist variability the higher the risk of the therapists' caseload (Saxon & Barkham, 2012). In considering ways to reduce this variability in outcomes, based on the findings above, therapist resilience may provide a potential moderator of the therapist effect.

Good reflective skills were associated with more effective therapists, and a secure attachment style of the therapist moderated the impact of poor reflective skills (Cologon et al., 2017). Studies investigating therapist stress and the therapeutic alliance, demonstrated

that patients treated by therapists reporting higher stress gave lower ratings on the strength of the alliance; whereas the therapists themselves did not. If combined with the results from Cologon et al. (2017), it could be hypothesised that increased self-reflection and the use of therapist ratings of alliance, may have a beneficial impact on therapist outcomes. Including these within an intervention package may improve the self-reflection of those therapists who are more limited in this area, thus equalising the impact it may have on patient outcomes between therapists. Indeed, lack of insight/self-reflection is one of the possible explanations for the counter-intuitive results of Nissen-Lie, Monsen et al. (2013) indicating that an advanced relational style is associated with poorer patient alliance. In this study, assessment of relational style was measured by therapist self-report; which contrasts with the positive findings of studies where relational style was externally assessed (Laska et al., 2013). If, indeed, those therapists who rated themselves as having advanced relational skills were lacking in self-reflection (or accurate reflection about the therapy alliance), this may account for the findings. This would be a consistent conclusion in light of the studies that investigated therapist professional self-doubt (PSD) and negative personal reactions to patients. Here, there were contradictory findings of PSD being associated with both more (Nissen-Lie, Monsen et al., 2013, Nissen-Lie et al., 2017) and less (Odyniec et al., 2017) effective therapists. Due to the stage in training of the Odyniec et al. (2017) study, however, it may be possible to conclude that a combination of robust professional self-doubt and self-compassion are elements helpful to the effective therapist.

Other than the contradictory results in the Nissen-Lie, Monsen et al. (2013) study of self-reported therapist interpersonal skills, generally level of interpersonal skills were an explanatory factor for therapist variability on outcomes (Laska et al., 2013). This may follow, given that shorter therapies necessitate a more rapid alliance over a shorter period arguably requiring higher levels of skills to develop and manage effectively. However, this finding is not consistent with studies in this review reporting greater therapist effects as a function of the number of sessions (i.e., more sessions led to greater therapist variability; Goldberg, Hoyt et al., 2016; Saxon, Firth et al., 2017).

Activities that focused on developing and maintaining the alliance (as rated by the patient) were found to have a positive impact on patient outcomes, alongside therapist comfort (Slone & Owen, 2015). In a study with a slightly different focus (i.e., feedback conditions), it was also interesting to note that the relationship between alliance and outcomes was greater for patients being treated by a therapist in the feedback conditions (i.e., being given feedback about the patient's perception of the alliance) (Zilcha-Mano & Errázuriz, 2015). This supports the idea that the alliance, as rated by the patient, is important in relation to outcomes, and may provide useful feedback to therapists during the process of therapy. This could potentially provide the opportunity for insight for those therapists that may struggle more with this skill, raising their skill level to that of their more reflective counterparts. There was one other study in the review that investigated feedback conditions and therapist effects but which found non-significant therapist effects and effect of feedback conditions on patient outcomes (Errázuriz & Zilcha-Mano, 2018). Feedback to therapists may be an important method by which therapists could be given accurate data/intelligence about the impact they are having on their patients' outcomes (and perception of alliance, as this relates to outcome). Indeed, this has been recommended in other studies of therapist effects (Green et al., 2014; Rousmaniere et al., 2016).

Supervision is an element of psychotherapy practice that has been well-documented, though not in the therapist effects literature until the last decade. Within this review, effective therapists attended the same amount of supervision as less effective therapists, but appeared to use supervision differently. They demonstrated a non-defensive attitude to exploring difficulties (Laska et al., 2013) and a proactive approach, involving good preparation and organisation levels (Green et al., 2014). The exploration of supervisor effects, adding a third level at which effects on patient outcomes may be seen, are not yet being conducted widely. One study was included in this review, with a non-significant effect size, potentially due to the small numbers of supervisors involved (Rousmaniere et al., 2016). Linked to this, more effective therapists were also likely to seek out opportunities to observe others and undertake CPD (Green et al., 2014). The amount of time therapists

spent deliberately practicing specific therapy-related skills was found to be predictive of patient outcomes (Chow et al., 2015). However, no single specific activity was associated with better outcomes. The role of deliberate practice in improving therapists' effectiveness was also recommended within four other studies within the review, though not empirically tested (Erekson et al., 2017; Firth et al., 2015; Goldberg, Rousmaniere et al., 2016; Nissen-Lie et al., 2016).

In order to bring about more consistent therapist outcomes, indications from the above line of enquiry suggest that guidance about how to better use supervision, opportunities to observe peers, undertake training, and practice specific skills may all potentially yield benefits. Given the complex nature of variability, and the potentially differing needs of therapists in terms of areas of deficit and skill, an approach such as this that provides a framework for the development of therapist-specific skills would appear to have applicability in the current context.

Therapist years of experience is an explanatory variable that has been tested within most therapist effects studies due to the ease with which it can be included. It has generally had non-significant results (Ali et al., 2014; Chow et al., 2015; Kraus et al., 2016) or controversially, has been related to poorer outcomes as a function of experience/training level (Erekson et al., 2017; Goldberg, Rousmaniere et al., 2016). However, there is also evidence within this review, that therapist effectiveness is stable over time, and there is no one domain in which effective therapists are particularly strong. Rather, they may demonstrate effectiveness over multiple domains (Kraus et al., 2016).

It is significant to note that there is evidence that the therapist effect increases as a function of therapy session number. Greater therapist variability in patient outcomes when patients attend eight or more sessions suggests that an overall therapist effect, if using a naturalistic dataset, will be significantly smaller than the therapist effect for the patients who have completed a full therapy treatment within the same dataset (Goldberg, Hoyt et al., 2016; Saxon, Firth et al., 2017). This adds further importance to the need for greater understanding and research into the field of therapist effects.

Study and Review Limitations

Most studies included in the review were limited by sample size, either by numbers of included therapists (ranging from 17 to 586), or numbers of patients per therapist (ranging from 1-455), but more often by both. Maas and Hox (2005) suggest that a sample size of over 100 at level 2 is required for accurate multi-level modelling, though they suggest a minimum of 50 as being acceptable. Schiefele et al. (2017) recommends a minimum of 1200 patients to reduce bias in estimations when detecting therapist effects. Only 16 of the studies reviewed met this minimum level of therapists, and 15 met the minimum patients. Just 12 out of the 28 studies met both minimum sample sizes at patient and therapist level.

Most of the 28 studies were based on naturalistic datasets, providing a high level of external validity and generalisability to 'real life' therapeutic practice settings. However, this design also reduces the controllability and transparency of patient allocation, adherence to therapeutic modalities, content of supervision provided and patient presentation/case mix. Where there were multi-site studies, there were also differences between sites/services that may have impacted on results, as this variable was not controlled for within analysis of these studies.

The present review sought to focus on therapist effect studies as they relate to possible explanatory or mediating factors of/for therapist effects. The premise of this review was to focus on factors that appear to operate at the therapist level (i.e., where there is variability between therapists, not accounted for by patient factors), and where there is an indication that the factor may have an impact on patient outcomes. By so doing, the review aimed to highlight areas of focus for an intervention package that could be provided to all therapists, with the objective of showing whether more consistent therapist outcomes (for patients) could be obtained across the service, whilst improving absolute patient outcomes. Due to the lack of robust studies explicitly comparing groups of therapists and the mediating factors for their outcomes, this limited the search to include studies that had minimal empirical testing of the recommendations or suggestions made that might explain the therapist effects identified. As a method of generating ideas from existing research, this

review has been helpful; but again, it highlights the infancy of our understanding or identification of the mechanism or content of the seemingly intangible 'therapist effect'.

Implications for Phase 3 Intervention Package

In considering how this literature review may inform a therapist-level bespoke intervention package that is practical in a service-based research environment, the following candidates appeared most strongly from the literature as potential areas of interest:

- Consideration of the content and quality of supervision rather than simply the quantity

- The structured use of patient feedback to therapists about their progress, possibly in relation to their view of the therapy alliance, as well as symptom outcomes, to aid self-reflection/reflective practice

- The use of deliberate practice of therapy skills and tasks, as needed by the individual therapist based on their strengths and deficits

- The use of the above elements within supervision to cultivate an atmosphere of compassionate critique of therapists' work

- Interventions that seek to increase therapist wellbeing and resilience

These areas were identified and discussed with therapists within the service in order to collaboratively refine the intervention package, and this stage of the process is presented in Chapter 6.

Research Questions

The findings from this pragmatic but systematic literature review, together with the driver for the research, (namely, seeking to improve the consistency of outcomes between therapists across a service environment), gave rise to the following research questions:

1. Can a package of therapist-focused, bespoke, research-based actions be implemented to reduce therapist variability and improve patient outcomes in routine practice, when controlled for and analysed across a stable group of therapists over time?

2. Can such a package yield a reduction in therapist variability and improved patient outcomes, when analysed at a service-wide level (i.e., accommodating the natural turnover of therapists within a service)?

Chapter 3

Study Context and Phase 1 Baseline

Overview

This chapter sets out the historical context and service setting in which the three study Phases reported in this thesis took place, with specific reference to the national programme of service delivery. It then specifies the research questions and aims, before providing details of the components of the service comprising the Baseline Phase (Phase 1), which is the first of the three study Phases. The chapter concludes with a brief outline of the benefits to the research of having the primary researcher embedded within the service leadership team during all three Phases of this longitudinal study. Ethical approval for Phase 1 and 2 of the overall study was granted by the East of England branch of the Health Research Authority (Ref: 17/EE/0251; see Appendix C).

Study Setting Context

The following sections set out the context for the study Phases. This includes information about the history and policy related to the type of service where the study was set.

History and Development of Improving Access to Psychological Therapies Services

The setting for the three study Phases was a large Improving Access to Psychological Therapies (IAPT) service within the National Health Service (NHS) in the UK. A series of reviews of the evidence for psychological therapies was undertaken by the National Institute for Health and Care Excellence (NICE) commencing in 2004, resulting in new guidance for the treatment of Panic Disorder (with or without agoraphobia) and Generalised Anxiety Disorder (2004a), Depression (2004b, 2009a, 2009b), Post-traumatic Stress Disorder, (2005a), and Obsessive Compulsive Disorder (2005b). Guidance on the use of computerised CBT for treating Depression and Anxiety (2006), and on pathways of care for 'common mental health disorders' (2011) were also developed following on from these reviews. The reviews provided guidance regarding the use of evidence-based

psychological therapies, and the implementation of these treatments was presented in a 'stepped care' model.

A stepped care service delivery model is defined by differing treatment components being available at different specified levels of intensity or 'steps'. It is a model of health care delivery utilised when there is a significant gap between demand for a treatment and supply of that treatment (Bower & Gilbody, 2005). The two key features of stepped care are: (1) offering a first step treatment that is the least restrictive or intensive whilst still providing sufficient clinical health gain; and (2) having a method of self-correction. In terms of treatment intensity or restriction, this can refer both to patient cost and inconvenience, but also, and with more relevance to this context, the amount of specialist therapist time required. This ensures a more economic service delivery, where only patients requiring higher levels of therapist involvement are provided with this more costly option. Self-correction within the system relates to procedures where health benefits are reviewed and monitored, and patients can be stepped up to treatment further up the stepped care system if they are not responding to lower step options (Newman, 2000) as well as being stepped down if appropriate. An example of the stepped care approach for depression is shown in Figure 3.1.

Figure 3.1

Stepped Care Treatment for Depression (Adapted From NICE, 2009b)
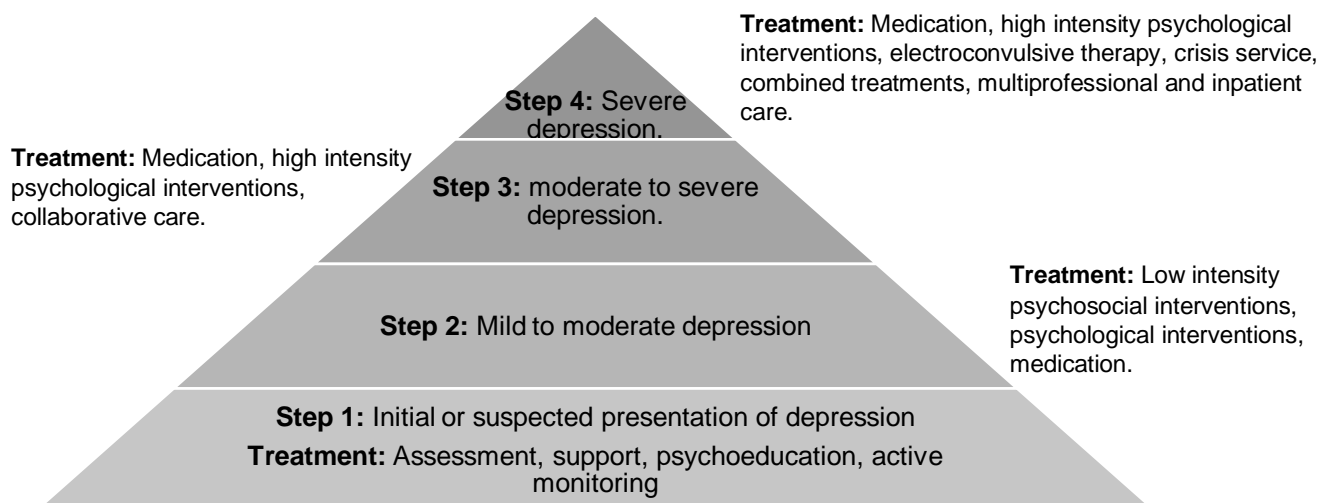
The NICE guidelines laid the clinical groundwork for the development of economic modelling which sought to make the financial case for the benefits of increased access to evidence-based treatment for common mental health problems. This work resulted in the refinement of an economic argument presented to the 10 Downing Street Strategy Unit meeting by Lord Layard and David Clark in 2005. The paper proposed building on existing mental health service improvements and evidence-based treatments, and focusing on further mental health service improvement to enable: patient choice as to a range of evidence-based psychological treatments; the implementation of local targets to reduce waiting times for both primary and specialist mental health services; ease of access to self-help facilities; a focus on returning and retaining adults in work; and tackling mental health stigma in schools and workplaces (Layard, 2005). In addition to the clinical argument for this commitment (supported by the NICE reviews), this was underpinned by a rationalised economic incentive reflecting the impact that common mental health problems have on peoples' ability to work, and therefore the ability for societies to fund other key health treatments (Layard & Clark, 2014, Layard et al., 2007). This argument laid out the benefits of providing evidence-based psychological therapies to those with mild to moderately severe common mental health problems both in relation to improving overall health, but also reducing sickness rates and unemployment - thus reducing costs to the national benefits system and increases to taxpayer contributions overall (Layard et al., 2007). To substantiate this, Layard and Clark set out a striking cost comparison: the cost to the Exchequer of an individual being on 'incapacity benefits' due to ill health was estimated at £750 per month (in 2007); however, the cost of a course of NICE-compliant psychological treatment was estimated at £750 *in total*. This calculation provided a strong incentive for the investment in 8,000 extra psychological therapists that was proposed at that time. The economic conclusion to the argument stated that based on their estimate, within 2-5 years *"the cost to the government would be fully covered by the savings in incapacity benefits and extra taxes that result from more people being able to work"* (p. 1). Based on these arguments, the UK government committed to increasing the availability of psychological therapies to people with

common mental health problems (i.e., depression and anxiety disorders). The initiation of two pilot sites (termed 'demonstration' sites) saw the development of two types of service model aiming to meet the objectives set out by Layard and Clark in 2005.

The demonstration sites, based in two differing geographical areas (Doncaster and Newham, England), were developed and implemented with some key features, including: maintaining a team approach to delivering evidence-based psychological therapies in a stepped care structure; out-reach to primary care settings; strong clinical leadership; best practice training, supervision and support for clinicians; inclusion of employment and housing advice within the team; and substantial data collection on clinical outcomes (via sessional measures), work and employment outcomes, and other key metrics.

Findings from the Demonstration Sites. The two demonstration sites were reviewed by a team from the Universities of Sheffield, Manchester and Southampton and covered the pilot period from June 2006 to April 2008 (when the pilot phase finished), as well as a further 12-month follow up period (Parry et al., 2011). The key findings from the demonstration sites that translated into the early IAPT service model, are summarised below.

Use of Stepped Care. The NICE guidelines present a stepped care approach to offering interventions for common mental health problems, with the least intensive evidence-based intervention offered based on the severity level of the presenting problem (for an example, see Figure. 3.1). This model of care has been applicable to the treatment of psychological therapy primarily due to the development of a range of differing intensity levels by which CBT can be provided. This has the bonus of providing consistency between steps as a result of these treatments sharing the same theoretical background and therapeutic techniques, whilst differing in intensity (Bower & Gilbody, 2005). This treatment might be in the form of guided self-help, with minimal therapist involvement, often delivered over the telephone (Richards et al., 2002); computerised-CBT (Proudfoot et al., 2004) with or without therapist feedback; or more traditional one-to-one 'high intensity' CBT sessions. However, the addition of other psychological therapy options that can be 'stepped up' to (such as

person-centred counselling or Interpersonal Psychotherapy), allows for the use of

therapeutic approaches that may be more effective in a case where a CBT approach has not

been successful (Newman, 2000; Ward et al., 2000). Both demonstration sites utilised a

stepped care model in their service delivery, reducing the amount of therapist input required

overall, and therefore making treating large numbers of patients both financially and clinically

effective, and operationally practical.

The Doncaster service model placed more emphasis on the use of guided self-help,

or low intensity CBT approaches, based on a pre-existing service model and a practical

limitation in the availability of high intensity therapists (Parry et al., 2011; Richards &

Suckling, 2008). This resulted in relatively few patients accessing a step up into high

intensity psychological therapies (just 5.2% accessed Step 3 therapy for their initial

treatment step, with a further 7% being stepped up after initially accessing Step 2 therapy

[Parry et al., 2011]). In terms of a pure stepped care model, this model therefore breached

the principle of self-correction, outlined above, as there was limited ability to increase the

intensity level of the treatment in response to patient progress/outcomes. However, the

model did provide the opportunity to test the viability, and patient acceptability, of widely

available, high volume, low intensity treatments in practice.

The Newham stepped care service model had an emphasis on brief interventions,

but initially saw these delivered by high intensity trained clinicians, as well as psychologists.

This was based on the existing service in the area, which had developed a primary care brief

psychological therapies service with clear pathways into more specialist mental health

treatments. The new model looked to expand this delivery into GP practices, improving the

accessibility of these treatments in the community. In addition, the service model aimed to

develop culturally appropriate self-help materials based on the population diversity, as well

as the use of large group-based psychoeducation based on CBT principles. An expansion of

step 2 interventions was implemented six months after the initial pilot began, utilising

assistant therapists, and introducing a telephone screening assessment undertaken by a

qualified therapist, directing the patient to the appropriate level of intervention. This change

was requested by the Department of Health due to the high costs per patient associated with the initial model, and aimed to utilise the Doncaster step 2 model but with greater ability to step up to higher intensity therapies. This change to the service model saw an increase in access to step 2 interventions as the initial treatment, from 7% to 66%; and an 8-fold increase in patient throughput per month (Parry et al., 2011). This represented a move to a purer stepped care approach and provided greater reach within the community.

Use of Self-Referral. The pilots sought to destigmatise the act of seeking mental health treatment by ensuring ease of access to treatment via self-referral, rather than via a GP or primary care clinician. It was anticipated that self-referral would greatly increase the number of people accessing the demonstration sites. Interestingly, however, the Doncaster model saw 90% of referrals originating from GPs across the pilot period. However, Newham achieved a significant increase in the number of self-referrals being received, particularly over the second half of the pilot, with 23% of referrals overall being self-referral, and 65% of GP referrals (Parry et al., 2011). The difference in demographics of the sites, with Newham having an ethnically diverse population (39% White compared to 88% for England as a whole), led to a focus in Newham directed towards increasing access from underrepresented groups, which likely accounts for this difference in access routes (Parry et al., 2011). The Newham site demonstrated that self-referrals more accurately reflect the ethnic mix of the population in comparison to GP referrals which tended to underrepresent ethnic minority groups (Clark, 2018).

Use of Sessional Outcome Measures. The level of investment in the demonstration sites, and the potential investment being proposed nationally for psychological therapies, relied upon both the effectiveness of the interventions, and their translation from RCTs to being provided in a real-world service context, where variables are more difficult to control. It was, therefore, imperative to the project that clinical outcomes were reliably monitored and reported to ensure the services provided the clinical improvements necessary to allow individuals to return to the workplace (where possible). The IAPT initiative settled on a Minimum Data Set (MDS) set out in the IAPT Implementation Plan (Department of Health

[DH], 2008a) and IAPT Outcomes Toolkit 2008/9 (DH, 2008b) and set a target of 90% data completeness across services. The MDS comprised patient demographic data (gender, ethnic category, date of birth), care pathway data (date of referral receipt, referral source, primary diagnosis, date of initial assessment and treatment appointments, first and last clinical outcome measures, reason for end of care pathway) and contact level, or session data. This sessional data included: therapist details, appointment date and purpose, interventions provided, current use of psychotropic medication, current step, employment status and a range of clinical outcome measures.

The recommended clinical outcome measures within the IAPT minimum dataset (i.e. required data to be collected and reported [DH, 2008b]), or patient-reported outcome measures (PROMs) were condition specific and comprised a measure of depression (PHQ-9, Kroenke et al., 2001; Spitzer et al., 1999), generalised anxiety disorder (GAD-7, Spitzer et al., 2006), anxiety-related avoidance (Phobia Scale, [DH, 2008b]), functional impairment (Work & Social Adjustment Scale, Marks, 1986; Mundt et al., 2002) and a fluctuating range of other disorder specific measures (these other measures were not used for measuring recovery until 2012). These core outcome measures (PHQ-9, GAD-7, WSAS & Phobia Score) are described as being chosen as *"the most suitable, free to access tools, and are widely used in practice"* (p.13, DH, 2008b); though they had limited use in the UK prior to the demonstration sites (Parry et al. 2011), and the Phobia Score had no prior use. Use of the PHQ-9 and GAD-7 was emphasised and used in isolation for calculating 'Recovery' and Reliable Improvement' rates in the National dataset until 2012 (DH, 2012a). From the outset, the IAPT initiative aimed to administer these measures at each session, no matter the level of therapeutic intensity, to be used *"by patients and IAPT workers to provide tangible evidence of treatment progression, and by supervisors to review clinical work. It is used by managers to facilitate effective service performance, and by service commissioners and others to demonstrate the direct return on the investment made in services, benchmarked against clear performance measures"* (p. 6-7, DH, 2011a). Detail about the structure and content of these measures is included in Chapter 6: Methods.

Routine sessional outcome monitoring allowed a more reliable evaluation of service effectiveness on an ongoing basis, ensuring that even those that terminated therapy early were included in outcome data, as the most recent score could be utilised as their post-treatment score. More traditional 'pre-post' score evaluations can have a bias toward overestimating treatment effectiveness, due to the loss of 'drop out' cases that do not have matched pre-post scores (Clark, 2018; Layard & Clark, 2014). However, neither approach addresses the outcomes of those who attend one session only, and these patients are excluded from both pre-post and sessionally-based clinical outcome analysis. Indeed, in IAPT definitions of recovery and reliable improvement, only those who have attended two or more *treatment* sessions are included, meaning those that attend one assessment and one treatment are also removed from the outcome calculations (DH, 2011a). Despite these limitations, this introduction of the use of routine sessional outcome monitoring ensured services could be better and more accurately evaluated as to the clinical effectiveness of the interventions their therapists were providing to the relevant patient population on an ongoing basis.

Use of sessional measures also allowed services to ensure that patients were receiving the appropriate intensity or 'step' treatment, by determining clinical 'caseness' and symptom severity (DH, 2008b). This was primarily determined through clinical caseness cut-off scores on the PHQ-9 (10 or above) and GAD-7 (8 and above). However, it was noted that a set of anxiety disorder specific measures (ADSMs) should be used to identify and measure the severity of other anxiety disorders. These disorder specific measures have changed over the course of the IAPT implementation and development (see Table 3.1).

Robust Clinical Outcomes. The IAPT initiative specified definitions of clinical Recovery and Reliable Improvement which are used to measure the clinical outcomes of services across England, this data being submitted to NHS Digital on a monthly basis. Monthly outcome reports of the performance of each IAPT service in the country, searchable by commissioning area, are reported quarterly, and available via the NHS Digital website (www.digital.nhs.uk).

Table 3.1

Disorder Specific Outcome Measures Used in IAPT 2008 - 2021.

Disorder	Recommended Measure 2008 (IAPT Outcomes Toolkit)	Recommended Measure 2021 (IAPT Manual v.5)
Social Phobia	Social Phobia Inventory (SPIN)	Social Phobia Inventory (SPIN)
Body Dysmorphic Disorder	None Specified	Body Image Questionnaire (BIQ) Weekly
Obsessive Compulsive Disorder	Obsessive Compulsive Inventory (OCI)	Obsessive Compulsive Inventory (OCI)
Post-traumatic stress disorder (PTSD)	Impact of Events Scale – revised (IES-R)	PTSD Checklist for DSM-5 (PCL-5)
Health Anxiety	Health Anxiety Inventory (HAI)	Health Anxiety Inventory (HAI)
Panic/Agoraphobia	Agoraphobia Mobility Inventory	Panic Disorder Severity Scale OR Agoraphobia Mobility Inventory
Phobia	Fear Questionnaire	'Phobia' No longer listed
Anger	None specified	'Anger' No longer listed
Generalised Anxiety Disorder	Generalised Anxiety Disorder -7 (GAD-7) & Penn State Worry	Generalised Anxiety Disorder -7 (GAD-7)
Depression	Patient Health Questionnaire-9 (PHQ-9)	Patient Health Questionnaire-9 (PHQ-9)

The current definitions of Recovery and Reliable Improvement are shown in Table 3.2. In calculating Recovery and Reliable Improvement, all patients who have attended a minimum of two treatment sessions are included, no matter the ending of therapy (i.e., including those who terminate therapy early, drop out of treatment, or are referred on to other services).

Table 3.2

IAPT Outcome Definitions (NHS England [NHSE] & National Collaborating Centre for Mental Health [NCCMH], 2021)

Recovery & Reliable Improvement	Definition
PHQ-9 Caseness	PHQ-9 >9
GAD-7 Caseness	GAD-7 >7
Moved to Recovery	A patient in caseness on *either* PHQ-9 *or* GAD-7/ADSM at first session, and moving to below caseness on PHQ-9 *and* GAD-7/ADSM at final session.
Recovery rate calculation	$$\frac{\text{Number of referrals that 'moved to recovery'}}{\begin{bmatrix}\text{Number of referrals that finished a course of treatment} & - & \text{Number of referrals that finished a course of treatment and started treatment not at caseness}\end{bmatrix}} \times 100$$
Reliable Improvement	Improvement of >5 points on PHQ-9 from intake to final score OR Improvement of >3 points on GAD-7 from intake to final score AND where neither PHQ-9 or GAD-7 score deteriorates from intake to final score.

A patient is deemed to have 'recovered', if either one or both of their initial PHQ-9 and GAD-7 (or other ADSM) scores are in caseness, and both final PHQ-9 and GAD-7 (or other ADSM) scores are not in caseness (NHSE & NCCMH, 2021). 'Caseness' refers to having scores on clinical measures indicating the presence of depression or an anxiety disorder, and the clinical caseness cut offs for the current IAPT measures are shown in Table 3.3. Reliable improvement measures movement on an outcome measure of an amount that indicates statistically reliable improvement (Evans et al., 1998; Jacobson & Truax, 1991).

Table 3.3

Anxiety Disorder Specific Measure Caseness Cut Offs (NHSE & NCCMH, 2021)

Anxiety Disorder Specific Measure	Clinical cut off (Referral is at caseness if it meets the following criteria)
Agoraphobia-Mobility Inventory (AMI)	≥2.3
Body Image Questionnaire (BIQ) Weekly	≥40
Generalised Anxiety Disorder Assessment (GAD-7)	≥8
Health Anxiety Inventory (HAI) (Short Week)	≥18
Obsessive-Compulsive Inventory (OCI)	≥40
Panic Disorder Severity Scale (PDSS)	≥8
PTSD Checklist for DSM-5 (PCL-5)	≥32
Social Phobia Inventory (SPIN)	≥19

Using a sample matching the stipulations for measuring IAPT recovery and reliable improvement, a recovery rate of 45.8% was achieved in Doncaster and 43.2% in Newham. Reliable improvement rates were 40.4% in Doncaster and 36.0% in Newham (Parry et al., 2008). In evaluating these outcomes across the two sites, large improvements were observed with pre-post treatment uncontrolled effect sizes of 0.98-1.17 (Parry et al., 2008). This compares favourably to studies of both low and high intensity CBT (Hofmann et al., 2012; Richards & Suckling, 2009). The outcomes observed in Doncaster and Newham demonstrated that clinical outcomes in practice could replicate the outcomes observed in RCTs, when specific conditions were met (such as standardised training, treatment delivery

and supervision). This was the first time that data using sessional measures provided evidence of clinical effectiveness, whilst addressing one of the potential issues of bias in completion rates (Clark, 2018, Layard & Clark, 2014).

Following the evaluation and success of the demonstration sites, £309 million was subsequently made available to fund the development of IAPT services across England, and by 2011 all areas of England had a functioning IAPT service. As a government initiative with stringent national reporting targets, guidelines for the operation of IAPT services were initially provided in the IAPT Implementation Plan (DH, 2008a). Following numerous revisions and additional appendices over the years, a comprehensive IAPT delivery manual was first published as *The Improving Access to Psychological Therapies Manual* ten years later, in June 2018 (NHSE & NCCMH, 2018). This describes not only the principles behind IAPT services, but details the therapies to be delivered, the workforce requirements for delivery, methods for measuring outcomes, and the performance target definitions. This document was developed through learning from services over the first ten years of the IAPT programme, following the achievement of the national target set at the outset for 50% of patients completing therapy within IAPT, to reach clinical 'recovery' (Clark, 2018; Layard & Clark, 2014; DH, 2012b). The manual is, therefore, presented as a blueprint for services to follow to achieve this target at a local level.

In the latest version of the manual (NHSE & NCCMH, 2021), IAPT services are described as being characterised by:

- Offering evidence-based psychological therapies, provided at the appropriate 'dose'

- Employing an appropriately trained and supervised workforce

- Utilising routine sessional outcome monitoring

IAPT Service Clinical Delivery

As stated earlier, IAPT services offer psychological interventions and therapies within a stepped care model of care delivery. At step 2, patients are offered low intensity brief psychological interventions based on CBT. Clinicians offering these interventions

(Psychological Wellbeing Practitioners [PWPs]) are trained to post-graduate certificate level in low intensity psychological interventions (Bennett-Levy et al., 2010; Richards & Whyte, 2011;), involving rigorous teaching, role play assessment and in-service elements to their training. The low intensity curriculum is nationally mandated, as with all IAPT training, to ensure that an equivalent standard and quality of therapy is offered across services (Bennett-Levy et al., 2010; DH & University College London [UCL], 2015). The training includes: engagement and assessment of patients with common mental health problems; evidence-based low intensity treatments for common mental health problems (such as behavioural activation, graded exposure, cognitive restructuring, panic management, problem solving, sleep management, supporting physical activity and medication adherence); values; diversity and context. In the year commencing April 2019 to March 2020, in excess of 458,000 patients completed a Step 2 treatment (NHS Digital, 2020). This compares to the year commencing April 2009 to March 2010 when just in excess of 92,000 patients completed treatment (Step 2 or Step 3) in total (DH, 2012b).

At Step 3, patients can access a variety of evidence-based psychological therapies depending on need and availability: Cognitive Behavioural Therapy (CBT); Person-Centred Experiential Counselling for Depression (PCE-CfD); Interpersonal Psychotherapy (IPT); Eye Movement Desensitisation and Reprocessing (EMDR); Couples Counselling for Depression (CCfD); Dynamic Interpersonal Therapy (DIT) (NHSE & NCCMH, 2021). Step 3 therapists are expected to have completed an IAPT compliant training which includes the minimum mandatory competencies. Curricula and competency frameworks have been developed by University College London (UCL) over the development of the IAPT program, and now cover all of the therapies listed above (Roth & Pilling, 2008; https://www.ucl.ac.uk/pals/research/clinical-educational-and-health-psychology/research-groups/core/competence-frameworks). Services are expected to offer therapies in line with NICE guidelines, including dosage level (i.e., number of sessions of therapy offered).

As high-volume services, supervision levels are intensive in IAPT services, with recommendations for staff to be provided with weekly clinical case management supervision

in addition to clinical skills groups at Step 2, and more traditional clinical supervision at Step 3 (NHSE & NCCMH, 2021). Supervisors and clinicians are encouraged to utilise patient outcome data within case management supervision, though guidance on how this is used has been less clear over the various guidance documents (e.g., IAPT Manual, [NHSE & NCCMH, 2018, 2021]; Reach Out Manual, [Richards & Whyte, 2011]). In the context of stringent training, supervision and operational recommendations for IAPT services, a high level of control, and therefore consistency of treatment offered and outcomes achieved, is aimed for across the programme nationally (Clark, 2018; Layard & Clark, 2014).

An analysis of the outcomes of 24 services involved in the first year of the roll out of IAPT demonstrated significant clinical improvements for patients treated in the services, though highlighted considerable variability between services (Gyani et al., 2013). This analysis used reliable recovery instead of IAPT recovery, a metric that combines the recovery and reliable improvement calculations and was applied to IAPT data for the first time in 2011 (Parry et al., 2011). This more robust measure of clinical change was incorporated into IAPT national reporting later in the development of the program (NHS Digital, 2014) but continues to be excluded from the core IAPT target metrics. This delay in applying well established procedures in calculating clinically significant change, accounting for whether the change is greater than the measurement error of the questionnaires (Evans et al., 1998; Jacobson & Truax, 1991; Parry et al., 2011), is notable in the IAPT guidance literature, and indeed appears not to be fully explained in existing manuals. Failure to apply this metric means that a patient could improve by just one point (moving from caseness to non-caseness) and be deemed as clinically recovered. Reliable recovery means that as well as meeting the recovery definition, the patient must also meet the reliable improvement definition. Based on this reliable recovery, the study calculated that 40.3% of the patients included in the sample achieved this metric, but this ranged from 23.09% to 56.6% between services. A total of 63.7% of patients demonstrated reliable improvement, again with a wide range between services (43.6% to 77.1%).

This study also reported the observation that those with higher scores on the PHQ-9 or GAD-7 experienced greater improvement on these measures, challenging assumptions that those with greater severity are less likely to benefit from therapeutic treatments. Other notable observations were: services that utilised the 'step up' option within the stepped care model had higher rates of reliable recovery; services that offered a higher number of sessions on average were more likely to recover than services that offered fewer sessions; patients treated in services with a higher proportion of qualified grade therapists were more likely to recover than those in services with a smaller proportion; and larger services (i.e., those treating larger numbers of patients) appeared to perform better – the greater the number of patients treated in the service, the more likely patients were to recover.

By 2012, a review of the first three years of the IAPT program outlined the key achievements of the initiative, including having treated more than 1 million patients, achieving a 90% data compliance rate, recovery rates of 45% and almost 4,000 new therapists having been trained under the program (DH, 2012b). The initial investment in IAPT has continued over multiple changes of government, with cross-party support ensuring that the key components of the IAPT program have been included in influential government policy documents over these years (e.g., No Health Without Mental Health, [DH, 2011a]; Five Year Forward View for Mental Health, [NHSE, 2016]), and into the latest publication The NHS Long Term Plan (NHSE, 2019).

Due to the significance of this investment in psychological therapies, and the findings from the two demonstration sites, multiple session by session patient outcome measures continue to be used routinely within IAPT services, making them the most highly measured mental health services in the NHS (Gyani et al., 2013). Clinical outcomes have continued to improve over the years since the first 3 years' review, with quarterly recovery outcomes shown in Figure 3.2.

Reflecting on 10 years of IAPT services (in 2018), Clark highlighted the key achievements of the program, with over 10,500 therapists trained within IAPT services, 93%

of services offering a choice of psychological therapy treatment, and 99% outcome data completeness for treatment episodes (Clark, 2018).

Figure 3.2

National IAPT Recovery Rates 2012 - 2021
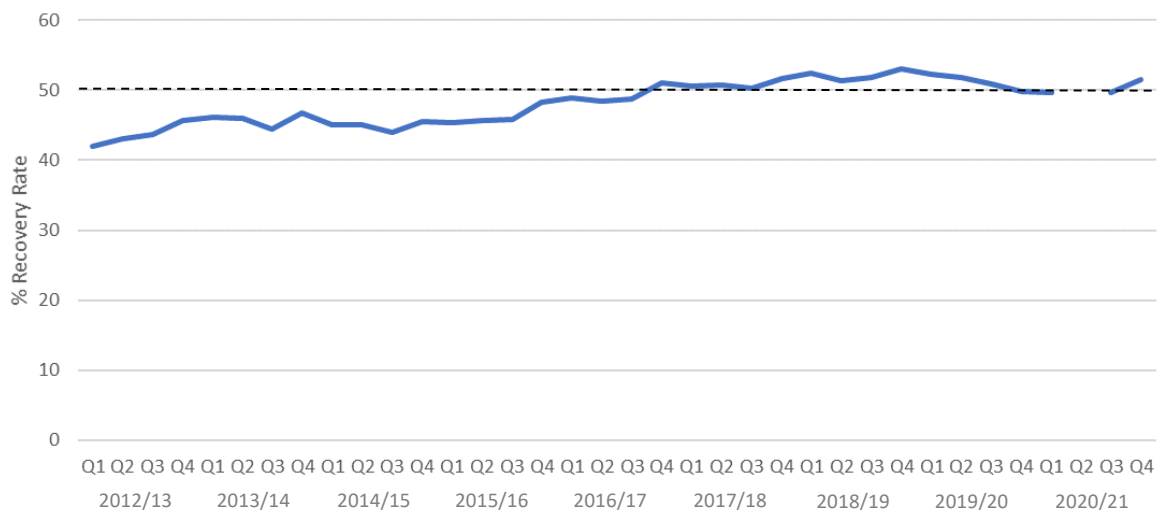
Services nationally were accessed by over 1 million people per year in 2019, and over half of those accessed a course of treatment as a result (NHS Digital, 2020). Outcomes, as shown in Fig 3.2, demonstrated that over 50% of those treated experienced clinical recovery, and approximately two thirds experienced clinically reliable improvement, matching the outcomes seen in the RCTs as a result of which the IAPT initiative was developed (Clark, 2018). Alongside these successes, as highlighted above, there remain some limitations to the IAPT approach to outcome measurement. The removal of patients who do not attend the minimum number of treatment sessions (2 sessions minimum) means that there is poor visibility of the outcomes of these patients. To put this in context, in 2019-20, 1,647,716 IAPT referrals 'ended', and 1,126,404 of these patients attended one or more appointments with an IAPT practitioner. A total of 606,192 of these patients completed treatment. However, 489,547 only attended one treatment session (in addition to any assessment sessions undertaken prior), and an additional 30,665 patients attended assessment appointments only (NHS Digital, 2020). This leaves patients who attend one

treatment session (and any number of assessment sessions), comprising 43.5% of completed referrals who attended IAPT sessions in the period, but whose outcomes are not included in the IAPT outcome calculations. The delay to incorporating reliable and clinically significant recovery calculations into the IAPT outcome metrics has been mentioned previously, and although this metric is now reported in IAPT annual reports, there is no mandatory KPI related to reliable recovery for services to achieve, (unlike with recovery and reliable improvement). Finally, the variation in outcomes between services that has been noted within IAPT reporting (e.g., IAPT August 2021 reported recovery rates range from 12% to 94%, [NHS Digital, 2021]) has not translated to exploration of within service variation by the IAPT programme, for example through the investigation of therapist effects, despite evidence of this in other studies using IAPT data (e.g., Firth et al., 2015, 2019; Saxon, Barkham et al., 2017; Saxon, Firth et al., 2017).

IAPT as Context for the Development of Practice-Based Evidence and Therapist Effects Research

In modern health settings there is an increasing acknowledgement of the need for an interactive model of research and practice, ensuring the outcomes of efficacy research underpins evidence-based practice (such as in IAPT services for example); but also, that effectiveness and practice research generates practice-based evidence, to further the field of psychotherapy in context (Castonguay et al., 2021; Barkham & Mellor Clark, 2003). We can conclude from RCTs that therapy is effective, with or without the control of therapist variability; and increasing evidence pointing to the impact of therapist variability on outcomes in practice suggests that further practice-based research needs to be conducted to identify some of the features of effective therapists (Firth et al., 2015; Okiishi et al., 2003; Saxon & Barkham, 2012, Wampold & Owen, 2021). Practice research tends to evaluate how to improve treatment or service delivery, and due to the samples involved being naturally occurring, has high external validity. It allows analysis of data within a service, investigating differences and variations, not only in patient subgroups, but between therapists (Okiishi et al. 2003). The IAPT initiative has shown how, in practice, the ability to control patient

variables is limited (compared to RCT settings) and therefore a focus on how to maximise therapist factors that contribute to improved patient outcomes would appear to be a valuable objective.

As discussed, IAPT services have been developed, in a large part, as a direct response to the increasing evidence base for the effectiveness of psychological therapies and in response to the difficulties and delays in people accessing such treatments (Clark, 2018). In this way, IAPT has served as a context in which to apply a highly manualised, highly controlled, 'RCT-like' service design, with the intention of replicating in practice, the outcomes seen in RCTs (Layard & Clark, 2014). Within a naturalistic setting, RCT conditions might be seen as a 'top down' service design, with strict mandates in relation to training, supervision, and assessment and therapy application (NHSE & NCCMH, 2021). The emphasis of the application of evidence solely based on RCTs within mental health and psychotherapy has historically been viewed as controversial within applied settings, most importantly due to the difficulty in predicting individual outcomes from these studies using data based on group means (Margison et al., 2000). But it is also due to issues of lower control over the evidence-base adherence of therapists in routine practice (Waller, 2009). Within RCTs, in order to control for therapist and patient factors, stringently applied manualised therapy is provided by a small number of highly supervised clinicians to a narrow population of patients, which has generally been viewed as having limited generalizability to routine practice (Castonguay et al., 2021). Outcomes from the IAPT programme demonstrate that, despite limited control over patient variables, it is possible to match the outcomes seen in RCTs using similar principles of top-down, highly directive service design and application (Layard & Clark, 2014). However, alongside the development and further refinement of IAPT service design over the years (discussed above), IAPT services increasingly offer the context for a form of practice-based research, allowing evidence developed through a 'bottom-up' process to be realised (e.g., Castonguay et al., 2013, 2021; Delgadillo et al., 2016). In addition, due to the potential sample sizes and controls on therapist variables, IAPT services also offer an ideal context in which to study therapist

effects, providing an opportunity to better understand therapist variability, and importantly, what might support the development of more effective therapists.

As discussed in Chapter 2, there is reasonable evidence indicating there are limited differences between therapies (Barkham & Lambert, 2021; Lambert, 2003), and indicating that the best predictor of outcome may be more related to what the therapist as an individual brings to the 'treatment' (Stiles et al., 1998). From studies that have been undertaken (both pre-dating, and including IAPT service data), significant variance has been found between therapists in relation to patient outcomes (Firth et al., 2015; Okiishi et al., 2003; Saxon & Barkham, 2012; Wampold & Owen, 2021). Indeed, some have argued that studies comparing therapies, when not accounting for the therapist as a variable itself, may result in erroneous conclusions that particular treatment modalities account for differences in outcome when they may not (Crits-Christoph & Mintz, 1991; Shapiro et al., 1989; Wampold & Owen, 2021). Viewed through the lens of the IAPT programme, this could have significant implications for the investment of public money in specific therapy modalities and therapy training priorities. Increasing evidence of the variability of therapists' outcomes irrespective of modality (Firth et al., 2015; Okiishi et al., 2003; Saxon & Barkham, 2012) has further reinforced a view that "s*tudies of effective versus ineffective therapists may be a useful way to pursue an understanding of how psychotherapy works and how to best train therapists to be successful*" (Crits-Christoph & Mintz, 1991, p. 25).

Study Aims

The systematic review of the therapist effect literature, including potential explanatory factors for therapist effects, identified a number of factors potentially related to therapist effects (see Chapter 2 and Appendix A). These factors were identified as warranting further exploration, establishing how using these factors in a therapist cohort might improve therapist outcomes and influence the therapist effect in practice.

Design Overview

The overarching approach adopted for the three study Phases comprised a longitudinal observational study consistent with the practice-based evidence paradigm,

located in a large psychological therapies service in a part of England, UK that comprises

large rural areas and a small number of urban centres. This service was part of the

Improving Access to Psychological Therapies initiative. The longitudinal study comprised

three 18-month Phases, resulting in a total study period of 4.5 years (see Figure 3.3). Phase

1 (Baseline Phase) comprised a retrospective analysis of 18 months of data, yielding both

the baseline patient outcomes and the therapist effect in the service during a period of

standard, albeit underperforming, practice. Phase 2 (Directive Phase) comprised data

collected during a period of top-down service change, led by the management team and

guided by NHS Improvement. Phase 3 (Collaborative Phase) involved the implementation of

a specific intervention package with the purpose of improving therapist skills and staff

wellbeing, introduced collaboratively with staff and implemented with an intentionally staff-led

approach. Further details regarding the development of the two intervention Phases (i.e.

Phase 2 and 3) are presented in Chapters 4 and 5.

Figure 3.3

Study Timeline
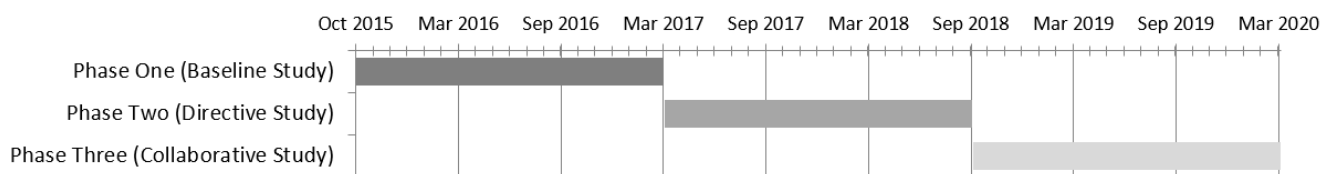
Although the primary comparisons were between the two intervention Phases, it was

necessary to understand the baseline therapist effect within the service prior to any

interventions being carried out. The following section describes the specific conditions or

service context during the 18-month period of data collection for Phase 1 (Baseline Phase).

Service Condition in Phase 1 (Oct 2015 – Mar 2017): Baseline Phase

The purpose of the Baseline Phase was to determine the variability in therapist effect

over time in a naturalistic setting, testing the impact of two subsequent sequential

intervention Phases – a Directive Phase and a Collaborative Phase – on the size of the

therapist effect and the clinical outcomes. The aim was to provide proof of concept, to a)

assess whether the variability in therapist effect could be reduced in the latter two Phases,

thereby producing improved and more consistent outcomes (i.e., less variability) for patients accessing psychological therapies services, and b) to increase understanding of *how* services can better support therapists, with a view to the potential for implementation of any useful intervention across multiple services nationally, reducing therapist variance whilst improving outcomes overall.

Phase 1 examines the therapist effects and patient outcomes present in the NHS IAPT service during a period of 'service as usual' as a comparator for the following Phase of service improvements. It provides the baseline therapist effects from retrospective data extraction and analysis, with no significant changes occurring within the service delivery. The service during Phase 1 was experiencing lower overall clinical performance than desired, which provided the context for the initiation of the overall study. The clinical recovery outcomes for the services' five teams during Phase 1, with England as a comparator, are shown in Figure 3.4, demonstrating the below average outcomes across all five teams within the service, particularly during February to November 2016. This poor performance generated the initiation of Phase 2 and Phase 3 interventions (Chapters 4 and 5, respectively), aimed at changing the conditions of the service to improve the outcomes for patients.

Figure 3.4

Service Recovery Rates Oct 2015-Apr 2017 (Data source Internal Service Data)
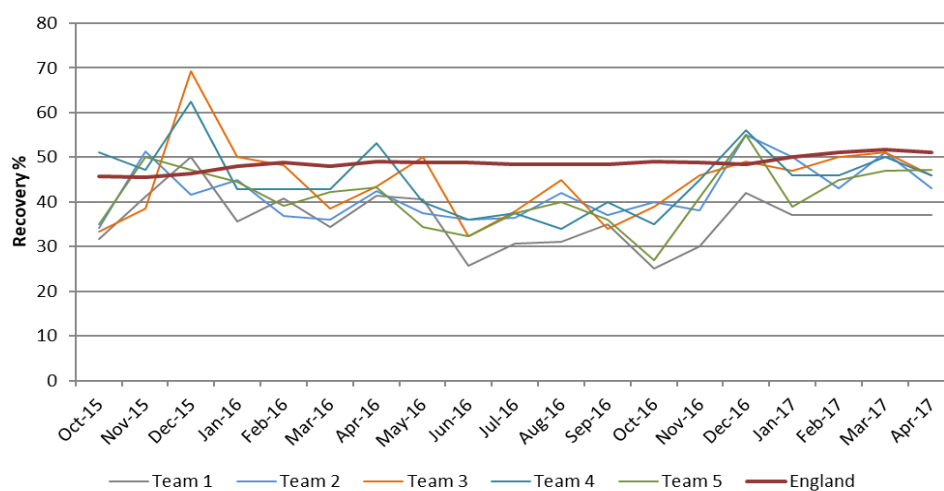
The recovery and reliable improvement outcomes for the service, on a three-month rolling basis, compared with the National rates, from January 2016 to January 2017, are shown in Figures 3.5 and 3.6, respectively.

Figure 3.5

Service Overall Three-Month Rolling Recovery Rate Jan 2016 – Jan 2017 (Data source NHS Digital Monthly IAPT Data)
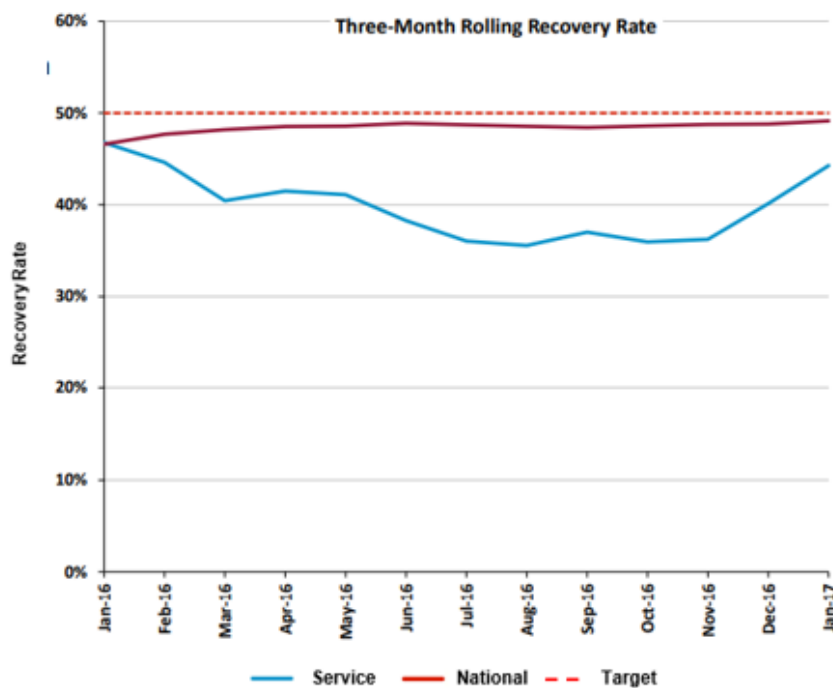
Figure 3.6

Service Overall Three-Month Rolling Reliable Improvement Rate Jan 2016 – Jan 2017 (Data source NHS Digital Monthly IAPT Data)
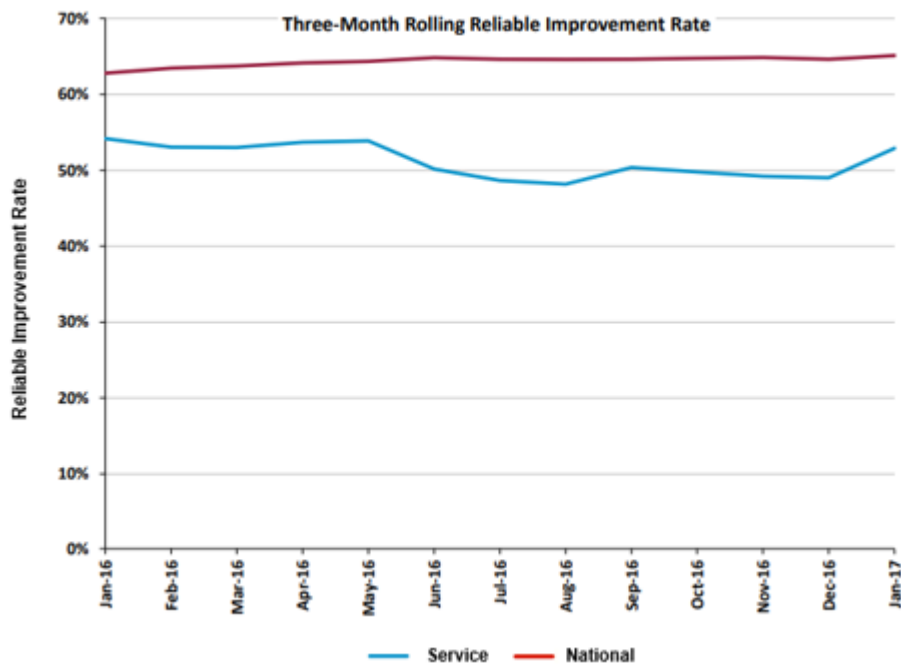
During Phase 1, the service ran as a standard IAPT service, as described earlier in this chapter, and comprised specific features relating to access, therapies, therapist training, clinical supervision, patient population, treatment pathway and service structure. Each is briefly outlined below.

Access

The number of referrals received across the service was relatively stable (Figure 3.7), though numbers of patients dropping out between referral and first appointment (attrition) was higher than the National average (Figure 3.8).

Figure 3.7

Service Referral Rates Jan 2016 – Jan 2017 (Data source NHS Digital Monthly IAPT Data)
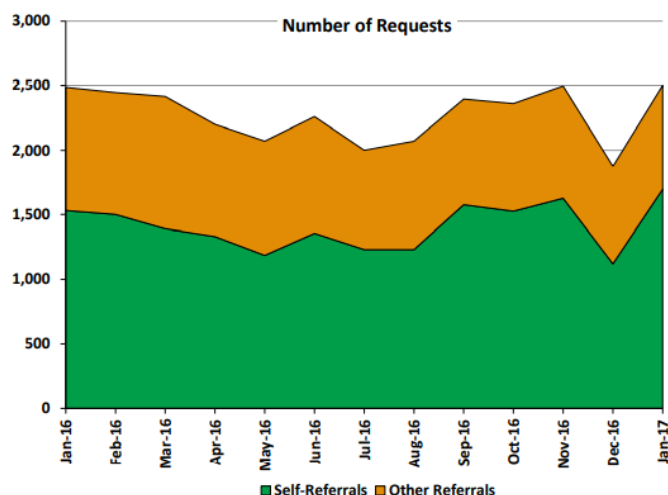
Figure 3.8

Service Attrition Rates Jan 2016 – Jan 2017 (Data source NHS Digital Monthly IAPT Data)
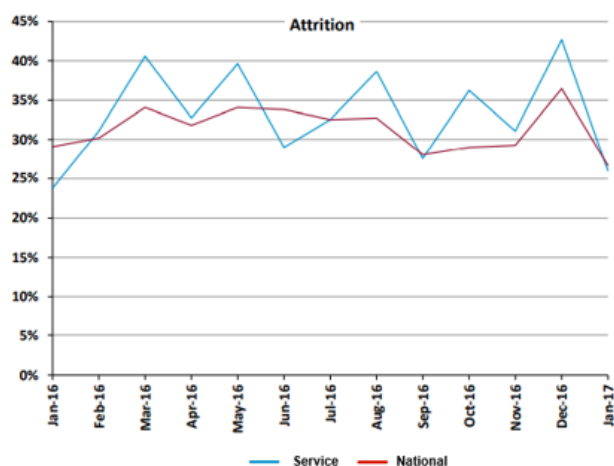
The service was structured to encourage patients to access the service via large introductory workshops. Upon referral, all patients would be invited to attend these workshops as a first step, with individual assessment appointments offered only to those patients who specifically requested this. This access point into the service was a feature of how the service model had been designed, in collaboration with commissioners, to provide a cost effective, easy route into the service. It was thought that attending a drop-in introductory session might be more accessible and less stigmatising than needing to call the service to arrange an individual assessment appointment. In addition, this allowed more patients to access the service at one time, with fewer staff costs, as fewer staff were required to run the workshops (2 staff per workshop of up to 100 patients).

Therapies

The research reported in this thesis focussed on Step 3 therapists and their patients within the service. At Step 3, the service over this period offered a range of IAPT recommended evidence-based psychological therapies: Cognitive Behavioural Therapy (CBT); Person-centred Counselling/Person-Centred Experiential Counselling for Depression (PCE-CfD); Interpersonal Psychotherapy (IPT); Eye Movement Desensitisation and Reprocessing (EMDR) and Couples Counselling for Depression (CCfD). Group-based therapies were also available at Step 3, but only patients accessing one-to-one therapies were included in the study. In contrast to some IAPT services, sessions were not capped and therapy was offered in line with NICE Guideline recommendations for the relevant diagnosis/presenting problem.

Therapist Training

There were 131 therapists within the service during this period, including trainee Cognitive Behavioural Therapists, qualified Cognitive Behavioural Therapists and qualified Counsellors. All CBT therapists were trained to post-graduate diploma level by an IAPT approved CBT course. Counsellors were qualified to meet either BACP (British Association for Counselling and Psychotherapy) or UKCP (UK Council for Psychotherapy) accreditation, and some were additionally trained in CCfD or IPT. Where EMDR was offered this was provided by qualified CBT therapists who were either trained in, or training in, EMDR. Some clinicians also had core professions as Mental Health Nurses, Occupational Therapists, Social Workers, or Probation Officers.

Clinical Supervision

Clinical supervision for Step 3 clinicians is generally less prescribed than for Step 2 clinicians, though the publication of the first IAPT Manual (NHSE & NCCMH, 2018) included the requirement for a minimum of 1 hour per week of individual clinical supervision for both Step 2 and Step 3 clinicians. During Phase 1, all qualified Step 3 clinicians in the service were provided with 1-hour individual clinical supervision per fortnight plus 1.5 hours' clinical supervision group of 3-4 clinicians per month. Trainee CBTs were provided with weekly

clinical supervision in addition to clinical supervision provided by the training provider.

Therapists with an additional therapy (e.g., CBT therapists also providing EMDR) were

provided with an additional 1-1.5 hours of supervision per month (either individual or in small

groups) for the specific modality. This was due to resources being focused on other areas of

the service and there being less capacity to provide weekly one-to-one clinical supervision.

Therapists preferred the inclusion of group supervision within their supervision provision, so

a blended model of individual and group supervision was used during this period.

CBT supervision was offered in a pyramid structure, with experienced staff

supervising more newly qualified staff, but those offering supervision did this as a small part

of their job role, resulting in many different supervisors across the service. All supervisors

were trained in IAPT supervision by a local provider. Counselling staff were supervised by

external supervisors as chosen by the individual counsellor, not all of whom were trained in

IAPT supervision. Although the service management encouraged the review of routine

outcome measures within supervision to guide discussions, there was no structured format

or guidance for doing so during Phase 1.

Patient Population

The service covered a population of approximately 1 million people and accepted

patients aged 16 years and over for assessment and interventions. Standard IAPT criteria

were used for assessing suitability for the interventions within the service; that is, patients

presenting with mild to moderately severe depression and/or anxiety; low to moderate risk

levels, low to moderate chronicity and low to moderate complexity (where complexity is

defined by number of concurrent needs/diagnoses). Patients presenting with stress, low

mood, depressive episodes, recurrent depression, panic disorder (with or without

agoraphobia), generalised anxiety disorder (GAD), post-traumatic stress disorder (PTSD),

health anxiety, social anxiety, obsessive compulsive disorder (OCD), adjustment disorder,

body dysmorphic disorder (BDD), phobias, and those with mixed presentations, were all

catered for within the services in line with IAPT guidance (NHSE & NCCMH, 2021). Based

on service data from March 2016 to March 2017, compared against the National average,

the service had comparable levels of intake severity on the PHQ-9 (Figure 3.9), GAD-7

(Figure 3.10) and WSAS (Figure 3.11).

Figure 3.9

Service PHQ-9 Intake Severity Compared to Overall National Scores (Data source NHS Digital
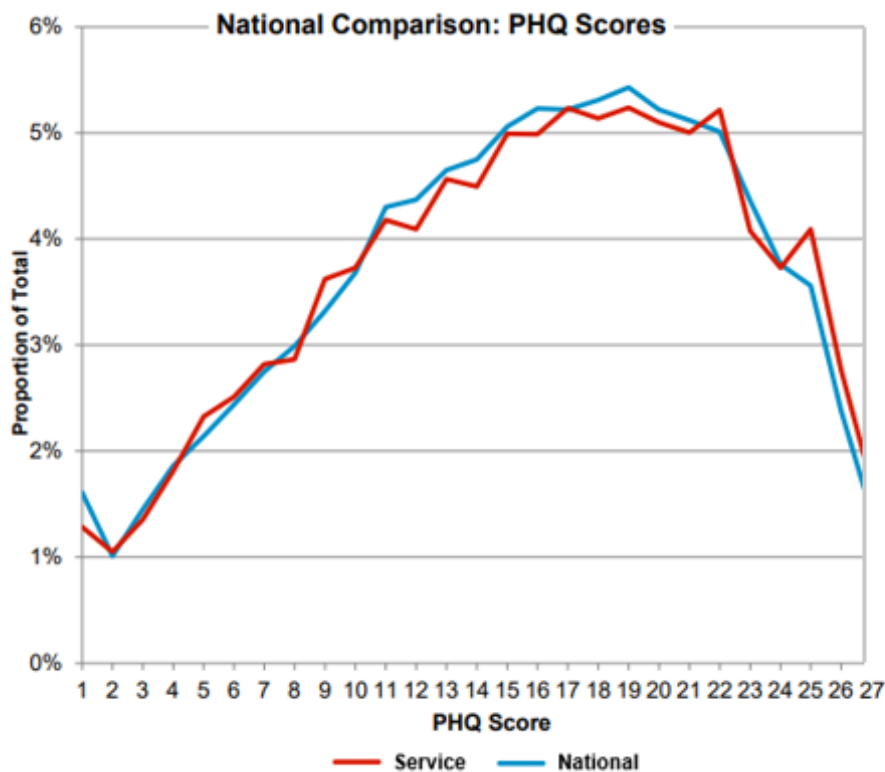
Monthly IAPT Data)

Figure 3.10

Service GAD-7 Intake Severity Compared to Overall National Scores (Data source NHS Digital
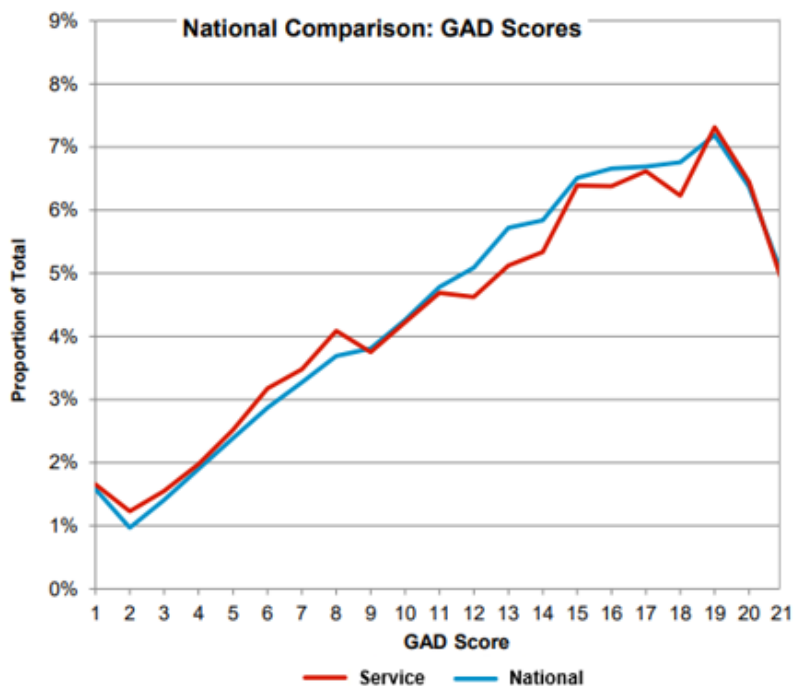
Monthly IAPT Data)

Figure 3.11

Service WSAS Intake Severity Compared to Overall National Scores (Data source NHS Digital

Monthly IAPT Data)
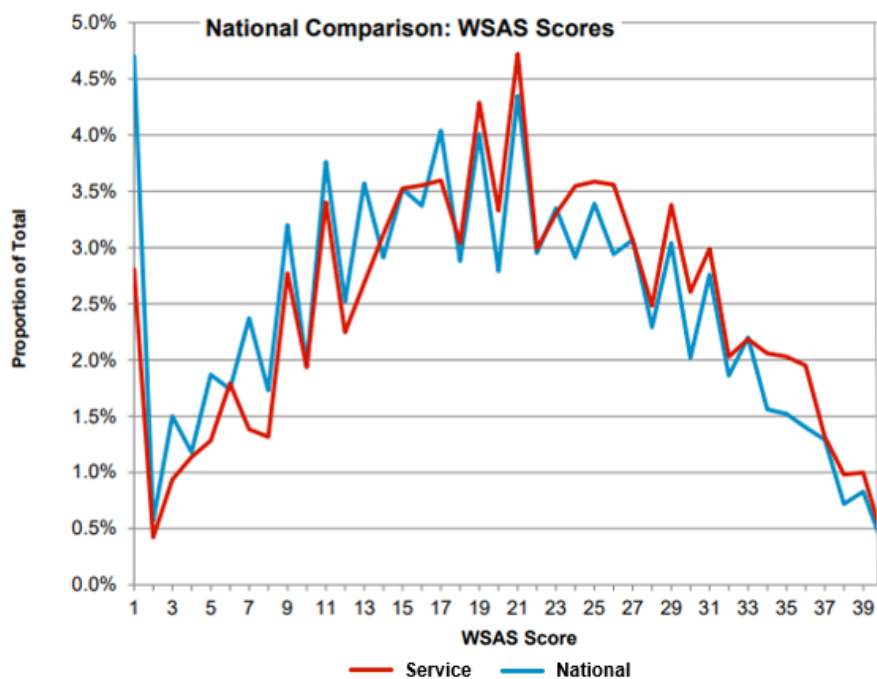
The demographics of the area are characterised by low levels of ethnic diversity

(Office for National Statistics, 2019); higher than the national average mean age (Office for

National Statistics, 2019); a wide range of socioeconomic deprivation levels (Public Health

England, 2019); and a predominantly rural geography, with only one urban city area (Office for National Statistics, 2019).

Treatment Pathway and Service Structure

The service accepted referrals either directly from patients, or via professional referrals. Patients could self-refer via calling the services' dedicated self-referral telephone number, through an electronic self-referral form submitted via the service website, by filling in a self-referral leaflet and posting this to the service administrative centre, or by attending one of the widely-advertised Introductory workshops provided by the service in the community on a regular basis. Professionals could refer via letter or an electronic 'professionals referral form' submitted through the service website. Referrals were accepted from a range of professionals, including primary care clinicians, secondary care clinicians, social service staff, educational staff, employers. Unless otherwise indicated on the professional referral (i.e., indication of high risk or severity), all patients would be invited to attend a workshop session as their first contact with the service. At this initial workshop, patients could then choose from a range of different Step 2 workshops focussing on a range of presenting difficulties (e.g., sleep problems, low mood, stress, and panic). Patients could request an individual assessment appointment as an alternative to the workshop. However, this was not offered routinely during Phase 1. Access to Step 3 therapy was generally only available to patients who had been through the workshop process but were showing limited response to this intensity of intervention.

The service had a separate Younger Persons therapy team, which was managed separately to the adult workforce. This team was managed and supervised within the Secondary Care Youth Service, with patients aged 16-25 years being seen by specific IAPT therapists who only worked with patients within this age group. The management team within the Secondary Care Youth setting had a limited understanding of IAPT and did not have the same leadership structure as the rest of the service during this period. Also during this period, waiting times for younger people exceeded that of adults (service data unavailable) and clinical outcomes were poorer for this group, see Table 3.4.

Table 3.4

Recovery Rates for Local Adult Teams Compared to Youth Team Jan – Dec 2016 (Data source

local service data)

Team	% of cases Recovered	Total No. of eligible cases
Whole Service	42	9375
Team 1	49	1186
Team 2	47	1748
Team 3	43	1421
Team 4	38	1537
Team 5	46	1627
Youth Team	33	1828

The adult teams had comparatively low ratios of management to clinicians, with an

average ratio of 1:16, and with some senior clinical staff required to undertake management

responsibilities as a result. Teams were managed on a geographical basis, with managers

required to manage a wide range of clinicians, including peer support workers, Step 2 and

Step 3 clinicians. Clinicians were employed by two organisations, but there was an

integrated management structure, meaning clinicians shared the same manager across

organisations. However, this also resulted in managers being required to navigate the

human resources policies of two different organisations. The clinical leadership was limited

to 1.5wte Psychologist posts and 2wte Senior Therapist posts across the service during

Phase 1.

Embedding the Research in the Service

As a study specifically designed to generate practice-based evidence in a

naturalistic, generalisable setting, it was a feature of the research that the primary

researcher was embedded within the leadership team of the service being studied. This

allowed in-depth understanding of the context of the service covering the baseline

retrospective data period prior to the intervention Phases. It also allowed the inclusion of the

management-led programme of service improvements as a distinct Phase (Directive Phase)

within the study design, as the full details and range of interventions included during this time could be identified and comprehensively reported. An additional benefit to this position of the researcher within the leadership team, was that this provided the necessary management engagement for the implementation of the third Phase intervention package (referred to as the Collaborative Intervention).

Alongside these advantages, there were also challenges to the project in having the primary researcher embedded within the service leadership team. The Phase 2 changes undertaken across the service were perceived as controversial to the staff group (see Chapter 4 for details) and management-led, or top-down. The purpose of the Phase 3 intervention was to develop a package of support that was initiated and supported by the staff, and so any association made with the primary researcher and the Phase 2 changes was potentially problematic. To mitigate the effects of this, a member of the staff group emerged as a champion of the Phase 3 intervention package and joined the research team as a collaborator. This balanced the embedding of the primary researcher in the service leadership team, by embedding a front-line clinician within the research team. In Phase 3 the research supervisors were also utilised to engage staff during the implementation of the interventions, to provide a wider context to the research intentions.

As a clinical leader within IAPT for a number of years, this dilemma between implementing top-down national directives, as well as aiming to support and develop staff's clinical skills and wellbeing, reflects the role of many lead clinicians within IAPT and other nationally mandated NHS services, adding to the generalisability of the research project as a whole. Although not without its challenges in a research context, being embedded within the service leadership was of significant value to the research project overall, not only because of the opportunity to conduct a large-scale practice-based longitudinal study of therapists, but for the other benefits outlined above.

Summary

This chapter has presented the context for this 4.5-year-long study against the backdrop of a large NHS IAPT service. The National IAPT initiative mandates high levels of

controllability of services and therapists, including training and supervision levels, use of clinical outcome measures, and adherence to NICE guidelines for treatment. However, being an open access, high volume service, this leads to the challenge of catering for a wide range of presenting problems, severity levels and complexity within the patient population. As a context for studying therapist effects, therefore, this provides a setting with high levels of service and therapist control, whilst providing a real world, practice-based environment. The specific elements of the IAPT service that provided the study setting have been presented, in particular the aspects of the service that were bespoke to the commissioning arrangements of the provider, and in contrast to standard IAPT practice. In all, the service largely complied with many of the main IAPT mandates, though there were some discrepancies in relation to Step 3 supervision arrangements, service access points, and the separation of youth staff teams from adult staff teams. It was noted that during Phase 1, the service was experiencing lower than target clinical outcomes (i.e., 'recovery' rates) which prompted the initiation of the Phase 2 and Phase 3 interventions. The benefits and challenges of the position of the primary researcher being embedded within the service leadership was presented, providing insight into the applicability of the research design to the role of clinical leaders in such services. Chapter 5, which follows, plots the development of the Phase 2 intervention package that created the service conditions in the subsequent periods.

Chapter 4

Development and Implementation of Phase 2 Directive Intervention

Overview

The previous chapter set out the context for the three Phases of the study with a

description of the service context during the Baseline Phase (Phase 1). This chapter

presents the process of development and implementation of the Directive Intervention

(Phase 2), a programme of service improvements. A description of the rationale for the

changes made within the service is presented. Details of an external review are included

which provided a series of recommendations aimed at improving the outcomes of the service

overall. Service changes that relate to the Step 3 part of the service are set out in detail and

comprise the Directive Intervention Phase of the current study.

Rationale for Phase terminology

Phase 2: Directive Intervention - Implementation of a programme of changes agreed

by the service management team, as recommended following a diagnostic analysis by NHS

Improvement Intensive Support Team (NHS Improvement, 2017). This programme of

change involved adjustments across all elements of the service and was a 'top-down'

process with minimal staff consultation, thus termed the 'Directive Intervention' Phase. In

making these rapid changes, the service aimed to improve the overall clinical outcomes of

the service as well as reduce waiting times to a sustainable level.

Timeline

A timeline for the three Phases of the study is shown below in Figure 4.1. Phase 2 is

highlighted black and comprises the 18-month period from the beginning of April 2017 to the

end of September 2018.

Figure 4.1

Study Timeline with Three Phases (Phase 2 Highlighted)
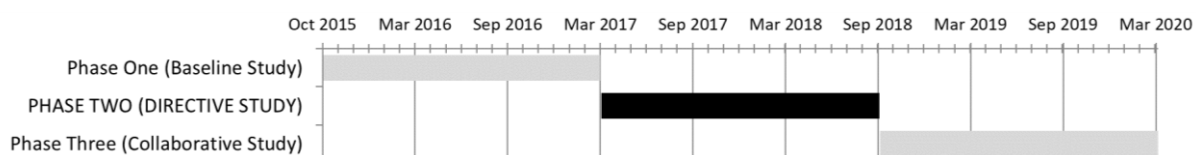
Rationale for Directive Intervention (Apr 2017 – Sept 2018)

As set out in Chapter 3, during the 18-month period prior to April 2017, the clinical outcomes of the service had been breaching National IAPT clinical outcome targets (NHSE & NCCMH, 2021). The service had undergone several diagnostic reviews to identify the reasons for this poor performance, including a review by the NHS Improvement Intensive Support Team (IST). Following these reviews, the service initiated a programme of change aimed at utilising the recommendations from these analyses to improve the clinical outcomes of the service, as measured by IAPT recovery and reliable improvement measures (DH, 2008b, see Chapter 3), ensuring that outcomes for patients were maximised. There were a range of aspects of the service that were identified by the IST review as needing to be adapted to meet these targets, all of which were included within the scope of the changes across the service. From a service, commissioner, and NHS England perspective, the aim of these changes was to reach the nationally mandated target of 50% of patients undertaking treatment in the service reaching clinical recovery, and 60% reaching reliable improvement (NHSE & NCCMH, 2021). There was an additional aim to bring waiting times for treatment to sustainable levels by addressing a backlog of waiting patients. For the purposes of the research project, the aim was to investigate the impact of these directive changes on the outcomes of Step 3 clinicians in relation to between-therapist variability and clinical outcomes for the Step 3 cohort overall.

Although the IAPT service was not included in local Care Quality Commission (CQC) Inspections, the NHS Trust under which the service operated was experiencing significant challenges, following a series of CQC inspections from 2014 to 2018 had identified key areas that needed to be improved (detail withheld to protect anonymity of Trust). This context provided a backdrop to the IST review of the IAPT service. Local commissioners and service senior management were expected to have regular and frequent update meetings with NHS England, supported by NHS Improvement, to report on the changes and improvements that were expected to be made. This structure epitomised the 'top-down',

directive nature of the change process within the service during this period within a context of ongoing poor overall Trust performance.

In developing the programme of changes to be implemented across the service, as well as the NHS Improvement performance report, intelligence was taken from key IAPT review literature (Gyani et al., 2013) and NHS leadership literature (Robinson-Hickman, 2010; Kings Fund, 2014; West et al., 2014) to inform the work. Although the IAPT Manual (NHSE & NCCMH, 2018) was not published until after the initial recommendations had been made, this document was used to inform the work as it developed over the period.

Recommendations from the IAPT Literature (pre-April 2017)

Key features of the IAPT model have been described in Chapter 4, with reference to differences between the standard IAPT model and the adaptations made in the local service context. Papers relating to the actions that services have been recommended to take to improve clinical outcomes are highlighted within this section. The Clark et al. (2009) evaluation of the demonstration sites highlighted the importance of the following: access via self-referral, data completeness, use of the stepped care model, and compliance with the NICE guidelines. In addition to these areas and including the results of semi-structured interviews with service staff and stakeholders, Parry et al. (2011) recommended ensuring the following: separate and clearly defined line management and clinical supervision roles within IAPT services; high levels of clinical supervision for staff; and clear processes in managing 'inappropriate referrals', particularly at Step 2 where staff have high case-load numbers catering for patients with mild to moderate mental health presentations.

Adherence to NICE Guidance and IAPT Supervision Model

The importance of adherence to NICE guidelines was a feature of all subsequent key reviews (Clark, 2011, 2018; DH, 2012b; Gyani et al., 2013). Gyani et al.'s (2013) review of the first year of IAPT utilised data from 32 IAPT services, including up to 19395 patients between Oct 2008 – Oct 2009, with a focus on aspects of services that improved clinical outcomes. Using logistic regression analyses of pre-post scores applying the definition of IAPT reliable recovery (see Chapter 3), they found no significant difference between

counselling and CBT for the treatment of depression. This is consistent with other IAPT service-based studies (e.g., Barkham, Saxon et al., 2021; Pybis et al., 2017) and suggestive of a limitation to the NICE recommendations for the treatment of depression (McPherson et al., 2009), which only recommends counselling for those who decline antidepressants, CBT, IPT, behavioural activation and behavioural couples therapy for depression (NICE, 2009a) - though this critique is out of scope for discussion within this thesis. For the treatment of GAD, in line with NICE guidelines, patients were 1.3 times more likely to reliably recover following a treatment of CBT compared to counselling, though it should be noted that the therapist effect was not included in this analysis.

The analysis also identified considerable variation in overall clinical outcomes between services, with reliable recovery rates varying from 23.9% to 56.5% and reliable improvement rates from 43.6% to 77.1% across the 32 services. To ensure compliance with NICE guidelines it was subsequently emphasised that a robust assessment (Clark, 2011) and compliance in recording the problem descriptor (or working diagnosis) was of key importance within IAPT services to ensure good clinical outcomes, with the percentage of patients with a recorded problem descriptor being associated with services with better clinical outcomes (Clark, 2018). In addition, ensuring patients received the appropriate dosage (number of sessions attended) was emphasised in later reviews, with services that demonstrated better clinical outcomes providing a higher average number of sessions (Clark, 2018; Gyani et al., 2013).

Good data quality regarding problem descriptors and appropriate dosage of treatment may to some extent be reliant on the presence of regular clinical case supervision, allowing a review of the treatment pathway for patients with reference to their problem descriptor and treatment type. As well as being a consistent key feature of IAPT services (NHSE & NCCMH, 2018, 2021), weekly supervision is cited in the IAPT reviews (Clark, 2011; Parry et al., 2011) as being an important element of IAPT services, though compliance with supervision levels is not a reported measure within IAPT services generally, nor is it reported within the reviews. A study by Westwood et al. (2017) involving 201 IAPT clinicians

(105 Step 2, 96 Step 3) found that higher levels of supervision in IAPT services predicted lower levels of burnout (measured by the Oldenburg Burnout Inventory) in Step 2 clinicians, though this was not a predictor for Step 3 therapists and was not linked to clinical outcomes.

Impact of Waiting Times on Clinical Outcomes

One of the key challenges that the IAPT initiative sought to address was the lack of access to psychological therapies for those with mild to moderate mental health problems (see Chapter 3). Prior to IAPT, if psychological therapies access was available at all, it was often associated with long waiting times (Bird, 2006) and challenges associated with increasing waiting times continued to be highlighted as an area of concern for IAPT services after their introduction (DH, 2012b). Data from IAPT reviews have demonstrated that longer waiting times for treatment are associated with poorer recovery outcomes (Clark, 2018). Parry et al. (2011) also highlighted a focus on variance outliers in waiting times as a recommendation from the experience of the demonstration site services. However, there is also evidence of waiting times having a limited impact on clinical outcome post treatment (Beck et al., 2015). Indeed, it is possible that a greater impact of waiting times on clinical benefit is due to the likelihood of a patient not attending therapy, or not accepting a referral for therapy, if they perceive that there will be a significant wait for the treatment (Snape et al., 2003).

Other service features observed by the Gyani et al. (2013) review to be associated with services with higher recovery rates were: a greater proportion of patients being stepped up into Step 3 treatment; services treating higher numbers of patients; and services with a greater proportion of higher paid staff (NHS Agenda for Change Band 7 and above).

Description and Explanation of Directive Intervention

The recommendations for the service following the review with the IST were provided in a summary report (20th April 2017) and were formally agreed and accepted on 19th July 2017. Due to the complexity associated with the identified changes (19 recommendations), the recovery/recommendations action plan was undertaken as a formal programme of projects, reporting through to a Task and Finish Group Programme Board chaired by a

representative for the relevant Clinical Commissioning Groups (CCGs). The full list of recommendations can be found in Appendix D.

The service changes can be understood within the context of the key learnings identified within the early IAPT reviews summarised above, and relevant changes impacting Step 3 either directly or indirectly are presented below according to these themes:

1. Improved NICE treatment compliance (Clark, 2011, 2018; Clark et al., 2009; DH, 2012b; Gyani et al., 2013; Parry et al., 2011):

    a) Movement to an initial one-to-one assessment for 100% of patients entering the service to ensure the suitability of treatment for each patient, including use of problem descriptors.

    b) Reduction of workshop-delivered interventions and increase in one-to-one therapy, thereby ensuring increased access for patients to the most evidence-based treatments.

2. Reduced waiting times for treatment (Clark, 2018; DH, 2012b)

    a) Amalgamation of the youth team staff and youth patient pathway into the general adult team and pathway, to provide equity of waiting times for patient groups.

    b) Implementation of a rapid brief intervention pathway for patients having waited longest.

    c) Development of full-service demand and capacity modelling and making improvements in staff productivity and activity monitoring.

3. Compliance with IAPT supervision model (Clark, 2011; Parry et al., 2011):

    a) Implementation of increased, regular, outcome focused case management and clinical supervision for all therapy staff, from 2 hours per month to weekly 1:1 supervision. Aim: to provide robust clinical oversight and assurance as to the quality of the treatment being offered.

    b) Creation of a dedicated supervisor team. Aim: to increase the consistency and quality within supervision across the Step 3 part of the service.

An overview of key projects is summarised below to provide the context for the staff morale and working environment during the Directive Intervention Phase (Phase 2). Information about the discreet workstreams and implementation process are included to illustrate the wide scope and breadth of the changes made across the service impacting all staff groups over this period, to ensure that this context is clear as a consideration for the outcomes of this practice-based research. It is likely that the embedded nature of the primary researcher allowed a greater depth of context and detail as to the status of the service and working environment for staff than would ordinarily be possible in this type of research and this is reflected in the commentary documentation sections based on the researcher's own research diaries and reflections.

Figure 4.2 presents the timeline for the key project areas as they were planned and undertaken during Phase 2.

Figure 4.2

Phase 2 Project Timeline

Phase 2 — Planned — Actual

ACTIVITY	2017									2018								
	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Move to 100% Assessment																		
Problem descriptor implementation																		
Reduction of workshop delivery																		
Youth/Adult team integration																		
Supervision model strucuture change																		
Fully recruited dedicated Supervisor team																		
Interim pathway development & completion																		
Introduction of demand & capacity modelling																		

Improved NICE Treatment Compliance

This area of improvement focussed on changes that supported the application of NICE guideline treatment compliance across the service (Clark, 2011, 2018; Clark et al., 2009; DH, 2012b; Gyani et al., 2013; Parry et al., 2011). This involved changes to both how this treatment was chosen (i.e., through one-to-one assessment) and the treatment delivery types (i.e., the reduction of group-based treatment and increase of one-to-one treatment).

Assessment for 100% of Patients Accessing Service. The service had been designed on a model using large, high volume introductory workshops to increase access to

low intensity interventions in the community. This introductory workshop model meant that patients could simply book on, or turn up, to an introductory session at a range of central venues in their community (e.g., community centres, colleges, supermarkets), and at the end of the session, sign up to one of a large number of community-based, Step 2 workshops, without the need for an assessment. Although individual assessments were available, patients were generally encouraged to attend an introductory workshop first. The introductory workshops included CBT-based psychoeducation about depression and anxiety, as well as a brief overview of the range of Step 2 treatment workshops on offer. Although there were benefits to the introductory workshop model in terms of accessibility and managing high volumes of patients, the structural flaw within the design reflected the lack of application of the IAPT model, which recommends an individual, holistic assessment for each person accessing the service to ensure NICE recommended treatment is offered in line with the problem descriptor (Clark, 2018; NHSE & NCCMH, 2021). The lack of assessments taking place in the service was also reflected in the lack of data completion in relation to problem descriptors (see Table 4.1 below). An unintended consequence of the introductory workshop model was the potential for patients that had higher intensity (Step 3) intervention needs being delayed in accessing Step 3 treatment. As well as this being non-compliant with NICE guidance, it was hypothesised that this could lead to a patient experience of being 'stuck' within the workshop programme, without an easy route to alternative treatment options such as one-to-one therapies.

The IST recommendation was to revert to a 100% patient assessment model, where all patients were to be provided with an individual holistic assessment upon accessing the service, and then directed by the clinician to the appropriate part of the service for their needs (NHSE & NCCMH, 2021). This assessment would involve the correct 'problem descriptor' being identified for each patient, indicating the potential treatment pathway when considered alongside symptom severity. Within the context of the service improvements, the change to a 100% assessment model aimed to improve the access to one-to-one therapies at both Step 2 and Step 3, and ensure a quicker patient journey, so that those with higher

intensity needs would be stepped up to Step 3 - a feature of higher performing services (Clark, 2018; Gyani et al., 2013). It was anticipated that this would result in a potential reduction in severity levels at Step 3, as patients would be stepped up prior to their symptoms deteriorating and may improve the clinical outcomes across Step 3 overall.

Key actions were:

- The cessation of direct access/introductory workshops and replacement of these with individual assessments for all patients
- A structured review of the assessment content and increase in time allowed for assessment (from 30 minutes to 45 minutes) to ensure sufficient time for assessment, mental health clustering, treatment decision making and provision of brief psychoeducation or self-help resources
- Development of a pathway for the clinical review of patients following workshops, if they had not clinically recovered following a group intervention

Implementation and Commentary. Direct access workshops (i.e., workshops accessed without a prior individual assessment) were phased out from August 2017 and ceased in Jan 2018, with an increase in assessments as shown in Table 4.1.

Table 4.1

Percentage of Patients Receiving a One-to-One Assessment

Month/Year	Aug 17	Sep 17	Oct 17	Nov 17	Dec 17	Jan 18	Feb 18
Trajectory (%)	40	40	40	40	60	100	100
Actual (%)	74	77	88	94	95	98	99

To ensure a robust assessment and timely access to self-help resources, assessments were increased in length from 30 minutes to 45 minutes. This saw an improvement in the use of problem descriptors, as shown in Table 4.2. A process of clinical quality assurance was introduced, with clinical supervisors utilising individual clinician caseload spreadsheets, which provided an overview of each patient and the associated problem descriptor and most recent clinical outcome measure. These caseload reviews were

used within supervision sessions on a weekly basis (see supervision changes section below). This ensured that any absence of these clinical tools being used was addressed at an early stage, and that the service was compliant with appropriate NICE treatments according to presenting problem and symptom severity.

Table 4.2

Percentage of Total Patients with Problem Descriptor Defined at Assessment

Year	2017					2018							
Month	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Trajectory (%)	20	30	30	40	50	60	80	100	100	100	100	100	100
Actual (%)	46	56	61	67	70	80	83	83	87	86	89	91	99

Feedback from staff during this time was mixed. Many staff welcomed the opportunity to cease the introductory workshops, which were large (some with over 100 patients in attendance) and as a result posed challenges when risk was identified, or behavioural issues presented. There was also a positive response to the increase in assessment length. However, this view was not universal and staff feedback included a preference for a minimum 60-minute assessment.

Reduction of Workshop-Delivered Interventions. To further ensure NICE treatment compliance, the IST report recommended a reduction in the volume of workshop-delivered post-assessment interventions and to increase the individual intervention provision. Step 2 treatment workshops were reviewed utilising outcome and drop out data from these interventions (Table 4.3), which resulted in one workshop being discontinued (behavioural activation only workshop), and a reduction in the numbers/frequency of workshops being offered overall. This was in line with the reduction in demand for these workshops due to the move to the 100% assessment model. A process of individual patient review was introduced following workshops for those patients who either dropped out or whose scores did not meet the recovered criterion by the end of the workshop sessions. This review was undertaken by the clinician who assessed the patient initially and directed them

to the workshop, with the aim of improving clinical continuity, patient experience, and clinical decision making at the point of assessment. Having a feedback loop to the assessing clinician when a decision had resulted in a poor experience for the patient enhanced therapists learning and potentially informed future clinical decision making. This review also enabled a step up to higher intensity (Step 3) treatment where appropriate and as such it was anticipated that this change would have an indirect impact on Step 3 clinical outcomes, ensuring NICE guidance compliance.

Table 4.3

Snapshot Analysis of Workshop Outcomes 2017 (Data source local service)

	Percent of patients attending Behavioural Activation only workshop (N = 68)	Percent of patients attending Worry workshop (N = 66)	Percent of patients attending Depression psychoeducation workshop (N = 59)
Recovery	14	19	33
Reliable Improvement	30	28	33
Completed all sessions	26	53	47
Dropped out	40	26	53
Referred but did not attend	34	21	Not measured

The workshops were all delivered, and in some cases developed, by Step 2 staff, and as such, there was some dissatisfaction at the changes outlined above. The analysis of the workshops as presented in Table 4.3 challenged some staff perceptions that the workshops were effective and well-liked by patients. There were anecdotal reports of some patients using the workshop programme as a means of social interaction, attending a succession of workshops based more on their availability rather than the utility of the treatment being provided. Some staff reported that the loss of variation, of providing both group and individual interventions, had led to a reduction in job satisfaction. There was a limited impact on Step 3 staff in terms of staff feedback.

Reduced Waiting Times for Treatment

This workstream focussed on changes that aimed to reduce the amount of time that patients were waiting to access their full treatment package (Clark, 2018; DH, 2012b). Changes were primarily structural and operational in nature and included actions that would impact the longer-term waiting times for the service, as well as addressing the existing backlog of waiting patients.

Youth/Adult Team and Pathway Integration. Part of the pre-existing service design had been the separation of a specific staff group to provide interventions for younger patients. The aim had been to improve integration with other youth services within the wider mental health trust and provide a more bespoke treatment package for this patient group. However, in practice, due to pressures elsewhere in the mental health system (mentioned above), and the separation of leadership and management of the youth team external to the IAPT structure, this resulted in IAPT staff resource being used to treat patients within other parts of the system. This resulted in longer waiting times and poorer clinical outcomes for younger people accessing the service, as compared to the patients accessing the adult pathway into the service (local data source unavailable). The agreed recommendations were to re-integrate the adult and youth IAPT pathways, bringing the youth IAPT team within the leadership of the rest of the IAPT service, and incorporating all IAPT youth patients within the overall IAPT waiting lists across the service, to ensure equity of access and provision for patients of all ages.

Key actions were:

- A staff impact assessment and informal consultation for Youth team IAPT staff to be conducted with a view to changing the management structure
- The redesign of the clinical pathway to bring the pathway for young people in line with the new adult pathway
- The introduction of youth IAPT staff into the core locality adult IAPT teams

- The amalgamation of treatment waiting lists to ensure an equitable service across all patient age ranges

Implementation and Commentary. The process of moving a staff group dedicated to the needs of younger people into the service at large involved a consultation process and a phased period of management transition. The identification of a named Lead Therapist responsible for overseeing clinical service development for young people in IAPT allowed the work that had been done within the youth IAPT team of developing age-appropriate ways of working with younger people, to be retained. Additional training was provided for staff within the adult IAPT service where they self-identified deficits in skills of working with patients within this age bracket. Staff who had previously worked in the youth IAPT team became Youth Champions within the core service, sharing skills and knowledge with their peers, though they were expected to work with IAPT patients of all ages. The integration of staff and patients into the broader IAPT service was completed by September 2017. Observations of the primary researcher were that the transition of staff from the Youth service to the IAPT service was viewed as controversial by the staff involved. The consultation process was viewed with some suspicion and some staff felt that the skills and specialisms of the youth staff group had been dismissed and/or unappreciated by management. In contrast to these views, however, there was also feedback from some youth staff that they recognised the long waits for their patients and welcomed an opportunity to reduce waiting times by sharing resources across the whole IAPT service.

Implementing a Rapid Brief Intervention Pathway for Longest Waiting Patients. The service had accrued a significant backlog of patients waiting for Step 3 treatment and without removing this backlog the service was unable to reduce or maintain improved overall waiting times for treatment. This project therefore worked on providing rapid access to brief interventions for the patients having waited the longest for treatment (over 6 months), to reduce overall waits to a sustainable level.

It was noted that this workstream had two primary associated risks: a) that the rapid access, brief intervention pathways may result in a brief reduction in recovery rates during

the lifetime of this interim clinical pathway, as the clinical intervention was not NICE guidance compliant and therefore may not result in recovery and; b) that the staff resource used to deliver the pathway would need to be backfilled in order to ensure no depletion in staff resource to carry out treatment as usual for those patients not being offered this briefer package of care. It should be noted that patients on the interim treatment pathway were excluded from this study.

Key actions were:

- The development of an interim pathway clinical intervention package
  - Design of a brief intervention based on '2+1' model (Barkham et al., 1999) and single session therapy skills (Talmon, 1990) and incorporating workshop access utilising digital methods (i.e. access to recorded webinar sessions) as well as peer support (Basset et al., 2010; Gillard, 2019)
  - Collaboration with staff on the development of this package to increase staff engagement
- A review of the clinical notes of all patients waiting over 6 months for treatment to identify their suitability for the newly developed brief interim pathway treatment package
- The implementation of staff training for those staff providing the interim pathway treatment package
- The completion of all interim pathway treatments within the agreed timescale (6 months from initiation)

Implementation & Commentary. A total of 1253 patients were identified through the review process as being appropriate for the interim pathway, based on length of time waiting and nature of presenting problem based on clinical notes. The patients were then contacted and offered the option of the brief interim pathway package of treatment. Any patients who declined this offer remained on the waiting list for treatment as usual. The planned trajectory

for moving patients through this pathway, and the actual numbers, are shown in Table 4.4.

All patients moved through the pathway by the end of March 2018.

Table 4.4

Reduction to Waiting List Backlog

Year		**2017**					**2018**	
Month		Sep	Oct	Nov	Dec	Jan	Feb	Mar
Number of patients remaining on waiting list backlog	**Planned**	1000	930	733	558	372	186	0
	Actual	1073	1258	973	579	247	137	0

This workstream was subjectively one of the most clinically controversial parts of the programme of service changes, both from a staff engagement and clinical governance perspective. The clinical risks and negative clinical impact to patients who had waited for over 6 months for treatment had to be weighed against the potential impact of offering a treatment package that was not in line with NICE guidance. For this reason, a working group of interested staff members was convened to determine the treatment package to be offered, and informed patient choice was included as part of the pathway design. Despite this, however, the pathway was generally viewed negatively by the staff cohort, and only a small proportion of the staff asked were willing to offer treatment within the pathway.

Demand and Capacity Analysis. Alongside the project to reduce the number of patients waiting more than 6 months, a workstream focusing on resolving the systemic issues that were contributing to the increasing waiting times within the service was established. This work involved a whole service demand and capacity modelling process to determine any resourcing issues that might result from the service changes (e.g., increase to number of individual assessments, reduction in group interventions). It also aimed to identify issues that were present in the service commissioning, as well as improving the productivity of the existing staff group (NHS Improvement, 2017).

The IST recommended that the service use a demand and capacity tool developed by their team (a version of this can be sourced at: https://www.england.nhs.uk/ourwork/demand-and-capacity/models/), to understand the underlying reasons behind the long treatment waiting times within the service. The tool allowed the service and CCGs to map the current productivity against potential capacity if service productivity were maximised. The tool also mapped the resource required to meet demand based on the new service model and reduced waiting lists.

Key actions were:

- The application of a demand and capacity modelling tool in all CCG areas to identify staffing gaps

- The use of staff productivity dashboards to effectively manage staff activity based on a target of 20 hours of attended clinical work per week (i.e., time spent with patients, excluding other tasks such as writing notes, onward referrals, safeguarding activity etc.)

Implementation and Commentary. The demand and capacity tool indicated that there was potential to improve the clinical productivity of the service, but also that there was insufficient resource to meet the clinical demand of the new service model. The implications of this were that if additional clinical resource was not incorporated into the service, then the waiting list would be unsustainable, and waiting times would continue to increase following the successful work of the project to reduce the number of longest waiting patients (see Table 4.4).

Staff activity dashboards were developed and used in line management of all staff. These dashboards included data such as the number of hours of clinical work undertaken (compared with the agreed target), alongside clinical outcome data (e.g., recovery and reliable improvement rates for the individual therapist). This information was provided to staff, with support provided where there were discrepancies in expected and actual activity. Staff were educated regarding the IAPT expectation of 20 clinical hours per week (NHSE & NCCMH, 2021) and consulted to identify the activities that were impacting on their available

clinical time. Additional administrative posts were created to reduce these non-clinical activities and to maximise clinical time. Additional clinical posts were agreed by commissioners and recruited during this Phase (data on numbers of additional posts agreed not available).

While the interim pathway was the most clinically controversial workstream, the demand and capacity workstream posed the most controversial and unpopular from a staff feedback and wellbeing perspective. Staff feedback regarding the report from the IST team that *"productivity amongst clinical staff is low"* (NHS Improvement, 2017, p.5) and *"productivity is extremely low"* (p.12) was not mitigated by the commentary that *"this may not reflect that staff are not working hard, but that the activity undertaken is not always value-adding"* (p.12). Feedback from staff was that the conclusion that they were not working productively, and the introduction of activity dashboards, significantly impacted their morale and wellbeing. This was reflected in the staff survey results (Table 4.5 and Table 4.6, for comparisons with Phase 3, see Chapter 7, Results All Therapists) over this period, though cannot be assumed to be a direct correlation. Staff also reported a feeling of anxiety about their job security in relation to the activity dashboards, with some staff reporting that they perceived this as a tool to remove them from their employment.

Table 4.5 and 4.6 below are data provided by the service, showing results from internal surveys carried out through the Directive Intervention (Phase 2) period, administered via anonymous survey to all staff (Step 2, Step 3, admin, management). During Phase 2 between 49-68% of staff who answered reported having felt unwell as a result of work-related stress in the last 12 months. This had dropped to 49% in the last three months of Phase 2. In relation to actions being taken by the service to improve staff wellbeing, in Phase 2 between 20-39% of staff who responded felt that the service did not take positive action in this regard.

Table 4.5

Staff Responses to Internal Survey: Levels of Work-Related Stress Apr 2017 – Sep 2018

	Apr - Jun 2017		Jul - Sep 2017		Oct - Dec 2017		Jan - Mar 2018		Apr - Jun 2018		Jul - Sep 2018	
	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%
Yes	48	61	31	63	69	62	57	68	48	67	21	49
No	31	39	18	37	43	38	27	32	24	33	22	51

During the last 12 months have you felt unwell as a result of work-related stress?

Table 4.6

Staff Responses to Internal Survey: Perceived Service Investment in Staff Wellbeing Apr 2017 – Sep 2018

Does your organisation/service take positive action on health and well-being?

	Apr - Jun 2017		Jul - Sep 2017		Oct - Dec 2017		Jan - Mar 2018		Apr - Jun 2018		Jul - Sep 2018	
	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%
Yes, definitely	11	14	6	11	6	5	10	12	9	12	3	5
Yes, to some extent	42	54	33	60	80	72	39	49	43	58	44	75
No	25	32	16	29	25	23	31	39	22	30	12	20

Compliance with IAPT supervision model

IST recommendations included an increase in case supervision in line with IAPT guidance (Clark, 2011; NHSE & NCCMH, 2021; Parry et al., 2011), and this project was expanded by the service management to include a review of supervision more generally. This included staffing (in line with the Parry et al. (2011) recommendation of separate and clearly delineated management and supervision roles) and supervision content (i.e., the

activities being undertaken within supervision, such as the use of caseload dashboards). The pre-existing supervision model involved many therapists (up to 30 in total) from across the service each offering supervision to a small number of other therapists, based on their number of years post-qualifying. Two 1-hour one-to-one supervision sessions plus one peer group supervision session were offered to each clinician per month. For counsellors, the one-to-one supervision was sourced from private external supervisors chosen by the individual counsellor. Supervision was focused solely on the discussion of patients that the therapists themselves highlighted and therapists' overall caseload was not necessarily reviewed as part of this process. Supervision did not utilise caseload data from the clinical system (IAPTUS) other than clinical notes, and clinical outcome measures were not routinely monitored as part of supervision.

Key actions were:

- The implementation of recovery-focused weekly individual case supervision
- The development of a dedicated supervisor team to improve the consistency and quality of supervision
- The design and implementation of robust assurance processes within supervision related to clinical decision making and treatment fidelity
  - Developing treatment package guidance to improve the fidelity to NICE guidance when offering treatment options, to be used in supervision
  - Implementing regular audits of supervision spreadsheets and treatment adherence

Implementation and Commentary. A specific group of Step 3 supervisors was formally recruited from the therapist cohort within the service, with their primary focus being the delivery and development of clinical case supervision across Step 3, including counselling. These supervisors had reduced caseloads (0.2wte of clinical work, 0.8wte of supervision tasks) allowing them to deliver supervision to a larger group of therapists, thus increasing the chances of improved consistency in supervision and clinical decision making

across the service. This group of supervisors was line managed by one lead therapist who provided support and oversight of the supervision provided by this group. Supervision was increased to weekly 1-hour one-to-one sessions, utilising caseload spreadsheets which provided a summary of the therapist's caseload including key clinical information such as: recent clinical outcome scores, problem descriptor, number of attended sessions to date, and date of last attended session. It was mandated that clinical treatment decisions, such as the intervention offered following assessment and the decision to step up/down or discharge a patient, must be discussed in supervision. This ensured greater consistency across the service and assurance that clinical decisions were based on evidence-based practice and NICE guidance. Peer supervision groups ceased, but a monthly clinical team meeting was introduced in each locality area, where Step 3 teams focused on clinically related (non-operational) issues or service changes/updates. The increase of supervision and change of structure was implemented by September 2017. However, the new supervisor team was not fully recruited until January 2018.

The change to the supervision structure resulted in opportunities for staff who were interested in supervision to undertake this as a designated role, with additional support and oversight from a Lead Therapist. Though this opportunity was welcomed by the staff who applied for these roles, the staff who lost the supervision activities within their therapist role gave largely negative feedback about this loss. Some counsellors in particular reported dissatisfaction with the loss to their supervision and supervisors which had been provided by individuals external to (i.e., not employed by) the service.

Summary

As described, this Phase involved significant change across multiple areas of the service with the aim of improving clinical outcomes. The main areas of positive impact anticipated at Step 3 were:

- A potential reduction in the severity of patients being treated at Step 3, due to both the early assessment provided for all patients at the beginning of their treatment and reduced waiting times.

- An improvement in the quality of therapy provided by clinicians due to the increased amount of supervision provided and improved consistency and quality of the supervision.

However, it should be noted that even the changes that were seemingly uncontroversial (e.g., providing staff with more clinical support through increased levels of supervision), were all initiated and implemented in a directive, high pressure environment. The changes themselves were largely identified and mandated by the IST, and although staff were interviewed as part of the initial diagnostic analysis by IST, there was little staff consultation involved in setting the recommendations or the service improvement process. Regular formal contact through updates at the Programme Board and other forums between service leadership, NHS England and Commissioners, set tight deadlines and targets for the service to demonstrate that progress was being made. The staff group experienced changes to their management, supervision, and the therapy that they were offering (for those offering 'interim' pathway treatment). Both management and front-line staff reported high levels of work stress over this time, reflected in regular temperature checking undertaken internally within the service over this period (see Table 4.5 and Chapter 7).

These whole service changes amounted to a service redesign; to sustain them for the long term a leadership and management review was undertaken midway through Phase 2. The results of the leadership review were implemented at the end of Phase 2, coming into full effect at the beginning of Phase 3 and are therefore included in the description of the Phase 3 Collaborative Intervention in Chapter 5.

Chapter 5

Development and Implementation of Phase 3 Collaborative Intervention

Overview

Chapter 5 provides details of the Phase 3 Collaborative Intervention, which was the final research Phase. The service context is briefly described and contrasted with that of Phase 2 (described in the previous chapter). An overview of the leadership models that informed the leadership and management review at the beginning of this Phase is presented. This is followed by a description of the collaborative development of the Phase 3 intervention, which involved therapist focus groups to determine a package of support that might contribute ultimately to improved patient outcomes. A description and overview of the application of deliberate practice within psychotherapy is presented, as this formed one of the elements of the intervention package in this Phase. The chapter provides information about the implementation of deliberate practice across the service in a stepped wedge design. Finally, an overview of the staff wellbeing aspect of the intervention package is presented. Ethical approval for Phase 3 was granted by the Health Research Authority and Health and Care Research Wales (Ref: 18/NS/0104; see Appendix E).

Rationale for Phase Terminology

Phase 3 involved the implementation of a bespoke intervention package at the therapist level aiming specifically to test the hypothesis that, through such an intervention, Step 3 therapist variability would be reduced, and Step 3 clinical outcomes improved. The process of development and implementation of this intervention package explicitly aimed to be carried out in collaboration with therapy staff, involving and engaging staff in the process as much as possible, hence being termed the Collaborative Intervention Phase. The main aspects of the intervention package were: the introduction of a 'deliberate practice-lite' (DP-lite) package, and a series of staff wellbeing events. The DP-lite package was implemented in a stepped wedge design primarily for practical reasons, however this enabled the amount of time therapists had been exposed to DP ideas specifically (as opposed to just the reflective practice skills groups) to be used as a variable in the analysis. Both aspects of the

package were supported by a management and leadership structural review implemented in

Phase 3. This restructure saw increased management support for a compassionate

leadership model (West et al., 2014) and also refocused the service leadership on staff

wellbeing and engagement. A timeline for the three Phases of the study is shown below in

Figure 5.1.

Figure 5.1

Study Timeline with Three Phases (Phase 3 Highlighted)
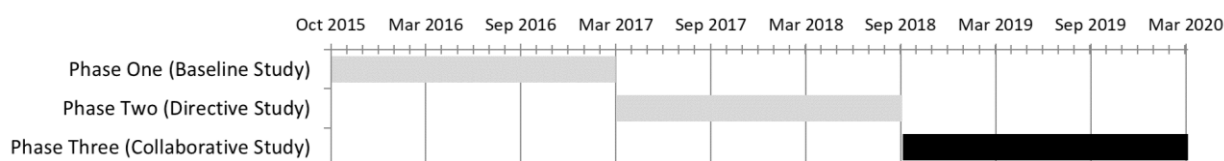
Phase 3: Collaborative Intervention (October 2018 – March 2020)

The Collaborative Intervention in Phase 3 comprised three main elements:

- A leadership and management model redesign

- The development of DP-lite training and skills groups

- A focus on staff wellbeing through the implementation of specific wellbeing events

The context for the development and implementation of the Collaborative Intervention

in Phase 3 was significantly different than in Phase 2. Having achieved the desired recovery

and reliable improvement rates (NHSE & NCCMH, 2021), the service was under less

external pressure to make rapid wholescale changes across the system. A recognition of the

impact that the changes, or the way in which they had been implemented, had made on staff

morale and culture, was felt across the senior leadership team, highlighted in part by the

leadership review undertaken at the end of Phase 2. Feedback from the staff survey (see

Chapter 4, Table 4.5 and Table 4.6) added to the anecdotal feedback received by

management from staff teams, as the impact of the previous 12 months was felt. A focus on

externally driven, target-based goals had left the service, and the senior leadership team,

ready for a change of culture and approach. Just as the context of directive change was

important for understanding the factors impacting the results of Phase 2, the context of the

gradual emergence of a more collective, compassionate leadership style provided an

important back drop to the approach taken in Phase 3. To some extent the approach taken by the primary researcher in developing the Collaborative Intervention package was a direct response to the directive approach taken by management in Phase 2. A lack of staff engagement at this time meant that, establishing the involvement of clinicians in a process of improving their clinical outcomes and reducing variability, would necessarily involve their active participation as co-creators of the processes by which these improvements might be realised. As a senior member of the clinical leadership of the service during Phase 2 and 3, the primary researcher aimed to utilise the research process itself as a mechanism by which the overall service management could be supported to begin the process of adapting the service culture to one that was more compassionate and quality-focused. This section does not aim to provide a comprehensive review of leadership models, but rather to provide a summary of the key models and papers utilised by the leadership team and primary researcher in approaching the development of the Phase 3 intervention. These ideas were central to the leadership review, and also the approach in which the Collaborative Intervention was created and implemented across the service.

Key Leadership Models Included in Review

One of the ways in which leadership theories diverge is the focus on either one individual leader and how they influence those they lead, or theories of shared leadership which are often more aligned with organisations that have, and require, distinct teams to carry out the organisational functions (Avolio, 2009). An example of a more individualistic leadership theory is cognitive leadership.

Cognitive Leadership. Cognitive leadership is an approach to leadership utilising the ideas from cognitive schema research and the psychological understanding of the self-concept (Lord & Brown, 2001). This theory suggests that leaders influence in two key ways: (1) based on values – that is, articulating values that are important or relevant to those to whom they lead as a means of motivating behaviour (e.g., values of care, quality or learning); and (2) based on the self-concept – that is, the leader articulates, models or in some way activates an identity that those being led want to aspire to, potentially creating a

collective identity for the team or organisation to work towards (Lord & Brown, 2001). Individual schemas and the team identity act as a prism through which the actions of leaders are seen and interpreted and are therefore key to the development of the team culture. This may also be seen within transformational or charismatic leadership theories, where a leader cultivates the gap between the actual and aspirational self in others, articulating a vision of a set of higher order values that others can relate to (Avolio, 2009).

Collective Leadership. Collective leadership, in contrast, focusses more on the team or the group, rather than an individual leader being the director of change and culture. West et al., (2017) *Caring to Change* paper defines collective leadership as having a number of key features:

- *"everyone taking responsibility for ensuring that there is high-quality, continually improving and compassionate care*

- *shared rather than dominating leadership in teams*

- *continual development of teamworking*

- *interdependent leadership with leaders working together across boundaries, prioritising patient care overall, not only in their area of responsibility*

- *a consistent approach to leadership across organisations, characterised by authenticity, openness, curiosity, kindness, appreciativeness and, above all, compassion"*

   (West et al., 2017, p. 9)

Other definitions emphasise the sharing of leadership across teams based on the expertise required for the particular task or problem at hand (Friedrich, 2009) where different members, or leaders may take the lead at different times depending on need and context. This approach has been seen as particularly useful in contexts where team working is fundamental, and problems require solutions quickly (Friedrich 2009). Important elements of this leadership model include the sharing of information, clear identification of the expertise of team members, leader accountability, and collective goals. Carmeli and Schaubroeck

(2006) studied top management teams and demonstrated a relationship between specific behaviours and organisation outcomes. Positive organisational outcomes were associated with the sharing of information across leaders, collaboration, and joint-decision making, (these being a significant predictor of final decision quality [r = .27] which was negatively related to organizational decline [r = −.56]).

However, collective leadership does not intrinsically involve the rejection of more vertical leadership at times. Pearce and Conger (2003) describe a type of collective, or shared leadership as "a dynamic, interactive influence process among individuals in groups for which the objective is to lead one another to the achievement of group or organizational goals or both. This influence process often involves peer, or lateral, influence and at other times involves upward or downward hierarchical influence" (p. 1). Instead, collective leadership has been described as including accountability of those (individual or multiple) taking a leadership role to those they are influencing at that time in pursuit of the collective goal (Yukl, 2009).

In collective leadership approaches then, there is a sense of shared responsibility, and the potential for a flexible, dynamic leadership model, with the clear identification of a leader or leaders, with the expertise required for the task at hand, that may change as new challenges arise. West et al. (2003) investigated the role of clarity of leadership within innovation. The study involved 3447 participants (clinical staff in primary care teams, mental health teams and cancer care teams) who responded to a self-report questionnaire asking for the respondents' perceptions of team functioning, innovation, leadership and effectiveness as well as outcomes on the Team Climate Inventory. They found a highly significant negative association with lack of leadership clarity across all teams showing that team processes were better in teams where fewer respondents reported a lack of clarity about leadership or conflict over leadership. West (2002) summarised the research evidence for the importance of clear objectives, high levels of staff participation, commitment to excellence and support for innovation, in creating conditions for high levels of team

innovation. Staff participation is encouraged by the sharing of information and sharing influence over decision making within the team (West et al., 2003).

Compassionate Leadership. The Kings Fund definition of collective leadership includes the final key feature, which involves the consistent application of authentic and compassionate leadership across an organisation (West et al., 2014). Compassion has been defined as involving the activities of: attending, understanding, empathising and helping (Atkins & Parker, 2012), all of which are elements of the activities of a compassionate leader/compassionately led organisation (West et al., 2017). The idea of authentic, compassionate leadership has gained momentum both in the UK through the work of West et al., (2003) and via the Kings Fund (Ham et al., 2011; Kings Fund, 2014; West et al., 2014; West et al., 2017), and in the US, in corporate settings, sometimes expressed as 'conscious' leadership (Dethmer et al., 2014; Kofman, 2002). In a setting, such as health care, where hierarchical leadership based on wielding positional power, or being heavily target-driven, is not a successful approach (Ham, 2014; West et al., 2014) leadership relies more on influence based on respect and authenticity (Linley et al., 2007). A compassionate, or conscious, leadership culture blends both individualistic approaches, based on positive psychology theories, focused on self-insight and identification of one's own strengths and blind spots (Linley et al., 2007); with values-based, collective aspirations where this insight is utilised for the good of the team, the organisation, and ultimately society at large (Dethmer et al., 2014; Kubátová, 2018).

West et al. (2017) describe compassionate leadership as being based on a commitment to learning, where a level of risk-taking is encouraged and where a lack of success in a new idea is viewed as an opportunity for further learning, rather than resulting in blame and threat, thus creating openness about failures and a willingness to innovate. Such leadership cultures have been found to result in better healthcare outcomes (West et al., 2014) and have been articulated as the new approach to leadership within the NHS (DH, 2008c) and more recently following the Lord Darzi Review of Health and Care (Darzi, 2018). This critiqued a continued top-down approach to health care reform and espoused a

supportive approach, encouraging staff involved in healthcare to be central to identifying and making health care improvements.

Leadership and Management Model Restructure Implementation

The existing senior leadership model, with a clear delineation and collaboration between operational and clinical roles, was an identified strength within the IST review (details of the review presented in Chapter 4). Nevertheless, the leadership team felt that, in order to maintain the improvements being undertaken within the service redesign, a review of the leadership and management structure of the service and senior and team management level would be beneficial. A review of literature (summarised above) and liaison with other IAPT services was undertaken to compare and contrast leadership models with a view to ensuring a sustainable management structure. The Kings Fund (2014) suggests a number of levels of support for NHS organisations that are struggling, including buddying, learning and clinical networks, and partnerships. Although the service was not in receipt of this type of formal support, an intention to learn from more successful IAPT services was adopted by the leadership team towards the end of Phase 2. This period of research and information gathering was followed by an informal consultation with the existing leadership team to share findings and offer recommendations and suggestions.

The leadership review began at the end of 2017 and the new structure was fully in place by September 2018, at the beginning of Phase 3. This saw the following changes:

- The development of a new leadership structure designed to promote collective leadership, based on an expertise-based, rather than purely geographical model.
- The creation of an additional Lead Therapist post.
- The creation of dedicated supervisor posts at Step 3 and Step 2 (Senior PWPs).
- The creation of a functional model splitting the leadership oversight of Step 2 and Step 3.
- Senior leadership self-directed training and peer-coaching in compassionate / conscious leadership.

The changes involved moving to an expertise-based organisational design, where teams were grouped together by common function (i.e., Step 2 and Step 3 as distinct teams within the locality area). This model tends to lead to a hierarchical, vertical chain of command with clear boundaries between teams and work streams (Robinson-Hickman, 2010). However, the service introduced horizontal overlays of cross service engagement, such as regular leadership cross-team forums, and cross-modality working groups, creating a robust organisational matrix and enabling internal collaborative working. This, and the creation of additional clinical and operational leadership roles at team levels within the service resulted in a flattening of the hierarchy and promoting collective leadership values (West et al., 2014). The aim of this model, which spread clinical leadership and line management structures, was to allow leaders (both operational and clinical) to concentrate on their strengths and expertise within a focused area of responsibility. Teaming clinical leaders and managers together aimed to encourage a culture of taking equal responsibility for the service whilst leading on separate areas. Figure 5.2 shows the previous, and redesigned, management structure of the service.

Figure 5.2

Management Structure Across the Service in Phase 1 and 2 Followed by the New Structure in Phase 3
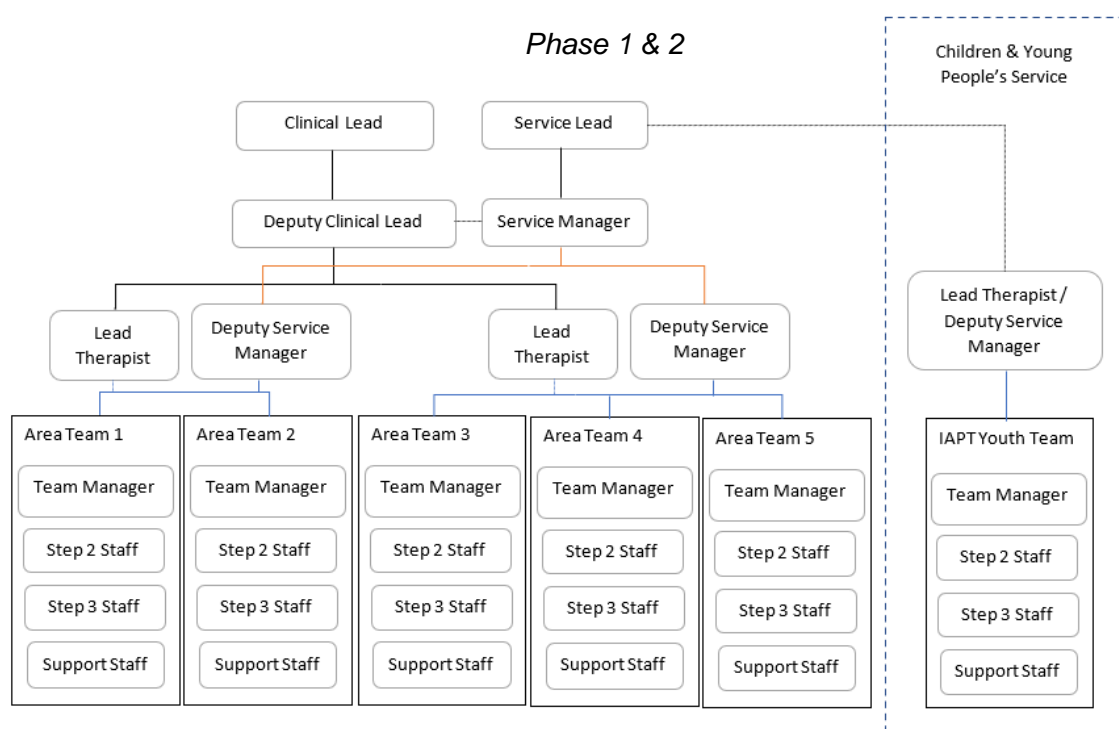
Phase 3
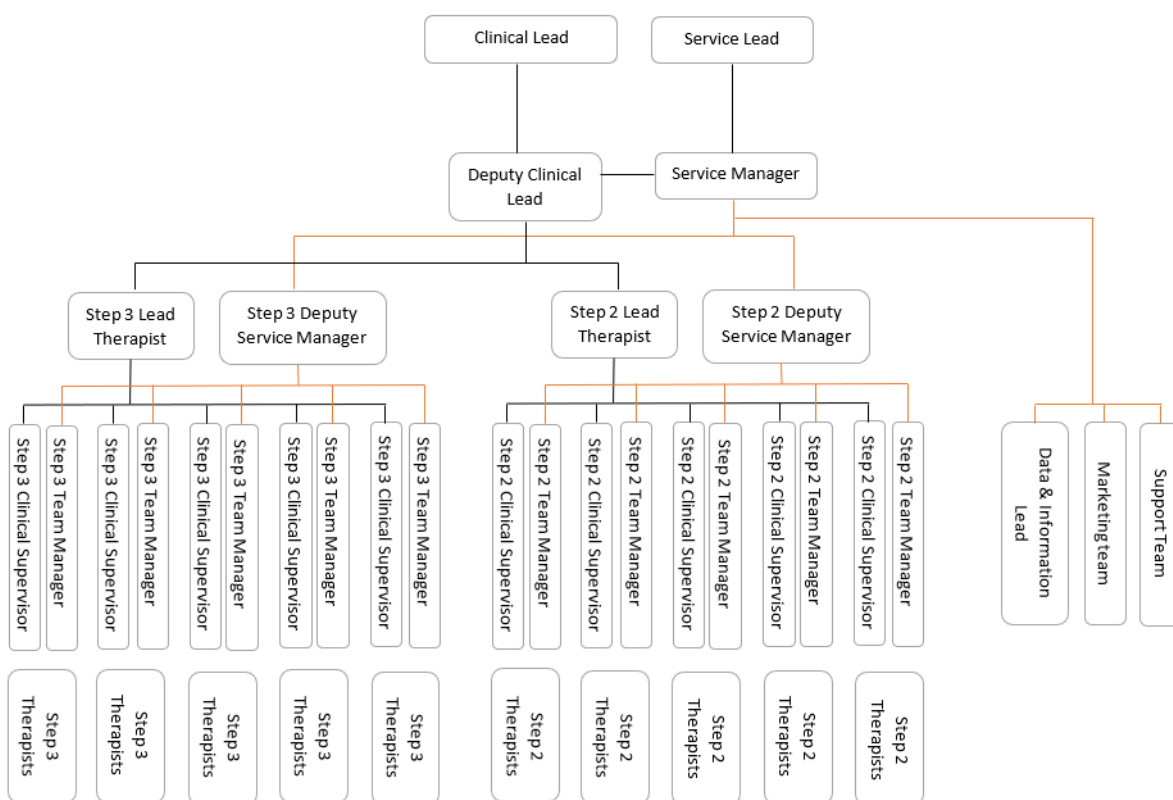
Developing and Implementing the Collaborative Intervention Package

Following the completion of the IST recommended changes, the development of the Collaborative Intervention was initiated. The purpose of this intervention was to focus on ways of reducing Step 3 therapist variability whilst improving patient outcomes utilising evidence from the therapist effect literature, as well as staff suggestions. This process benefitted from the leadership review outlined above, as the move to a more collaborative service context resulted in the leadership team as a whole being supportive of staff involvement in the development of a 'bottom up' intervention. The leadership review also supported an increased focus on staff wellbeing and staff support, with the leadership and management team committing to learning about compassionate, conscious leadership (Dethmer et al., 2014; West et al., 2014), initially through training and peer coaching activities during this Phase.

A systematic review of the therapist effect literature as it relates to explanatory or moderating therapist factors was undertaken to initially inform ideas for this bespoke

intervention (see Chapter 2). Following this literature review, a series of focus groups was undertaken to refine the final bespoke intervention package.

Focus Groups. Because the current study utilised the live and dynamic nature of an NHS IAPT service, it was important to involve staff in the development and implementation of the key study intervention in the final Phase. This contrasted with the perceived, and actual, 'top-down' nature of the changes made by the service during Phase 2 and was fully supported by the leadership team who were aware that a change of approach and culture in the service was needed, based on staff surveys (see Chapter 3).

A series of focus groups were set up early in the development of the Phase 3 intervention package, comprising in total 20 Step 3 therapists over three sessions which took place in June and July 2018 (see Appendix F). The purpose of the focus groups was to help develop the intervention that would be implemented in Phase 3. The focus groups therefore focused on two aspects:

1. Introducing a summary of the ideas from the research literature in relation to therapist effects, particularly where these linked to explanatory factors. Specifically, the following factors were presented to therapists within the focus groups as being likely to be associated with more effective therapists:
   - Higher resilience levels
   - Lower therapist stress
   - Quality of the therapists learning environment
   - Professional self-doubt and self-compassion
   - The nature of learning (e.g., deliberate practice)
2. Gathering feedback from therapists about what they felt contributed to 'good' therapy, the things that impacted on their own therapy practice and contributed to their patients' outcomes and what they would like to see in a package of therapist support.

The focus groups were audio recorded and reviewed to identify themes both within and across the groups. This was an informal process of thematic analysis with the purpose

of identifying areas important to staff and involving staff in the development of the intervention package from the first stages. Taking place following a period of change within the service, all groups highlighted the stressful nature of the working environment and lack of time and space to think as factors that were preventing them from being as effective as they could be. Interestingly, some of the assumptions disproved in the literature, such as experience leading to better outcomes, were mentioned by therapists as perceived contributing factors to effective therapy. The following themes were generated during the focus groups, in response to the questions asked:

What contributed to good and poor outcomes or pieces of therapy work that you thought did/did not go well?

- Feeling stressed in the workplace (worse outcomes)

- Not having enough time to reflect on and plan for sessions (worse outcomes)

- When I feel organised and prepared things tend to go better (better outcomes)

- When I feel connected to my patient and that things 'flow' (better outcomes)

- Feeling confident and that I know what I am doing (better outcomes)

- Focusing on interpersonal aspects of the therapy (better outcomes)

- Being more experienced (better outcomes)

How could you be supported better/ideas for package of intervention?

- Less change in the system/service so that stress levels can reduce

- More time to reflect on therapy work, not more micro-management

- Learning from each other – more time to get together in groups

- Group therapy for staff to improve staff wellbeing

- Fewer meetings, less disruption to therapy work

- Focussing on skills development, more training opportunities

Is there anything that might get in the way of the ideas you have generated?

- Not enough time

- Scepticism about how meaningful support might be and how long it will be committed to by management

Conclusions. It was clear from the focus groups that there remained a high level of change fatigue and scepticism about the motivations of management as a result of the prior period of change within the service. Therapists across all focus groups requested more time to *"think"*, "*prepare*", "*learn*" or "*reflect*", despite the increase in individual supervision introduced 12 months prior. There were numerous mentions of wanting more time to learn or reflect with peers. It was evident that the conclusions from the therapist effects literature was not known to therapists, who were surprised to learn that years of experience did not result in better patient outcomes. There was, however, sympathy with the findings suggesting that burnout and resilience may be factors in explaining therapist effects, perhaps consistent with the experiences of the focus group staff at this time.

This Phase of the study aimed to initiate a therapist-level intervention, informed by literature and the collaborative involvement of therapists, that could be implemented at a service-wide level. A systematic review of literature related to the explanatory factors associated with therapist effects (see Chapter 2) had highlighted several potential avenues:

- Consideration of the content and quality of supervision rather than simply the quantity

- The structured use of patient feedback to therapists about their progress, possibly in relation to their view of the therapy alliance, as well as symptom outcomes, to aid self-reflection/reflective practice

- The use of deliberate practice of therapy skills and tasks, as needed by the individual therapist based on their strengths and deficits

- The use of the above elements within supervision to cultivate an atmosphere of compassionate critique of therapists' work

- Interventions that seek to increase therapist wellbeing and resilience

The focus groups provided further suggestions as outlined in the themes above. Staff were particularly interested in increasing the time they had in groups to reflect on their

practice, so this was incorporated into the package from an early stage. In order to maximise the learning opportunity of these groups, the research team agreed that asking therapists to use deliberate practice within the groups would provide a suitable framework for learning, based on the literature. In addition, the themes of change fatigue and high stress coming from the therapists suggested that a focus on staff wellbeing within the intervention may complement the practice groups, aiming to minimise the potential for burn out (Delgadillo, Saxon et al., 2018; Pereira et al., 2017; Saxon & Barkham, 2012), and to provide a balance within the intervention package, as suggested both by staff in the focus groups, and in the deliberate practice literature – that is, the need for rest as well as practice (Ericsson, 2006). These two elements – the use of deliberate practice reflective groups and staff wellbeing activities – were identified as the key features of the therapist level intervention package for Phase 3.

Summary of Deliberate Practice Literature

The concept of deliberate practice (DP) was introduced by Ericsson et al. (1993) and further developed within many fields (e.g., sports, music) as a means of developing expertise through a specific type of practice (Ericsson & Lehmann, 1996; Ericsson & Pool, 2016). There has been some controversy as to what the defining characteristics of DP are, but there is largely a consensus that it must include: "*(1) individualized learning objectives; (2) ongoing feedback regarding performance and learning; (3) involvement of a coach; and (4) successive refinement through repetition most often conducted alone*" (Miller et al., 2020, p.11; Ericsson & Lehmann, 1996). The launch of DP into the arena of psychotherapy has produced much interest (Miller et al., 2007; Tracey et al., 2014) as well as scepticism (Clements-Hickman & Reese, 2020), but remains an area of growth in the literature. At the time of the scoping review of DP, two key studies had been published on its application within psychotherapy practice – Chow et al. (2015) and Goldberg, Babins-Wagner et al. (2016).

As discussed within the systematic literature review (Chapter 2), Chow et al. (2015) investigated the role of therapist DP on patient outcomes and found that the amount of time

spent improving therapeutic skills was a significant predictor of outcomes, although no one single activity led to better outcomes than any other. There was a suggestion that the individual review of therapy recordings may provide a strong likelihood of learning, as this was rated as requiring more mental effort by more highly effective therapists. The authors concluded that therapists should target the following areas to improve their outcomes: improving outcomes of at-risk cases; creating social experiments in naturalistic settings to test, recalibrate, and improve empathic accuracy; practicing fundamental therapeutic skills (e.g., rehearsing difficult conversations); using role play of case vignettes; setting aside time to reflect on therapy sessions; and planning ahead with treatment.

In a longitudinal observational study of a health care agency seeking to improve their patient outcomes using ROMs and DP, a significant improvement was observed over the 7-year period of the study (Goldberg, Babins-Wagner et al., 2016). The agency introduced ROMs across the service, initially inviting therapists to use the measures, and after four years, introducing the use of ROMs (though not the outcome of ROMs) as a performance measure. Therapists were provided with a two-hour consultation with an external DP consultant each month to focus on cases that were not progressing and to develop a plan to better meet the needs of the patient. The agency aimed to create a culture of feedback and improvement, and this developed increasingly over time. Multilevel modelling was used in the analysis of the data from the agency, and a significant improvement in outcomes for patients receiving therapy at the agency over the period was observed. This improvement was not accounted for by other patient or therapist variables such as patient intake severity, or therapists initial experience level. The authors suggest this study as a potential model for improving outcomes in mental health care settings and the study provided a helpful demonstration of potential service level improvements that could inform the Phase 3 intervention.

Therapists within the service were already receiving additional feedback on their patient outcomes and own performance in the form of individual patient improvement/deterioration ROMs to guide supervision, and their own recovery and reliable

improvement rates for their overall caseload, introduced as part of the Phase 2 interventions. This change had been largely experienced as being management-led and more related to performance management than clinical quality. Accordingly, further use of ROMs as feedback was not deemed an appropriate or useful avenue for clinicians within this context. In support of this decision, it has been noted that routine outcome feedback alone is not sufficient to produce improvements/expertise, but *how* this feedback translates into skills development, change to practice or DP itself is equally, or possibly more, important (Miller et al., 2015; Tracey et al., 2014). Given that the consultation with therapists appeared to suggest that therapists wanted more time to reflect on their practice, ideally with peers, it was decided that utilising a *deliberate practice-lite* model might be both acceptable to therapists and supported in the literature. The use of DP, reflection and learning as both a concrete package of support and cultural shift for the service, was felt to hold potential.

Service Constraints

The challenge of how to make this intervention practical and affordable for the service, in the context of an NHS/government funded service with no additional financial resource, was considered by the research team and clinical and management representatives within the service. There was understandable anxiety about the impact of implementing some of the package outlined by Goldberg, Babins-Wagner et al. (2016) as this had initially resulted in staff resignations within their cohort. This was something that the service was keen to avoid, within the context of 18 months of significant service change and low morale reported in the staff group. However, this study was viewed as a positive example within the context of the other literature considered, and the staff consultations undertaken. It was observed that to make a meaningful change in the service culture and promote the value of active learning experiences, this would involve significant support from managers and supervisors (Miller et al., 2017). A benefit of the research taking place within, rather than external to, the service meant that there was significant leadership and management commitment to the study, which was viewed as a strength of the research design.

Deliberate Practice-Lite (DP-Lite) Package

The research team decided on the following as a practical *DP-lite* package as part of the Phase 3 bespoke intervention:

- Monthly 1-1.5hr deliberate/reflective practice peer groups

Therapists were invited to form small peer groups of 3-5 people. Mixed modality groups were encouraged but not mandated. These groups were specifically for therapists to practice micro-skills within a group setting, benefitting from feedback from the group in a way that was most supportive to the therapist.

- Identification of individualised goals and micro skills within individual supervision.

Supervisors were provided with an additional information/training session on the principles of DP, with a focus on the identification of micro-skills. They were encouraged to support their supervisees to identify particular micro-skills that they and the supervisee felt would benefit from further development. These micro-skills would then become the focus of DP-lite peer group sessions over the coming weeks/months, until the supervisee felt ready to move on to a new skill. Utilisation of feedback from listening to therapy tapes (a pre-existing routine, though infrequent, activity in one-to-one supervision in the service) and ROMs was encouraged within the training and at monthly supervisor meetings held within the service.

To support the implementation of the above elements of the package, the following additional elements were included across the 18-month period:

- Initial 2hr DP training session for each team of therapists at point of implementation

This session was delivered by the primary researcher and other members of the research team. The session included a summary of the literature underpinning the development of the package, and more detailed information about how DP can be applied in psychotherapy practice. Clips of a video interview with Tony Rousmaniere were used, including examples of DP in action (Psychotherapy Expert Talks, 2017).

- 2hr follow up review and training session for each team 4-6 months following implementation.

Each team was visited by the primary researcher or embedded co-researcher for a follow up session between 4 to 6 months following the initial training session. The session focussed on obtaining feedback from therapists about whether they had used the ideas of DP; obstacles or reasons for not using DP; examples of using DP; reflections or feedback on DP and the new groups. The session also included a refresh of the rationale and purpose of DP.

- 2hr top-up training session via webinar for the whole service 12 months following initial implementation

A webinar was conducted by the primary researcher and embedded co-researcher 12 months following the initial implementation of the DP-lite/reflective groups. The webinar recapped the rationale and purpose of DP and primarily focussed on the practical application of DP in action. The researchers provided examples of the application of DP in their own practice as well as utilising feedback from the attendees to troubleshoot and share best practice.

- Research attendance at supervisor meetings every 3 months

The primary or co-researcher would attend a pre-existing clinical supervisor meeting approximately every 3 months to specifically discuss the DP-lite package, provide any additional information and answer questions that supervisors may have. This functioned to troubleshoot any difficulties and share learning across the service.

It should be noted that there were distinct elements of the package that did not meet the general consensus definition for pure DP (Ericsson & Lehmann, 1996; Miller et al., 2020, p.11):

(1) individualized learning objectives – the importance of developing clear personal goals was included in the training sessions for therapists and supervisors, though this was not monitored during the study other than through self-report feedback at the review sessions.

(2) ongoing feedback regarding performance and learning – one of the reasons for having the DP time allocated in peer groups was to enable therapists to obtain feedback

from their peers on their skills. In addition to this direct feedback, the feedback initiated in Phase 2, provided as part of supervision through ROMs and therapist outcomes of IAPT recovery and reliable improvement measures continued in Phase 3. However, this feedback was not directly linked to DP.

(3) involvement of a coach – supervisors were involved in supporting therapists to identify goals and micro-skills. However, they were not present to give feedback during the DP-lite groups. Supervisors were senior clinicians within the service, but this did not equate to them being expert therapists or experts in psychotherapy (a distinction discussed by Norcross & Karpiak, 2017), nor indeed necessarily having superior patient outcomes compared to their supervisees.

(4) successive refinement through repetition most often conducted alone – the importance of repetition and refinement of skills was included in the training and follow-up sessions. However, therapists were provided with DP time in groups, rather than individually.

In addition, it should be noted that, due to the purpose of the research to be, in part, an investigation into the applicability of the introduction of such a package into a service through a bottom-up collaborative approach, none of the therapists, supervisors or DP-lite groups were required to evidence how or if they used DP either within the groups or individually. The purpose of the feedback sessions was to gauge the overall up-take of the ideas within a naturalistic setting, but therapists were not asked for specific information in relation to the amount to which they were using or adhering to the practices as taught.

Stepped Wedge Design and Method

A stepped wedge design was used for the implementation of the DP-lite element of the Phase three intervention package.

Stepped Wedge Design. Stepped wedge trials have been used within health care service evaluations since 1987 (Gambia Hepatitis Study Group), and have been viewed as an alternative to parallel cluster trials (Hemming et al., 2015). The design involves an initial control period with all clusters not receiving the intervention, followed by a gradual (stepped) introduction of each cluster within a trial to the intervention condition. Data is collected

throughout this process, until all clusters have moved to the intervention condition, and for a period of time afterwards.

Hemming et al. (2015) described contexts which might lead to a rationale for a stepped wedge design. This design is particularly useful when interventions are being tested in a live service setting, particularly where there is pressure to implement a change, but without satisfactory prior levels of experimentation or testing. The design also allows the potential for adaptations to be made over the course of the phased implementation, therefore improving the intervention based on learning from each cluster. Other reasons for using such a design are when there might be logistical implications in attempting to implement a particular intervention across a large system, or where there are geographical constraints.

A stepped wedge approach to the implementation of the DP-lite intervention was chosen for a number of reasons, but with the overall aim of maximising the effectiveness and learning from the DP framework being introduced. Due to the size of the service, and the need to introduce the DP framework to clinicians in a meaningful way, it was not felt to be practical to introduce the DP training to the whole staff group in one session. Instead, training was provided to smaller subsets of therapists, where a more interactive session could be utilised. Staggering these sessions allowed feedback and learning to inform the next introductory training session for the next group. It also allowed time for the clinical supervisor training and update sessions to be undertaken. The 'cluster' that a therapist belonged to was also available as a predictor variable, allowing for the variable: 'amount of time exposed to DP-lite condition' within the subsequent analysis, and other comparisons between different cluster groups.

The service comprised five geographically based teams who were all included in the study. The implementation of the DP-lite aspect of the package was introduced across the service in three clusters (Figure 5.3). The first cluster included one 'early implementor' team, with clusters two and three comprising two teams concurrently. Thus, by the introduction of cluster three, all teams across the service were included in the DP-lite package.

Figure 5.3

DP-Lite Stepped Wedge Implementation Timeline

Year		2018							2019										2020		
Month	09	10	11	12	01	02	03	04	05	06	07	08	09	10	11	12	01	02	03
Cluster 1 Intervention period		X				X				X									X
Cluster 2 intervention period						X				X									X
Cluster 3 intervention period										X		X							X

X = introduction or review meeting with staff in cluster

▨ = intervention condition period

☐ = control condition period

The control period for all clusters began with the implementation of an additional hour of skills development time, under the title of *peer reflection/skills session* and commenced service-wide from September 2018. An early implementor team was chosen to be the first cluster for the first stage of the stepped wedge implementation, based on their geographical location, which was reasonably self-contained and disconnected from the other teams; and the staff team attitude, which was generally positive about change and development.

At each introduction session, the primary researcher presented the principles of therapist effects, an overview of the literature and an introduction to DP. They were joined by an embedded member of the research team who was a member of staff that had asked to be involved in the research at an early stage. This member of staff was a counsellor and had an interest in DP within psychotherapy. A member of the University research team supported the primary researcher and staff member to represent the academic and research arm of the implementation. Generally, the University member attended in person, although for the early implementor training session they attended via a video-conferencing facility. Regular reviews were structured into the implementation of DP-lite groups across the service, as shown in Figure 5.3.

Cluster One. The first DP training session for the early implementor team took place in October 2018 and was well received. There was some debate regarding how the team would divide into smaller groups and the benefits or otherwise of being in modality specific or cross-modality groups. Though the cross-modality nature of DP was stressed by the research team, the therapists were not given mandatory instructions on how to form their groups, and were encouraged to come to a decision as a peer group through a collaborative process.

A review was conducted by the embedded co-researcher on the research team. Two main reflections/learning points emerged from this early review:

- There were differing views as to whether DP would be helpful, some were using it, some had chosen not to. The majority had chosen to use the ideas.

- Therapists reported finding it difficult to stay focused on the task at hand in the DP-lite groups, and the suggestion of having a recording form was raised.

Cluster Two. Two teams were included in cluster 2 (the second implementation step), though the introductory session for each team was conducted separately to maximise the potential for questions and discussion. Both sessions took place in February 2019 and included the same information presented to cluster 1, described above, with some minor amends to the order and pace of the session based on learning from the previous session. The reviews for the cluster 2 teams took place in July 2019 and resulted in the following feedback:

- Difficulty keeping to the task of the groups; easy to fall into comfortable/familiar habits more akin to group supervision/case discussion.

- As not all staff had been present at roll-out session, there was a variety of levels of understanding of the concepts.

- Consistent group membership proved difficult due to annual leave. Larger groups had been formed and therapists were satisfied that these were manageable and useful.

- Tension between individual learning and group learning was identified, that is, some therapists were unsure if they should be identifying micro-skills that the whole group would benefit from practicing, or if they should just be focussing on their individual learning needs.

- Misunderstanding as to how quickly new micro-skills had to be developed, with some therapists reporting that they felt they should practice a new micro-skill every one or two sessions.

The reviews focused on real world examples of DP in action with an emphasis on picking one micro-skill for practice over a period of several months rather than a different practice each time. Use of a recording form was felt to be useful and this was circulated to all teams (including cluster three teams, who had already had their roll out session by this time). The recording form was developed with members of the early implementor team who had initially suggested it. Examples were added and a *Hints and Tips* sheet also circulated at the same time, summarising some of the learning points generated by the first three team reviews (both the recording form and 'hints and tips' sheet can be found in Appendix G).

Cluster Three. The implementation of the intervention for cluster three occurred in June 2019, prior to the cluster 2 review sessions, and in hindsight may have benefitted from some of the learning from these sessions. There was more emphasis on the practical approach to DP in these roll out sessions, though this would have been emphasised more if it had occurred after the review sessions.

Refresher Training. A live staff training webinar was undertaken on 30th October 2019, 12 months after the initial roll out in the early implementer group. Therapists across all teams participated in the webinar, which included a summary of the evidence base in therapist effects literature, a recap on the theory and application of DP, and a focus on practical examples of using DP in the groups.

Collaborative Intervention: Continuing Professional Development Training and Staff Wellbeing Sessions

The other aspect of the Phase 3 Collaborative Intervention package was a focus on staff wellbeing. Eight events were organised throughout the 18-month period based on staff feedback and suggestions (Figure 5.4). A combination of Continuing Professional Development (CPD) events and wellbeing events were included. The four CPD events within the package were not dissimilar to those provided in Phase 1 and Phase 2, as this was a standard part of the annual training opportunities offered to staff to support their development. A new aspect to the schedule of events was the inclusion of four staff wellbeing sessions of either a full or half day duration. These were launched in May 2019 with a full day of Wellbeing activities, including a session on resilience and self-care, led by staff members designated as Wellbeing Champions. The half day wellbeing sessions were led by each team separately to allow teams to choose their own activities. The types of activities chosen by the teams included: talent show; arts and crafts activities; guided walks; shared lunches; yoga; mindfulness.

Figure 5.4

Continuing Professional Development and Staff Wellbeing Events Timeline

STUDY PHASE THREE: CPD & staff wellbeing Sessions																		
Year	2018			2019												2020		
Month	10	11	12	01	02	03	04	05	06	07	08	09	10	11	12	01	02	03
CPD Event				▓	▓						▓		▓					
Wellbeing Event								■		■			■			■		

Summary

The naturally occurring programme of service improvements developed by IST and the senior service management team provided the context for the Phase 2 Directive Intervention package as described in Chapter 4. As the service consolidated these changes, a period of relative stability provided the opportunity to initiate consultation and development for the Phase 3 Collaborative Intervention package, maximising staff involvement within the

research project. This package also included a leadership and management restructure and focus on compassionate/conscious leadership approaches. A stepped wedge approach to the implementation of the *DP-lite* part of the Phase 3 package allowed a more practical and collaborative introduction to this new framework of learning, adding to the staff wellbeing events delivered across the 18-month period. The two discreet intervention packages; directive, externally-driven service improvements in Phase 2; and a bespoke, research-informed, collaboratively developed package in Phase 3, allowed analysis of the impact such interventions may have on therapist variability and clinical outcomes across a service. The unique position of the researcher being embedded within the service leadership team provided the additional information regarding the nature of the implementation of these packages, in addition to the content of the interventions themselves, enriching the understanding of the context for the intervention Phases of this project.

Chapter 6

Methods

Overview

This chapter presents the methods undertaken to investigate the research questions identified in Chapter 2. The entire research project spanned a 4.5-year period of data collection and intervention development and implementation. The study context and setting have been presented and discussed in Chapter 3 alongside the Baseline Phase information, and the development and implementation of the intervention Phases have been presented in Chapters 4 and 5. This chapter will therefore focus on the collection and preparation of data samples and the statistical methods used in the analysis.

Study Structure

The study comprised three distinct 18-month long Phases, in which therapists were working in specific service conditions, as set out in Chapters 3, 4 and 5. The first Phase provided baseline patient outcomes and a baseline estimation of the therapist effect when the service conditions were as standard, or *'service as usual'*. The second Phase was the implementation of the *Directive Intervention*, a period when service conditions were characterised by top-down change, a directive management style and service transformation or redesign. This Phase provided patient outcomes and estimations of therapist effects immediately prior to the implementation of Phase 3, the *Collaborative Intervention* Phase.

Across these three Phases, two samples were created: the *Core Therapists* sample and the *All Therapists* sample:

- The Core Therapists sample comprised therapists that had a minimum of 10 patient episodes within *each* study Phase (i.e., they appear in each Phase dataset over the 4.5-year study period). This allowed direct comparison of the same therapists across the three Phases.

- The All Therapists sample comprised *all* eligible therapists in each Phase (whether or not they appeared in multiple study Phases). This sample was used to investigate the

generalisability of findings from the Core Therapists model in a wider sample of
therapists and patients.

The two samples (Core and All Therapists) were divided into three data subsets based on
patients treated during each of the three 18-month Phases. Figure 6.1 describes the study
structure and data samples.

Figure 6.1

The Study Structure
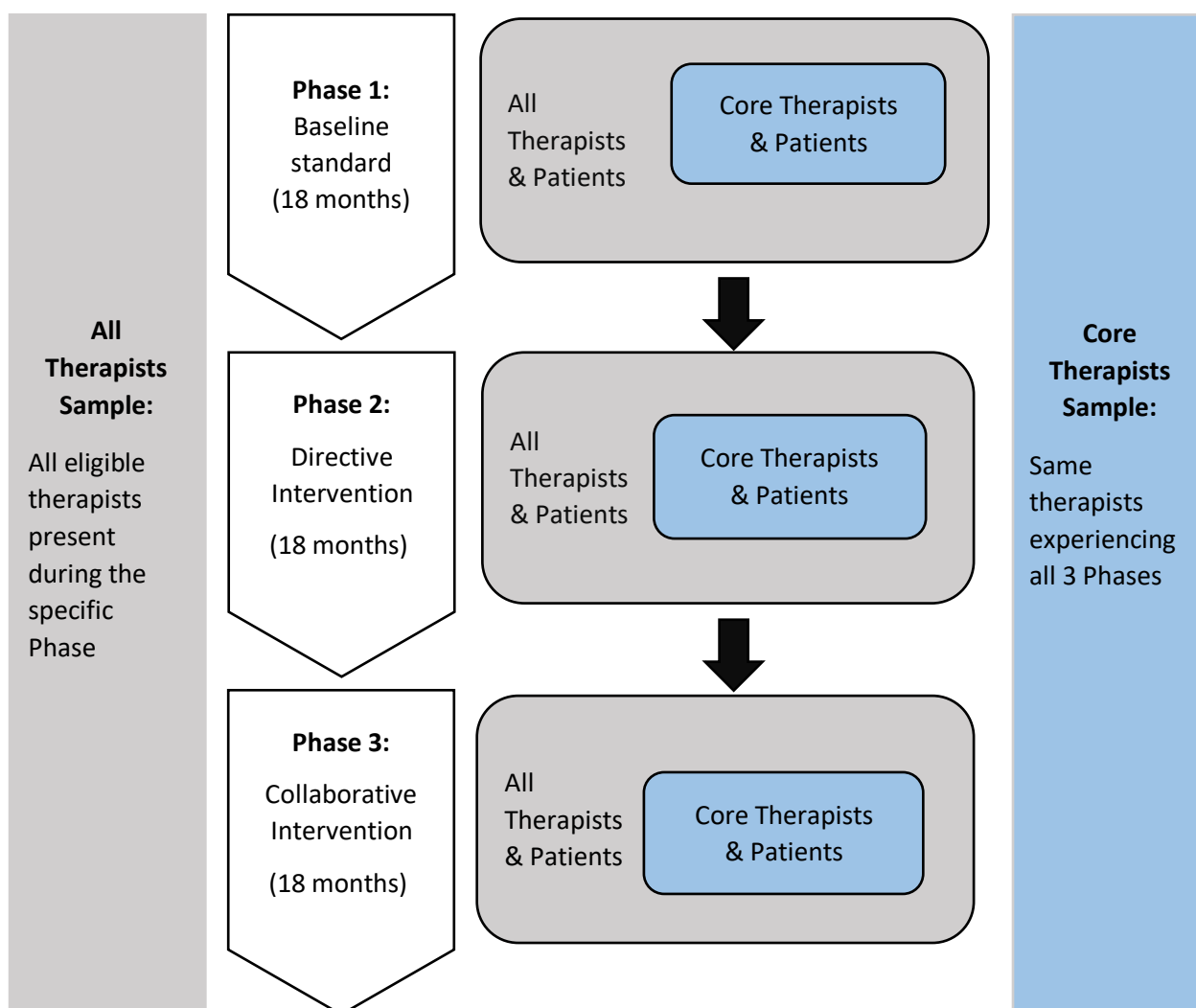
Figure 6.1 illustrates the two main samples: All Therapists shown in grey, and Core
Therapists shown in blue. Each sample is broken down into the three Phases of the study as
shown in the central boxes. The nature of the Core Therapists sample being contained
within the All Therapists sample is depicted by the blue boxes within the grey boxes for each
Phase.

Study Data

Data Collection

The dataset used for the analysis was extracted from the clinical patient management system, IAPTUS, (www.iaptus.co.uk), which is software that is used in two thirds of IAPT services across England. This system allows the clinician and administrative team to record all patient information, including demographic information (gender, sexuality, age, location, employment status, ethnicity, primary spoken language, physical health status including disabilities), session clinical notes, clinical outcome measure scores and other relevant metrics (including suicide risk level, use of psychotropic medications, number of previous therapy episodes, satisfaction feedback). The system allows for the extraction of key data at a patient, therapist or team level, on all metrics recorded within the database. Ethical approval for Phase 1 and 2 of the overall study was granted by the East of England branch of the Health Research Authority (Ref: 17/EE/0251; see Appendix C) and for Phase 3 by Health Research Authority and Health and Care Research Wales (Ref: 18/NS/0104; see Appendix E).

Preparation of Study Data

All recorded clinical information (excluding clinical notes and patient or therapist identifiable information) for each clinical contact was extracted from the patient recording system, checked and prepared to allow each patient episode to be summarised as a line of data, nested under the therapist/s who treated them.

Several exclusions were applied to the dataset:

- Patients had to have at least three contacts with the service to be deemed to have been in receipt of a minimal dose of therapy. This decision was made based on knowledge of the service model where many patients are given an extended two session assessment. As it was not possible to distinguish these extended assessments from therapeutic interventions within the dataset, all patients with two contacts or less were excluded from the dataset.

- Patients had to have started (first therapy contact) and completed (final contact) their therapy episode with the specific therapist, within the 18-month data period for each Phase.

- Patients who did not have initial and final clinical measures recorded on PHQ-9, GAD-7, WSAS, and Phobia Scale or missing data on IMD quintile or discharge reason were excluded.

- Patients on the rapid brief intervention pathway during Phase 2 (created to address longest waiting patients) were excluded.

- Patients treated by locum or subcontracted therapists who were not embedded in the service, and therefore not subject to the intervention/service conditions, were excluded.

- Within each Phase dataset, therapists, and patients of therapists, who had treated less than 10 patients in that specific Phase period were excluded (see Sample Size section below).

- For the Core Therapists datasets, only therapists, and patients of therapists, who had treated a minimum of 10 patients in each Phase were included in the dataset (i.e., the therapist met the inclusions for each Phase and was therefore included in the datasets of all three Phases).

Trainee therapists were included in the datasets, as it is standard for services of this size and type to have an ongoing proportion of therapists in training within the staff group. As the number of trainees was stable across the study period, to ensure the generalisability of the results, trainees were not excluded from the study if they met other inclusion criteria. To check the potential impact of trainees on the final results, the primary analysis was also undertaken on a subset of the data with trainees removed, and the results of this are also reported.

Missing Data

Although IAPTUS allows for the collection of a wide range of patient variables (see above), not all fields were routinely completed by the service. The following patient variables

were therefore excluded from analysis due to data quality issues: medication, sexuality, employment status, ethnicity, primary spoken language, physical health status. A therapist variable that was collected but was excluded from the dataset due to levels of missing data was: years since qualification (as a measure of therapist experience).

Sample Size

Currently there is no recognised method for a priori calculations of sample size for multilevel models (Snijders & Bosker, 2011). This is due to there being too many unknowns in advance, particularly the effects on outcome of the different levels in multilevel models and the relationships between levels. In multilevel modelling, consideration needs to be given to the sample size of each level (i.e., the number of therapists and the number of patients per therapist and overall). Maas and Hox (2004, 2005) suggested that the number of therapists was most important in a reliable estimation of the size of therapist effect. They conclude that a sample comprising over 100 therapists would produce the most reliable estimates although 50 may be acceptable. Schiefele et al. (2017) recommended a minimum patient sample of 1200 with some flexibility regarding the number of therapists and patients per therapist required to achieve the 1200 total.

In the current study, in addition to wanting to include as many therapists as possible, it was also important to ensure each therapist had a sufficient number of treated patients to provide a reliable estimation of the therapists general practice performance, and in providing reasonable reliability of the patient level variables. To balance these considerations, after applying the patient exclusions above, only therapists who had treated 10 or more patient episodes were included in the main analyses. To assess the reliability of the therapist effects found, a sensitivity analysis was undertaken which included all Core Therapists with one or more patients. However, as this study is limited to therapists in a single service, sample sizes are unlikely to meet recommendations. Therefore, analyses may be under-powered to produce the most reliable estimates of therapist effects.

Therapist Samples: Core Therapists and All Therapists

The primary analysis for each Phase included only those therapists who were constant across all three Phases (the Core Therapists sample). These were the therapists who treated patients across the full 4.5 years of the study, were present in the Baseline Phase, and were exposed to both interventions. This allowed comparisons to be made between therapist effects and therapy outcomes in each Phase of the study for the same group of therapists over a 4.5-year period.

For each Phase, a further analysis was undertaken, including all therapists in the relevant Phase (with sufficient numbers of cases, i.e., 10) regardless of whether they were present in another Phase. This All Therapists sample allowed for the assessment of the impact of the interventions on the wider cohort of therapists in each Phase. These therapists may have experienced one or more of the service conditions, and this dataset more realistically represents the outcomes of a service where the impact of therapists as new starters and leavers on outcomes cannot be controlled. In addition, by including a larger sample of therapists in each Phase, more reliable estimates of therapist effects can be produced, representing the overall therapist effect in the service at that time. The development of the datasets, across all three Phases, is shown diagrammatically in Figure 6.2.

Figure 6.2

Core & All Therapists Datasets CONSORT flow diagram

Phase 1   Phase 2   Phase 3

| Assessed for eligibility (n patients = 6200 n therapists = 129) | Assessed for eligibility (n patients = 5769 n therapists = 156) | Assessed for eligibility (n patients = 4164 n therapists = 191) |

Excluded (n patients = 4218 n therapists = 48)
- *<3 contacts & missing data (3843 patients, 8 therapists)*
- *<10 patients completed (64 patients, 33 therapists)*
- *Locum/subcontracted (311 patients, 7 therapists)*

Excluded (n patients = 3542 n therapists = 76)
- *<3 contacts & missing data (3164 patients, 17 therapists)*
- *<10 patients completed (238 patients, 52 therapists)*
- *Locum/subcontracted (140 patients, 7 therapists)*

Excluded (n patients = 1897 n therapists = 117)
- *<3 contacts & missing data (1174 patients, 83 therapists)*
- *<10 patients completed (134 patients, 20 therapists)*
- *Locum/subcontracted (589 patients, 14 therapists)*

Included in All Therapists dataset
(n patients = 6476; n therapists = 126)
Accounting for duplicate therapists

| Phase 1 All Therapists dataset (n patients = 1982 n therapists = 81) | Phase 2 All Therapists dataset (n patients = 2227 n therapists = 80) | Phase 3 All Therapists dataset (n patients = 2267 n therapists = 74) |

Excluded (n patients = 3103 n therapists = 91)
- *Therapist did not appear in all phases*

Included in Core Therapists dataset
(n patients = 3373; n therapists = 35)

| Phase 1 Core Therapists dataset (n patients = 930 n therapists = 35) | Phase 2 Core Therapists dataset (n patients = 1226 n therapists = 35) | Phase 3 Core Therapists dataset (n patients = 1217 n therapists = 35) |

The Core Therapists Sample. The initial Phase 1 dataset (prior to exclusions being applied) comprised 6200 patients who were referred and completed their Step 3 therapy within the service between October 2015 and March 2017, as well as 129 Step 3 therapists. The initial Phase 2 dataset contained 5769 patients, as well as 156 therapists and the initial Phase 3 dataset comprised 4164 patients and 191 therapists. Following the exclusions shown above, retaining all therapists who had treated a minimum of 10 patients in each Phase, and excluding any therapists not appearing in all Phases, yielded the Core Therapists datasets comprising 35 therapists and 930, 1226 and 1217 patients for Phase 1, Phase 2 and Phase 3, respectively. The following features of the datasets were observed (Table 6.1).

As shown in Table 6.1, more patients were retained in Phase 2 and Phase 3 samples, and a slightly higher proportion of patients received CBT compared to counselling interventions in each Phase, with Phase 1 being most evenly split between these two treatment types. Between 31% and 34% of patients were male across the Phases, and the average age ranged from 41 to 46 years. Of the 35 therapists in the datasets, 69.2% were female and 63% of the sample were CBT therapists, with the remaining 37% being counsellors. Six of the Core Therapists (CBT only) were in training during part of Phase 1, comprising 17% of the total therapists in the sample.

Table 6.1

Features of Patients in Core Therapists Sample

Patients	N	Modality of treatment		Gender		Age (years)	
		CBT (%)	Counselling (%)	% Male	% Female	Mean (SD)	Range
Phase 1	930	471 (50.6)	459 (49.4)	34.1	65.9	45.8 (15.25)	17 - 94
Phase 2	1226	691 (56.4)	535 (43.6)	30.8	69.2	41.5 (15.93)	16 - 94
Phase 3	1217	650 (53.4)	567 (46.6)	33.2	66.8	40.7 (15.69)	16 - 90

The All Therapists Sample. The Phase 1 dataset (prior to exclusions being applied) comprised 6200 patients who were referred and completed their Step 3 therapy within the Phase period, treated by 129 Step 3 therapists. In Phase 2 this initial dataset comprised 5769 patients and 156 therapists, in Phase 3, 4164 patients and 191 therapists. Following initial exclusions, retaining all therapists who had treated a minimum of 10 patients, the final datasets comprised 1982, 2227 and 2267 patients, in Phase 1, 2 and 3, respectively. Numbers of therapists retained in Phase 1, 2, and 3 were 81, 80, and 74 respectively. Table 6.2 shows the patient features of the datasets. As with the Core Therapists datasets, a higher proportion of CBT was accessed compared to counselling, and to a larger degree in the Phase 2 and 3 datasets than for the Core Therapists. The split of gender of patients remained consistent, with more female patients being treated. The average age of patients ranged from 39 to 44, which was slightly less than the Core Therapists samples, although the pattern across Phases is the same.

Table 6.2

Features of Patients in All Therapists Sample

Patients	N	Modality of treatment		Gender		Age (years)	
		CBT (%)	Counselling (%)	% Male	% Female	Mean (SD)	Range
Phase 1	1982	1017 (51.3)	965 (48.7)	35.1	64.9	44.3 (15.80)	17 - 94
Phase 2	2227	1416 (63.6)	811 (36.4)	33	67	39.7 (15.96)	16 - 94
Phase 3	2267	1583 (69.8)	684 (30.2)	32	68	38.9 (15.31)	16 - 90

Table 6.3 shows the therapist descriptives within the All Therapists sample. There was an overall reduction in the proportion of CBT therapists compared to counsellors across the three Phases, with Phase 1 being most similar to the Core Therapists sample in this respect. A total of 31% of therapists were male in the Core Therapists group, a higher

proportion than across the All Therapists samples. The percentage of therapists that were in training was consistent across Phases, and with the Core Therapists sample in Phase 1.

Table 6.3

Features of Therapists in All Therapists Sample

Therapists	N	CBT (%)	Counsellor (%)	Male (%)	Female (%)	Trainee N (%)
Phase 1	81	50 (62)	31 (38)	25	75	13 (16)
Phase 2	80	55 (69)	25 (31)	24	76	12 (15)
Phase 3	74	57 (77)	17 (23)	22	78	13 (17)

Measurement

The primary focus of the study was patient depression outcomes and therapist variability following the Phase 3 intervention in comparison to Phases 1 and 2. The measures available are those collected routinely by IAPT services nationally (see Chapter 3 for full details).

Primary Outcome: PHQ-9

Across all Phases and study samples, the primary outcome measure used was pre-post change on Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001; Spitzer et al., 1999). This is a nine-item measure of depressive symptoms, capturing the primary features of depression included in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 2013), although it is not a diagnostic measure. It comprises questions establishing: feelings of interest/enjoyment; depressed mood or hopelessness; sleep; fatigue; appetite; feelings of failure; concentration; retardation of movement/agitation; suicidal ideation. Each item is scored from 0-3 based on the frequency of symptoms experienced over the preceding two weeks (*"In the last two weeks how often have you been bothered by…"*), to produce a total score ranging from 0 to 27, with higher scores indicating greater severity.

PHQ-9 is an established measure for depression with good internal consistency (Cronbach α = .89), test-retest reliability (intraclass correlation = .84), and sensitivity and specificity (see Kroenke et al., 2001 for details). The outcome measure used within the analysis was the amount of change on the PHQ-9 that a patient demonstrated from first to last Step 3 therapy session, derived by subtracting the final PHQ-9 score from the initial PHQ-9 score. The reliable improvement definition for PHQ-9 score change (Evans et al., 1998; Jacobson & Truax, 1991, see Chapter 4) was utilised as a primary measure of clinical improvement, in line with service definitions.

Secondary Outcomes

Three other outcome measures were routinely collected by the service: GAD-7, WSAS, and a Phobia Scale. Although, the aim was not to produce models predicting these three outcomes, patient and therapist outcomes for these measures were compared across the three Phases. In addition, therapy completion, that is, those patients who completed therapy compared to those who had an unplanned ending or were referred to another service during therapy, and number of attended sessions was also compared.

Generalized Anxiety Disorder-7. (GAD-7; Spitzer et al., 2006). This is a brief measure of symptoms of Generalised Anxiety Disorder utilised in primary healthcare settings. It comprises seven questions each with 0 to 3 rating response options related to the frequency of specific symptoms over a two-week period (i.e. *"Over the last two weeks, how often have you been bothered by…"*). Options are 0 = "*not at all*", 1 = *"several days"*, 2 = *"more than half the days"*, 3 = *"nearly every day"* and total scores have a range 0 to 21. Questions focus on the following symptoms: feelings of nervousness/anxiety; controllability of worry; range of worry thoughts; ability to relax; restlessness; irritability; fear of future consequences. The GAD-7 has a strong internal consistency (Cronbach α = .92), and test-retest reliability (intraclass correlation of .83). See Spitzer et al. (2006) for details.

The Work and Social Adjustment Scale. (WSAS; Marks, 1986; Mundt et al., 2002). This is a brief measure of functional impairment in relation to a specified disorder. There are five questions within the measure, each scored on a 0-8 scale from 0 (meaning *"no*

impairment at all") to 8 (meaning *"very severe impairment"*). The questions focus on specific

areas of life functioning: work/employment; home management; social leisure activities;

private leisure activities; and the formation and maintenance of close relationships. The

WSAS demonstrates good internal consistency (Cronbach α range from .70 to .94), test-

retest reliability (correlation of .73), and correlates with the severity of depression (.76) and

obsessive compulsive disorder symptoms (.61). For further details see Mundt et al. (2002).

Phobia Scale. The Phobia Scale, (DH, 2011a) was developed specifically for use in

IAPT services to provide a measure of specific anxiety, alongside the GAD-7. This has been

somewhat superseded by the use of a wide range of anxiety disorder specific measures

within IAPT services, but remains a measure that is used routinely for all patients who

access IAPT services. The Phobia Scale is a measure of avoidance associated with anxiety

symptoms triggered by specific situations. There are three questions, each scored on a 0 – 8

scale, from 0 (meaning "*I would not avoid it"*), to 8 ("meaning *"I would always avoid it"*). The

areas of focus are: social situations due to a fear of embarrassment or humiliation; situations

due to a fear of having a panic attack or other distressing symptom; situations due to a fear

of a specific object or activity (e.g., spiders, heights etc.).

Control Variables

In order to assess therapist variability, patient variables available at intake and

significantly associated with outcome were controlled for in the analysis. Potential patient

control variables available were: intake scores on PHQ-9, GAD-7, WSAS, and Phobia score

and patient demographics comprising age, gender, and deprivation.

For deprivation, the English Index of Multiple Deprivation (IMD) (Department for

Communities and Local Government, UK Government, 2015) was used. This is a commonly

used measure of deprivation level in a geographical area based on postcode across a range

of seven specific domains: income, employment, education/training/skills, health/disability,

crime, barriers to housing and services, and living environment. Each area in England is

given a rank from 1 (most deprived) to 32844 (least deprived). These are separated into

deciles but are usually analysed as quintiles with quintile 1 (deciles 1 and 2) being the most

deprived and quintile 5 (deciles 9 and 10) being the least deprived. The IMD quintile of the home postcode of each patient included in the study was generated using the English indices of deprivation 2015 look up guide (Ministry of Housing, Communities & Local Government) and this data was included as part of the anonymised dataset provided to the research group.

Therapist Variables

The therapist variables available were limited to their gender, whether they were in training or qualified during the period, and the core therapeutic training of the therapist/therapy type delivered. The professional groups and therapy types delivered in the service were CBT and counselling. Years since qualifying, to provide a measure of therapist experience, was collected but excluded as a variable during the preparation of the dataset due to levels of missing data (52.3% missing data).

Data Analysis

The primary aim was to test whether the variability between therapists, the therapist effect, could be reduced in a service without any negative effect on overall patient outcomes. The primary analysis therefore, used multilevel modelling (MLM) to estimate the size of therapist effect in each study Phase. Overall patient outcomes were defined as pre-post change on PHQ-9 and these were compared across Phases using ANCOVAs. Markov Chain Monte Carlo (MCMC) procedures were used to provide a measure of uncertainty around estimates of the therapist effects when they were compared across Phases (Browne, 2015; Browne & Rasbash, 2009; Snijders & Bosker, 2011). The analytical methods are described in more detail in the following sections.

Primary Analysis: Multilevel Modelling (MLM)

Multi-level modelling (MLM) was used to analyse each of the datasets. MLM is the appropriate analytical method where there is a hierarchical structure in the data, whereby units on lower levels are nested (clustered) within units on a higher level (Raudenbush & Bryk, 2002; Snijders & Bosker, 2011). This is often the situation in large organisational contexts, such as healthcare or education. For example, in education, students are nested

within teachers, teachers are nested within schools, and schools are nested within education authorities or commissioning groups. Each level in this system can impact on the educational attainment (e.g., exam grades) of any individual student. In the current study of a single IAPT service, patients are nested within therapist, a 2-level structure.

MLM recognises that the outcomes of patients treated by the same therapist will be similar in some way, and different from the outcomes of patients treated by a different therapist. Therefore, the outcomes of patients are not independent but are determined to some extent by the therapist providing treatment. If a hierarchical, nested structure exists in the data but is not accounted for in the analysis, this would violate the assumption of the independence of observations, which can lead to an underestimation of the standard errors of regression coefficients and result in a potentially inflated statistical significance (Heck & Thomas, 2020, Raudenbush & Bryk, 2002). By recognising the hierarchical, nested structure of the data MLM is able to control for this lack of independence and estimate the effect on patient outcome of higher-level units, namely therapists (Snijders & Bosker, 2011). In addition, MLM allows more complex models to be developed which may more accurately model the situations and interactions often found in the context of a psychological therapy service. For example, explanatory variables for all levels can be included (i.e., both patient and therapist variables) and also cross-level interactions.

The rationale, methodology, and statistical underpinnings of MLM approach are fully explained elsewhere (e.g., Hox et al., 2017; Raudenbush & Bryk 2002; Snijders & Bosker, 2011), but MLM can be summarised as an extension of multiple regression analysis, where instead of fitting a single regression line through all datapoints producing a fixed intercept, the regression lines and intercepts are allowed to vary between higher level units (e.g., therapists). Multilevel models are often referred to as random intercept models because each higher-level unit has their own regression line through their data and their own intercept. This is illustrated in Figure 6.3.

Figure 6.3

Level 2 Variance – Random Intercept Model
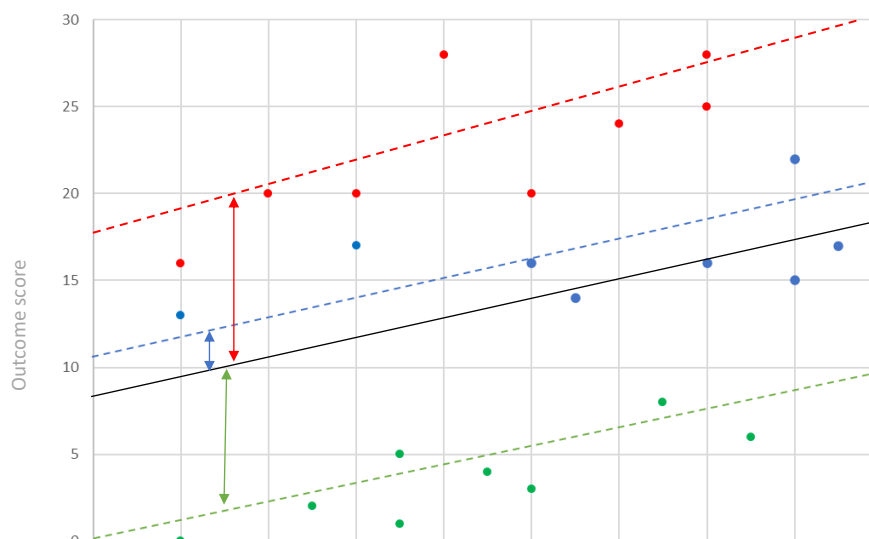
Figure 6.3 shows the lines of best fit for the outcomes of the patients of three different therapists (Red, Blue and Green), plotted on the same graph, with the overall fixed intercept model line included (black solid line). By grouping patients within therapists, each therapist has their own intercept and these vary to some degree. If the lines of best fit for the three therapists in Figure 6.3 mapped exactly on to each other and had the same intercept, then this would indicate no variability between therapists and no therapist effect.

Below is an equation for a simple multilevel model. The first line is similar to a single level regression equation, except there is a subscript 'j' representing the therapist in addition to the subscript 'i' representing the patient. The first line therefore reads that the outcome for patient 'i' seen by therapist 'j' is the average outcome for therapist 'j' (β_{0j}) (the intercept for therapist 'j'), plus the patient residual (e_{ij}) representing other unmodelled contributory factors associated with patient outcome. The second line indicates that therapist 'j's outcome equals the average therapist outcome plus a therapist residual (u_{0j}), the unmodelled therapists' contributory factors.

A multi-level model equation is shown below:

Equation 2

Multilevel model equation

$$y_{ij} = \beta_{0j} + e_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2)$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

The third line informs us that this therapist residual is normally distributed and has a mean of zero and a variance (σ_{u0}^2). The bottom line of the equation indicates that the patient residual is also normally distributed, with a mean of zero and a variance (σ_e^2). Therefore, unlike a single level regression model, a multilevel model splits the variance in patient outcome between the patient level (σ_e^2) and the therapist level (σ_{u0}^2).

Calculating the Therapist Effect. Taken together, the patient and the therapist variance give the total variance in patient outcomes not explained by the model. Therefore, the proportion of this total variance that is at the therapist level is an estimate of the degree to which differences in patient outcomes are due to the variability between therapists. In MLM, this proportion is termed a variance partition coefficient (Rasbash et al., 2017), which is equivalent to the intra-class correlation co-efficient (ICC). The variance partition coefficient is usually multiplied by 100 and expressed as a percentage and termed the 'therapist effect' (Wampold & Owen, 2021). The equation is presented below:

Equation 3

Therapist effect equation

$$\textit{Therapist Effect} = (\sigma_{u0}^2 \,/\, (\,\sigma_{u0}^2 + \sigma_e^2\,)) * 100$$

In the current study, any reduction in therapist variability was primarily based on an observed reduction in the therapist effect across the Phases. Where the variance at the therapist level is non-significant (i.e., when its value is not greater than 1.96 times the value of its standard error), the sample can be assumed to have no significant therapist effect. Non-significant

therapist effects indicate no single therapist's outcomes are significantly better or worse than the average therapist of the group and, in effect, the therapist regression lines and intercepts in Figure 6.3 are very close together and the differences between them are not having a significant effect on overall patient outcomes. From a service and patient perspective, no significant therapist effect would be the ideal position. A method to test for the reliability of therapist effect estimates is discussed below.

Random Slopes. Figure 6.3 shows that the only variation between therapists is in their intercepts. However, the possibility of therapists having different slopes on the predictive variables can also be modelled using random slopes. Considering Figure 6.3, the therapists regression lines would not be parallel but would have different slopes. A significant random slope would indicate that the effect a patient variable had on outcome varied between therapists.

Therapist Residuals. As described above (*Equation 2)*, the multilevel model determines the residuals for both the therapist (u_{0j}) and patient (e_{ij}) and these are assumed to have a mean of zero and some variability (variance). The residual (sometimes called the 'error' term) is the remaining effect that individual therapists and patients have on the outcome not due to the variable/s in the model or variability at another level of the model. As well as providing a test of model assumptions, therapist residuals represent the extent to which an individual therapist's outcomes differ from the average therapist by a positive or negative amount. By ranking and plotting therapist residuals with 95% confidence intervals, in a 'caterpillar plot', therapists can be identified as being above or below average and significantly lower or higher in effectiveness (Goldstein & Healy, 1995; Saxon & Barkham, 2012).

Figure 6.4 shows an example of a caterpillar plot with therapist residuals shown with their 95% confidence intervals,) and therapists ranked from less to more effective. As with the Phases in the current research, the outcome variable used in the above model was amount of change on the PHQ-9, so positive residuals denote more positive change (i.e. better patient outcomes), and negative residuals less change (i.e. worse patient outcomes).

The dotted line on the graph, for the residual of zero, represents the average therapist change - in this case 6.8 points on PHQ-9. To identify therapists who are significantly different to other therapists, those therapists whose residual CIs do not cross this zero residual line could be considered as the lower or higher effectiveness groups, denoted on the graph above with bold markers. Those whose CIs cross the average could be considered to be in the average effectiveness group. In Figure 6.4 six therapists on the right can be identified as above average after controlling for variables in the model, while ten therapists on the left can be considered below average, with the remaining 46 therapists considered average.

A visual indication of the extent of therapist variability is given by the difference between the therapist on the extreme left and the therapist on the extreme right. If the difference between them is less and the plotted points are more level, there would be less therapist variability; a smaller therapist effect. For a non-significant therapist effect the 95% CIs of all of the therapists would cross zero indicating that no therapists are having a significantly greater or lesser effect on patient outcome than the average therapist. Within this study, a reduction in the variability of therapists' outcomes would be observable by the reduction in the numbers of therapists whose CI's did not cross the average line.

Figure 6.4

Example of a Caterpillar Plot Showing Therapist Residuals with 95% Confidence Intervals (CIs)
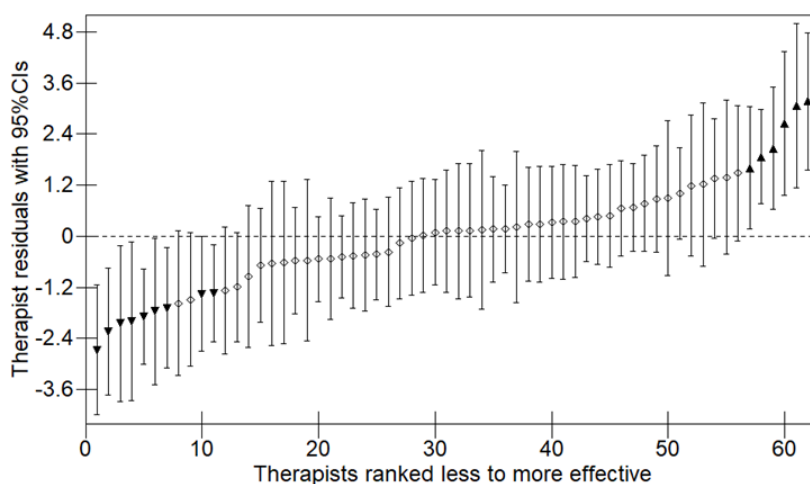
To further develop this graphical use and application of caterpillar plots, the residuals in the models in the study Phases (values on the y-axis) were transformed into actual change in patient outcome scores thereby showing more practically the impact that individual therapists have on their patients' outcomes relative to other therapists in the sample. This was achieved by transforming the residual regression line from a value of 0 (residual) to the value of the average therapists PHQ-9 change and adding or subtracting each therapists residual to provide the change attributable to each therapist.

Multilevel modelling and the comparison of higher-level residuals provide a better and fairer means of comparing the effectiveness of higher-level units (i.e., therapists) than a comparison of higher unit outcomes in themselves, as patient variables (case-mix) are controlled for and the 95% CIs provide a measure of uncertainty (Goldstein & Healy, 1995; Goldstein & Spiegelhalter, 1996). The approach was developed in the UK in the context of education and the evaluation of school outcomes as a response to the critique of school league tables and allowed for the modelling of school effects whilst controlling student variables (Goldstein & Healy, 1995; Goldstein & Spiegelhalter, 1996). In part, due to the need for large samples of data containing individual therapist's variables where a nested structure can be recognised, the methods have only been applied in the field of psychological therapy and psychological therapy services over the last two decades (e.g. Lutz et al., 2007; Saxon & Barkham, 2012). For a critique, see Wampold and Owen (2021).

Testing Model 'Goodness of Fit'

During model development, how well the model fits the data can be assessed by comparing the -2*loglikelihood ratios produced by each model. Reductions in the -2*loglikelihood values indicate improvements in model fit. A single level model containing significant predictors of outcome is created and the reduction in the -2*loglikelihood ratio when the therapist variability (the random intercept) is added is noted and tested for significance against the chi squared statistic for the additional degrees of freedom, which in a random intercept model is 1 representing the additional parameter, the therapist level variance. Statistical significant would indicate that recognising and including the nested

multilevel structure in the model creates a better model fit for the data than a single level model.

Markov Chain Monte Carlo

As noted above, Markov Chain Monte Carlo (MCMC) estimation procedures can be used to provide a measure of uncertainty around the size of therapist effect. It is not possible to put 95% CIs, usually derived from means and standard deviations, around the therapist effects which are based on two variance values. However, MCMC methods are able to provide some measure of uncertainty about effects using 95% probability intervals (PrI). MCMC is a Bayesian approach that uses the model estimates produced by the default modelling procedure, Iterative Generalised Least Squares (IGLS), as 'priors' in a simulation 'chain'. This is an iterative process that continues until model values stabilise and a best fitting model is produced. Each variable coefficient and model parameter had therefore a chain of values which can be summarised, using the 50th percentile value (median) to derive final estimates (Browne, 2015; Browne & Rasbash, 2009). The 2.5% value and the 97.5% value in the chains provide a 95% PrI. Chains will be produced for the therapist and patient level variances and these can be combined to produce a chain of therapist effects, which will have a median value and a 95% PrI. As these 95% PrI are not based on means and standard deviations they cannot be used to test the statistical significance of differences between therapist effects across Phases. However, due to concerns about sample size, they provide an indication of the *reliability* of the estimates of therapist effects in each Phase which allows an assessment of whether differences between phases are reliably different or not. That is, if the PrIs of two Phases did not overlap this would be indicative of reliably different therapist effects. It should be noted that unlike 95% CIs, the average value in 95% PrIs does not need to be, and is unlikely to be, at the midpoint.

In this proof of concept study with a sub-optimal sample size, change in the therapist effects between Phases using standard IGLS procedures was the primary outcome. Bayesian methods were used as a secondary analysis as the only means to assess the reliability of the effects found.

Testing Model Assumptions

Multilevel models assume that the residuals at each level are normally distributed and are homoscedastic (i.e., the variance across different values of a predictor variable is similar). Both assumptions were tested for the final models. This was done for both levels by visual inspection of quantile-quantile (q-q) plots to test for normality and plotting residuals across variable values to test for homoscedasticity.

Developing the Study Models

A multilevel model, with patients at level 1 and therapists at level 2 and including significant predictors of pre-post change on PHQ-9, was developed for each of the six data samples. MLwiN version 3.02 (Rashbash et al., 2009) was used for IGLS and MCMC modelling. Although subsequent versions of MLwiN were available over the period of the study, the same version was used for all analyses to ensure consistency across the three Phases.

Initially, for each sample a single level regression model was created which included significant patient variables determined by z-scores (variable coefficient is greater than 1.96 times the standard error). The therapist level was then added and the therapist level variance assessed for significance (z-score) and improvement in model fit assessed using the reduction in the -2*loglikelihood ratio. Each predictor variable was then checked for a significant random slope, again using reductions in the -2*loglikelihood ratio to test for significant model improvement. Interactions between variables were checked and significant interactions retained.

Models were produced for each Phase within each of the two therapist samples and it was important that the patient variables present in each model were the same in order to make comparisons of therapist effects between Phases. This was achieved by comparing all six models, identifying common, consistent variables and assessing the significance and effect on model fit of other variables and excluding those that had little effect and may have been particular to only one Phase. (Full models containing all significant variables are presented in Appendix H).

Following the identification of common patient predictors, these variables were included in each model and the therapist level introduced and tested for significance and improvement in model fit as above. Therapist effects were then calculated for each model. Finally, each model was run using MCMC and 95% PrIs were calculated for each therapist effect in order to assess the reliability of each effect. For the six final models, model assumptions of normality and homoscedasticity were tested and assessed.

Additional Analysis

In order to test the impact of including trainee therapists in the Core Therapists sample all therapists who were trainees in Phase 1 were removed from this and all other Phases of this sample. This produced a sample of 29 Core Therapists, with 818 patients in Phase 1, 1047 in Phase 2 and 1005 patients in Phase 3. MLMs were produced for each Phase and compared with the original Core Therapists models to identify any impact of including trainee therapists in the original sample. In addition, a sensitivity analysis on a larger Core Therapists sample was conducted by including all therapists who had treated a minimum of one patient in each of the study Phases. By so doing, this increased the therapist and patient sample size for the primary analysis (Core Therapists), providing a further check of the reliability of the results. By expanding the inclusion criteria in this way, this produced a Core Therapists sample of 53, with 1104 patients in Phase 1, 1472 patients in Phase 2 and 1412 patients in Phase 3.

Comparative and Descriptive Analysis

IBM SPSS Statistics Version 25.0 (2017) was used for all other analyses in the study. These analyses included a comparison of the main features of the dataset of each Phase of the study for both the Core Therapists and All Therapists datasets (e.g., average initial/final scores on each measure, reliable improvement rates, score change on each clinical measure, average number of sessions attended, percentage of cases completed vs dropped out/referred on).

In addition, reliable improvement rates and mean PHQ-9 change scores were calculated for each of the stepped wedge clusters (see Chapter 5) at each implementation

time period. Samples for comparison were identified by therapist cluster of the stepped wedge, and patients whose therapy sessions were contained within a specific 'step' time period. Though the stepped wedge design was utilised for practical implementation reasons, this also provided an opportunity to acquire indicative descriptive data during this implementation. As this was just one component of the Phase 3 package of improvements, this was not designed to be part of the main analysis.

Summary

This chapter has provided details of the structure of the dataset across the three Phases of the study. Details of the data analysis that were undertaken on all datasets was described with reference to relevant statistical methodology. Primary analysis focused on determining the presence or absence of a therapist effect, determined by z-scores, and through an assessment of improvement in model fit and illustrated by caterpillar plots. Changes in the size of therapist effect could be observed between Phases, and MCMC PrIs were used to provide a measure of uncertainty around each therapist effects and were plotted to inform any assessment of differences between the effects found. Additional analyses to test the impact of trainees within the sample, and a sensitivity analysis using differing number of patients and therapists, were also conducted. The following two chapters provide the results of each of the study Phases. Chapter 7 presents the results of the primary analysis of the Core Therapists sample. Chapter 8 presents the analysis of the All Therapists sample which was used to test the generalisability of the findings of the primary analysis across a whole service context.

Chapter 7

Results 1 Core Therapists

Overview

The results presented in this chapter relate to the patient outcomes of the 35

therapists who were present in all three Phases of the study (referred to as 'Core Therapists'

sample). Due to the nature of the study taking place over a 4.5-year time period, it was not

possible to ascertain which therapists included in the Phase 1 baseline would appear in all

three Phases, therefore the extraction and analysis of this dataset was only possible at the

end of all three Phases. By controlling for therapists in this way, it was therefore possible to

investigate the therapist effect over time, under three different service conditions. These

Core Therapists were present in the service for the full duration of the study (i.e. 4.5 years)

and had treated the minimum of 10 patients in each of the three Phases over the 4.5-year

period. As a consistent therapist cohort across the study, any differences found in the

therapist effect for this group at each study Phase could be more reliably assumed to be, at

least in part, due to the service conditions of the specific Phase, rather than due to the effect

of new therapists entering the service or therapists leaving the service. The primary

comparison was between Phase 2 and Phase 3, as these Phases represented two

sequential service interventions / improvements. Therefore, any improvements in Phase 3

outcomes may be more reliably assumed to be attributed to the addition of the Collaborative

Intervention package, rather than the generic act of implementing a service improvement.

Comparisons between Phase 2 and Phase 1, and Phase 3 and Phase 1 were also

undertaken and represent secondary outcomes. The second results chapter (Chapter 8) will

focus on the generalisability of any findings in the Core Therapists dataset, by presenting the

results of analysis of all therapists included in the study (referred to as the 'All Therapists'

sample), including the Core Therapists.

Results

The following results begin with a descriptive analysis of the patient level data and

other relevant variables (the Core Therapists Patient and Therapist Features are presented

in Chapter 6). Pre and post clinical scores and reliable improvement rates are also presented and compared across Phases to assess any change in overall outcomes. The identification of the significant variables for each Phase are then presented with each Phase Multilevel Model (MLM). Therapist residuals produced by MLM models are presented graphically showing the variability across the therapist sample. MCMC estimations of 95% PrIs around the therapist effects are presented to inform an assessment of the reliability of differences between the change in therapist effect across the Phases. Additional analysis of a broader data sample of Core Therapists is presented as a sensitivity analysis, as well as a comparison of therapist effects with trainees excluded from the Core Therapists sample. Finally, descriptive analysis of the stepped wedge clusters is provided.

Core Therapists: Patient Level Descriptive Analysis

Table 7.1 shows the main patient level variables for each Phase of the study.

Table 7.1

Core Therapists Sample - Patient Level Descriptive Analysis

Study Phase	Pt N	Mean Patient Age in years (SD)	Patient Gender (%)		Referral Source (%)				Patient IMD Quintile Group (%) (1 = highest deprivation, 5= lowest deprivation group)				
			Female	Male	Self	GP	Mental Health Service	Other	1	2	3	4	5
1	930	45.8 (15.25)	65.9	34.1	59.2	36.0	3.3	1.5	10.4	23.3	34.6	23.7	8.0
2	1226	41.5 (15.93)	69.2	30.8	69.5	26.8	3.2	0.5	10.9	21.3	35.7	20.2	11.9
3	1217	40.7 (15.69)	66.8	33.2	77.4	19.3	2.2	1.1	13.7	20.7	30.5	21.1	14.0

n.b. Referral Source 'other' includes: prison/probation services, employers, voluntary organisations

The average age of patients in the datasets reduced significantly over the study period, (t(3370) = 30.92, p <.001). The difference was significant between Phases 1 and 2 (t(2154) = 6.26, p < .001) and between 1 and 3 (t(2145) = 7.55, p< .001), but not between

Phases 2 and 3 (t(2441) = 1.32, p= .19). The majority of patients were female and this was the case in each Phase, between 65-69%, with no significant difference between Phases, X^2 (2) = 2.88, p = .237.

The proportion of patients referred from particular sources was significantly different between the Phases, X^2 (4) = 80.13, p <.001, and comparisons between pairs of Phases were also significant (all p-values <.001). The proportion of those who self-referred increased across all Phases, from 59.2% in Phase 1 to 77.4% in Phase 3, with a reduction in referrals originating from GPs from 36% to 19.3% of patients. This is reflective of the Phase 2 service changes, where GPs had been encouraged to ask their patients to self-refer rather than directly refer themselves. Deprivation scores (based on patient home address) differed significantly across the Phases X^2 (8) = 32.52, p <.001, specifically between Phase 1 and 3, X^2 (4) = 27.4, p <.001 with an increase in the proportion of patients being in the 'least deprived' and 'most deprived' groups, and a decrease of those in the other three groups. Table 7.2 reports the initial and final scores and pre-post change scores on each of the clinical outcome measures, across the three Phases.

Table 7.2

Core Therapists Sample – Pre/Post Clinical Scores

Study Phase	PHQ-9 (SD)			GAD-7 (SD)			WSAS (SD)			Phobia (SD)		
	Pre	Post	Change	Pre	Post	Change	Pre	Post	Change	Pre	Post	Change
1	14.7	8.6	6.1	13.1	7.4	5.6	20.1	14.3	6.2	8.9	6.2	2.7
	(6.06)	(6.58)	(6.32)	(5.23)	(5.59)	(5.74)	(9.19)	(10.37)	(9.60)	(6.57)	(6.35)	(5.81)
2	14.7	8.2	6.6	13.4	7.5	5.9	17.6	12.1	5.7	8.3	5.5	2.9
	(5.77)	(6.21)	(6.05)	(4.95)	(5.27)	(5.45)	(9.22)	(9.37)	(8.57)	(6.40)	(5.82)	(5.26)
3	15.2	8.6	6.5	13.4	7.7	5.6	19.3	12.9	6.1	9.1	6.1	3.0
	(5.61)	(6.00)	(5.86)	(4.76)	(5.22)	(5.39)	(8.69)	(9.60)	(8.23)	(6.31)	(5.97)	(5.20)

Differences between initial clinical measures were small, though there was a significant difference between mean initial WSAS scores between Phases, F(2,3370) =

21.78, *p* <.001. The difference was significant between Phase 1 and 2 (*t*(2154) = 6.26, *p* <.001), Phase 2 and 3, (*t*(2441) = - 4.46, *p* <.001), and Phase 1 and 3, (*t*(2145) = 2.29, *p* <.05). Initial Phobia scores were not normally distributed, and there was a significant increase in mean scores, *K-W*(2) = 10.04, *p* = .007. In pairwise comparisons there was a significant difference between Phase 2 and Phase 3, (*p* = .005, *r* = - .06), but not between Phase 1 and Phase 2 (*p* = .199, *r* = .04) or Phase 1 and Phase 3 (*p* = 0.837, *r* = - .02).

Average PHQ-9 change increased from 6.1 in Phase 1 to 6.6 in Phase 2 and 6.5 in Phase 3. However, ANCOVA tests on each mean change score found no significant differences in mean change scores on any clinical measure between the three Phases. Table 7.3 compares the three Phases on other important outcomes for evaluations of service and therapist performance, the percentage of patients whose scores reliably improved on the PHQ-9, the number of sessions attended and whether therapy was completed.

Table 7.3

Core Therapists Sample – Clinical Indicators: Reliable Improvement, Therapy Completion and Number of Sessions

Study Phase	PHQ-9 Reliable Improvement Rate N (%)	Therapy Completion Rate N (%)	'Number of Sessions' Mean (SD)		
			Attended	Missed	Offered
1	463 (49.8)	601 (64.6)	7.2 (3.73)	1.8 (1.99)	9.0 (4.54)
2	671 (54.7)	960 (78.3)	7.2 (4.03)	1.9 (1.97)	9.2 (4.90)
3	681 (56.0)	931 (76.5)	8.2 (4.26)	2.2 (2.13)	10.4 (5.23)

There was a significant increase in the PHQ-9 Reliable Improvement rate, of 6.2% points over the three Phases (X^2 (2) = 8.74, *p* =.013), increasing from just under 50% in Phase 1 to 56% in Phase 3 (X^2 (1) = 8.06, *p* =.005). Therapy completion rate (number of patients who completed therapy rather than those who were stepped up, or discontinued therapy early) increased significantly (X^2 (2) = 58.05, *p* <.001). There was a significant increase in completion rate between Phase 1 and Phase 2 (X^2 (1) = 49.53 *p* <.001), from

64.6% of patients, to 78.3% of patients completing, and between Phase 1 and Phase 3 (X^2 (1) = 36.37, p <.001), with 76.5% of patients completing therapy in Phase 3. There was however, a small but non-significant decrease in completion rate between Phase 2 and Phase 3. The average number of attended sessions also increased significantly (K-W(2) = 43.85, p < .001). Differences were significant between Phase 2 and 3 (p = .000, r = - .12) and Phase 1 and 3 (p <.000, r = - .11), but not between Phase 1 and Phase 2 (p = 1.0, r = .01). Missed sessions increased significantly, (K-W(2) = 25.57, p <.001), between Phase 2 and 3 (p = .001, r = - .07) and Phase 1 and 3 (p = .000, r = - .10), though in practice this amounts to a maximum increase in missed sessions of less than half a session. Taking attended and missed sessions together it suggests that in Phase 3 more sessions were offered overall.

Identification of Common Predictor Variables

In order to compare therapist effects across the three Phases, the common patient (case-mix) variables significantly predictive of outcome in each Phase were identified in a preliminary analysis. A multilevel model for pre-post change in PHQ-9 was developed for each of the three Phase samples (See Chapter 6 for detailed method). The full models for each Phase are in Appendix H, tests of model assumptions are in Appendix I).

Comparing the three models, the same variables were significant or non-significant across all models, with the exception of IMD level, which was a significant predictor variable in Phase 1, with an effect of -0.96 PHQ change points if the patient was in the lower deprivation quintile groups (quintiles 4 & 5) but not significant in the other Phases. This may be due to changes that occurred, in terms of the percentages of the different IMD quintiles referred to the service between the Phases (see Table 7.1). However, including the IMD variable in the Phase 1 model made little difference in the 'goodness of fit', with minimal reduction in the -2*loglikelihood ratio from 5811.147 to 5809.774. Therefore, for consistency it was excluded. Initial Phobia score was non-significant in Phase 2, however was significant in the other Phases, so was included to ensure consistency and the ability to compare models. The interactions between variables were also tested in each model but only the

interaction between initial PHQ-9 and initial WSAS score was found to be significant and it was significant in each Phase. Table 7.4 presents the patient variables significant in all three Phases.

Table 7.4

Final Significant Predictor Variables in Each Model

Predictive Variables	Phase 1			Phase 2			Phase 3		
(grand mean centred)	Value	*S.E.*	*p*	Value	*S.E.*	*p*	Value	*S.E.*	*p*
Initial PHQ-9	0.56	*0.04*	*0.00*	0.59	*0.03*	*0.00*	0.59	*0.03*	*0.00*
Initial Phobia	-0.14	*0.03*	*0.00*	-0.05	*0.03*	*0.08*	-0.08	*0.03*	*0.00*
Initial WSAS	-0.06	*0.03*	*0.02*	-0.10	*0.02*	*0.00*	-0.10	*0.02*	*0.00*
Interaction PHQ-9 * WSAS	-0.01	*0.00*	*0.04*	-0.01	*0.00*	*0.01*	-0.01	*0.00*	*0.00*

The positive coefficient for initial PHQ-9 is a statistical function of outcome being pre-post change on PHQ-9. For each additional point on PHQ-9 at intake, the change between intake score and outcome score increases by a smaller amount, 0.56 of a point. Therefore, the actual outcome score is higher (poorer outcome) by 0.44 of a point for each additional point at intake and outcome. Greater severity at intake for PHQ-9 is therefore associated with a poorer clinical outcome score. Having a higher Phobia score and a higher WSAS score (greater severity of phobia symptoms and functioning) were also associated with less pre-post change on PHQ-9.

The interaction between Initial PHQ-9 and Initial WSAS (shown in Table 7.4) was significant in each model. This indicates that a higher PHQ-9 score and a higher initial WSAS score had a negative effect on outcome greater in addition to the effects of the two variables individually.

Therapist Effect Across Phases

Table 7.5 shows the predictor variables in Table 7.4, in the context of the multilevel models including the level 1 and level 2 variances and the therapist effect in each Phase.

Figure 7.1 shows the model in its original form and is presented as an example of MLwiN output.

Table 7.5

Estimates from Multilevel Models

MLM Values	Phase 1		Phase 2		Phase 3	
	Value	S.E.	Value	S.E.	Value	S.E.
Intercept: Average therapist PHQ-9 Change	6.13	0.30	6.66	0.26	6.86	0.21
First PHQ-9	0.56	0.04	0.59	0.03	0.59	0.03
First Phobia	-0.14	0.03	-0.05	0.03	-0.08	0.03
First WSAS	-0.06	0.03	-0.10	0.02	-0.10	0.02
Interaction PHQ-9 / WSAS	-0.01	0.00	-0.01	0.00	-0.01	0.00
Level 2 (therapist) Variance	1.50	0.66	1.32	0.52	0.47	0.29
Level 1 (patient) Variance	29.38	1.39	26.77	1.10	25.22	1.04
Therapist Effect	4.9%		4.7%		1.8%	

Figure 7.1

MLWin Output Showing Core Therapists Phase 1 Model

$$\text{PHQ Change}_{ij} = \beta_{0j} + 0.564(0.037)(\text{First PHQ9-gm})_{ij} +$$
$$-0.139(0.033)(\text{First Phobia-gm})_{ij} +$$
$$-0.065(0.027)(\text{First WSAS-gm})_{ij} +$$
$$-0.006(0.003)(\text{First PHQ9-gm}).(\text{First WSAS-gm})_{ij} + e_{ij}$$
$$\beta_{0j} = 6.134(0.302) + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \quad \sigma_{u0}^2 = 1.502(0.658)$$
$$e_{ij} \sim N(0, \sigma_e^2) \quad \sigma_e^2 = 29.382(1.387)$$
$$-2*loglikelihood = 5811.147(930 \text{ of } 930 \text{ cases in use})$$
UNITS:
      Therapist Code: 35 (of 35) in use

The total variance reduced across the Phases, from 30.88 in Phase 1, to 25.69 in Phase 3, a 16.8% reduction in overall outcome variance. The proportion of variance attributable the therapists reduced most, with a 68.7% reduction in therapist variance in Phase 3 compared to Phase 1. The patient variance reduced by 14.2% over the same period. Therefore, outcomes became more consistent in Phase 3 due to both patient and

therapist factors but, as a proportion of the variance in Phase 1, therapist factors showed the larger reduction, hence the smaller therapist effect. Also, the reduction in therapist variance indicates that in Phase 3, therapists are more similar to one another in relation to their patient outcomes. In Phase 3 the therapist level variance was not significant as indicated by the large standard error (0.29).

The therapist effect, is also shown in Table 7.5. This indicates that the percentage of the outcome variance that is due to differences between therapists is reduced from 4.9% in Phase 1, to 4.7% in Phase 2, and with a further reduction to 1.8% in Phase 3. In Phases 1 and 2 the therapist effect was significant at the 0.001 level, as indicated by improvements in model fit when therapist variability was introduced. For Phase one the -2*loglikelihood ratio of the model reduced by 18.807 which when compared to the chi squared statistic for the additional degrees of freedom (1 for the additional parameter) was statistically significant (p <.001). For Phase 2 the reduction was also significant (X^2 = 21.906, p <.001). The therapist effect in Phase 3 was not statistically significant at the 0.001 level, but was significant at the 0.05 level, as indicated by the reduction in -2*loglikelihood ratio from 7403.194 in a single level model to 7398.754 when therapist variability was introduced (X^2 = 4.44, p = 0.035). Therefore, although the variance at the therapist level was not statistically significant, modelling the nested structure still improved model fit in Phase 3 but to a lesser degree than in the other Phases.

Table 7.6

ANCOVA Comparisons Between Phases: PHQ-9 Outcomes

Comparison of Phases	df	*F* value	*p* value
Phase 1 – Phase 2	1	0.392	0.531
Phase 2 – Phase 3	1	0.097	0.756
Phase 1 – Phase 3	1	0.096	0.756

While the variability between therapists reduced across the Phases, Table 7.5 also shows that the average therapist PHQ-9 change (the model intercept) for each Phase

increased over the course of the study. However, an ANCOVA comparing patient outcomes on PHQ-9 between Phases found that the improvement in PHQ-9 change was not significant (F(2, 3373) = .156, p = .856). See Table 7.6 for comparisons of all Phases. This would indicate more consistent therapists and a mean patient outcome being maintained across Phases.

Core Therapist Residual Charts

Figure 7.2 charts therapist residuals generated by the model with their 95% confidence intervals (see Chapter 6 for an explanation). The caterpillar plots illustrate the reduction in variability with the points more level in Phase 3 compared with Phase 2 and Phase 1. Calculating the residuals for each therapist within the model allows the identification of groups of therapists whose patient outcomes are reliably and significantly higher or lower than the average therapist. For this study, in Phase 1 and Phase 2 there are two therapists (denoted in green) showing significantly higher effectiveness than the average therapists (denoted in blue) and one therapist (denoted in black) showing significantly lower effectiveness than average. The chart for Phase 3 shows the confidence intervals of all therapists crossing the average residual line, indicating that no therapist is significantly lower or higher in terms of effectiveness than the average therapist, illustrating that there is no significant therapist effect.

To demonstrate how this therapist variability impacts patient outcome scores on the PHQ-9, the residual value of zero has been transformed into PHQ-9 change scores representing the average therapist pre-post change score, i.e. the model intercept value. These charts, shown in Figure 7.3, have greater utility in a service or individual therapist context as they illustrate the difference between therapists in terms of actual PHQ-9 change scores. For example, in Phase 1 the therapists on the extreme left and right have 95% CIs that do not overlap each other so can be considered significantly different. The therapist on the extreme left of the chart has an average change of 4.03 points on PHQ-9, while the therapist on the extreme right of the chart has an average change of 8.17 points, a difference of 4.14 points.

Figure 7.2:

Ranked Therapists Showing Therapist Residuals with 95% Confidence Intervals (CI)

Phase 1: Baseline

0 = 6.134
(PHQ-9 change)

Phase 2: Service intervention

0 = 6.665
(PHQ-9 change)

Phase 3: Research intervention

0 = 6.865
PHQ-9 change

Legend

◆ Average effectiveness   ◆ Higher effectiveness   ◆ Lower effectiveness

Figure 7.3:

Ranked Therapists Showing Average PHQ-9 Change for Each Therapists Caseload with 95% Confidence Intervals (CI)
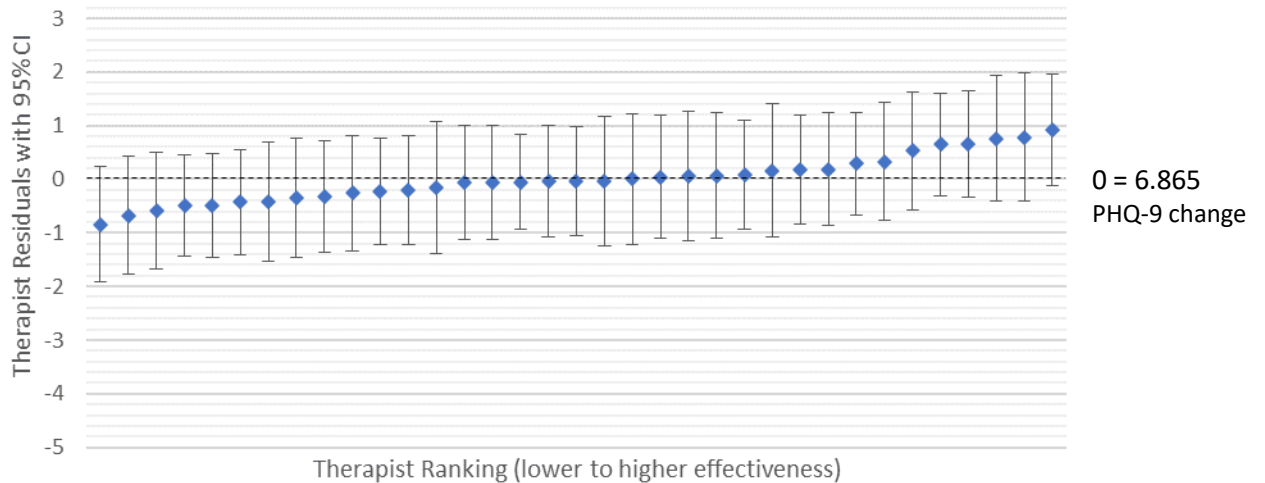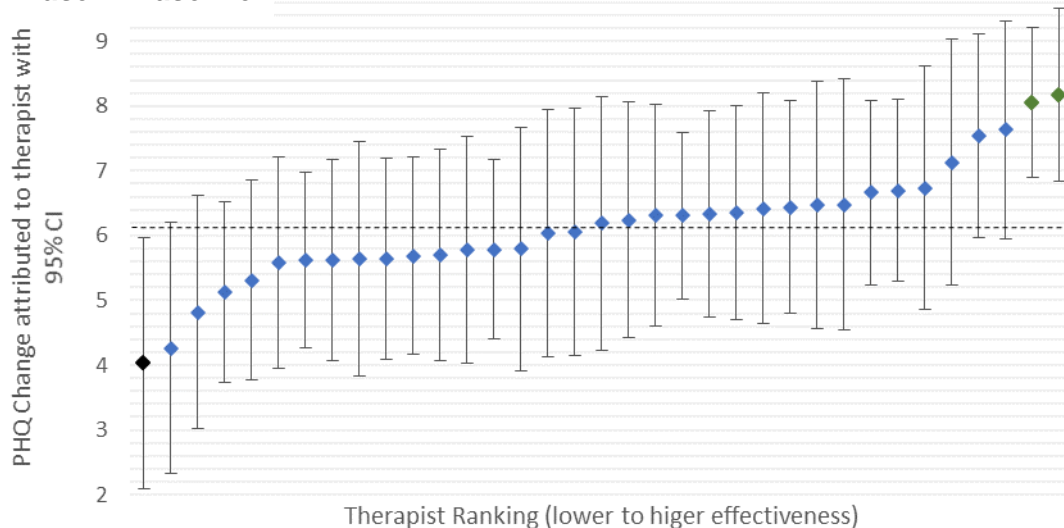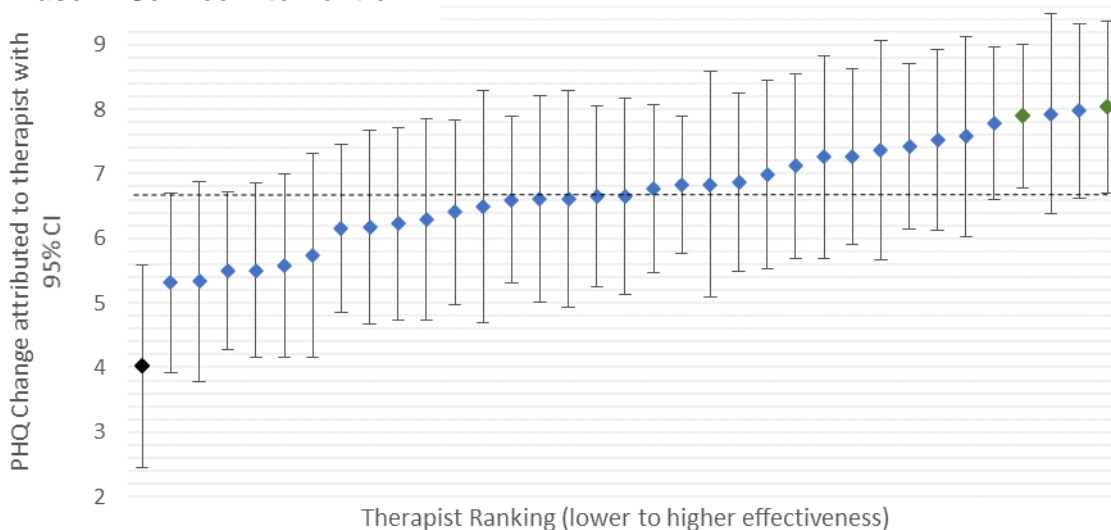
Phase 1: Baseline
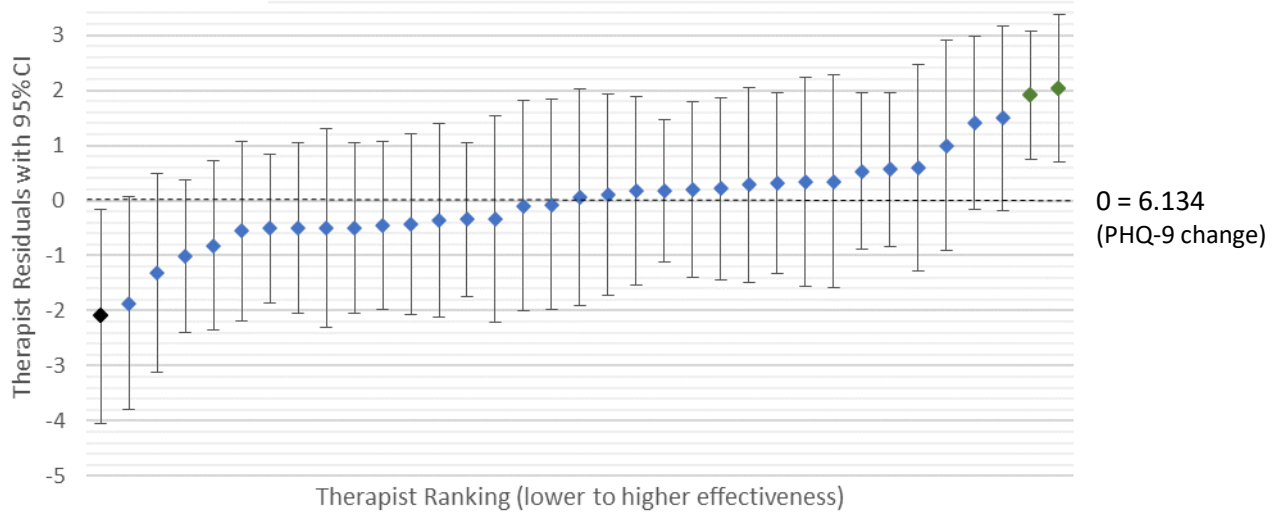
Phase 2: Service intervention

Phase 3: Research intervention
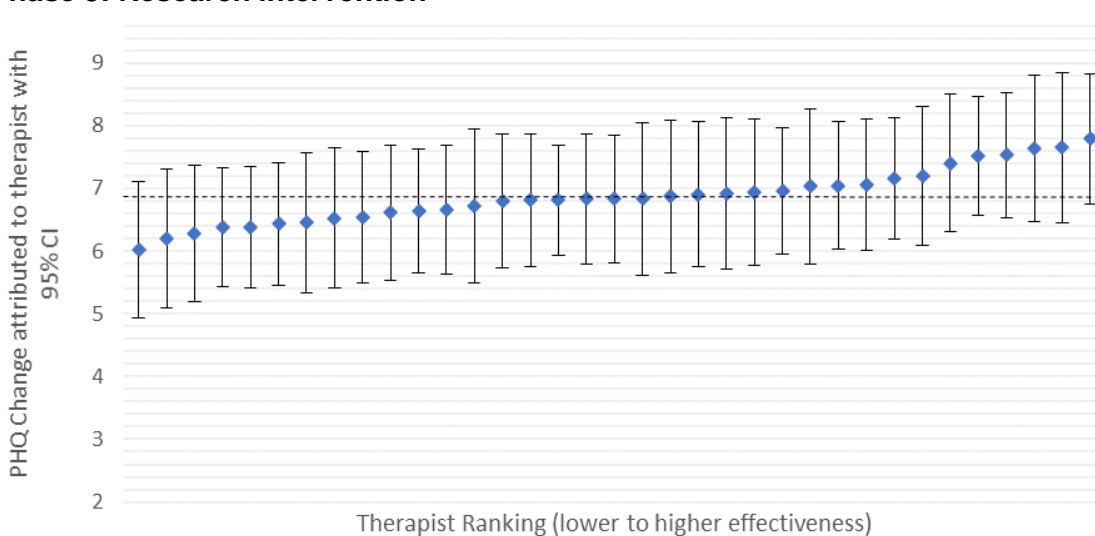
Legend

◆ Average effectiveness   ◆ Higher effectiveness   ◆ Lower effectiveness

Therefore, on average, the most effective therapist obtained more than double the pre-post change for their patients than of the least effective therapist.

In Phase 3 as well as there being no therapists whose CI's do not across the average therapist line, the difference between the two extreme therapists in each chart becomes narrower. Phases 1 and 2 range from 4 to 8 PHQ-9 points of change, with Phase 3 ranging from 6 to 7.8 PHQ-9 points of change. The 'levelling out' of average score change shows more improvement at the lower change end (2 PHQ-9 points), with just a 0.2 of a PHQ-9-point reduction at the upper change end. This indicates that the change has been in uplifting the effectiveness of the therapists at the lower effectiveness end in previous Phases. The average therapist (residual) line across the Phases shows a slight increase from 6.13 to 6.86, but as previously presented, this is a non-significant change.

Interrogation of the movement of the 'top' five and 'bottom' five ranked therapists in each Phase identified that there was movement within the therapist cohort. The top two ranked therapist in Phase 1 remained in the top five ranking in Phase 2, but were replaced by other therapists in Phase 3, with the second top ranked therapist in Phase 1 dropping to the bottom five in Phase 3 (fifth from bottom). There was one therapist who moved from the bottom five in Phase 1, to the top five ranking in Phase 2 and remained in the top five ranking in Phase 3. Interestingly, there was one therapist who was ranked third from the bottom in Phase 1 who moved to the third top ranked therapist in Phase 3. This would suggest individual differences between therapists in how they were affected by the conditions in the Phases, although the overall trend was to 'level' the therapists relative to one another.

MCMC Estimations

As previously mentioned, the estimates of therapist effects may be somewhat unreliable due to the limited number of therapists included. Therefore, in order to assess the reliability of the effects found MCMC estimations were derived for each model. These produced 95% Probability Intervals (PrIs), for the therapist effect in each Phase (see Chapter 6 for detail). The estimates and PrIs were plotted as shown in Figure 7.4 and

demonstrate the reduction in the therapist effect across Phases. However, as would be expected due to being underpowered, the 95% PrIs for each estimation are wide and overlap, which would indicate that the reliability of this reduction is yet to be established.

Figure 7.4

MCMC Therapist Effect (with 95% Probability Intervals) for Each Study Phase
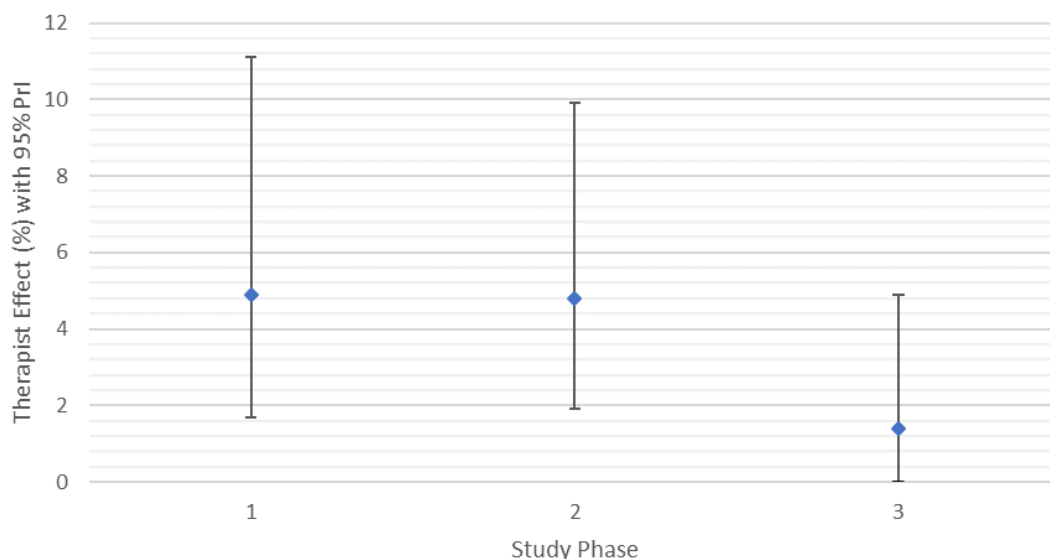
Additional Analysis

Three additional analyses were undertaken on the datasets; firstly, a sensitivity test with the trainee therapists removed from the sample; secondly, a sensitivity test on the reliability of the therapist effects found in the Core Therapists data, and thirdly an assessment of the descriptive data available from the stepped wedge part of Phase 3.

Therapist Effects with Trainee Therapists Removed. This assessment aimed to test the impact that trainee therapists may be having on the findings presented above. MLMs were calculated for the samples for each Phase, removing the therapists who had been in training during Phase 1 from the sample in all Phases. This reduced the therapists in the Core Therapists sample to 29, with 818 patients in Phase 1, 1047 in Phase 2 and 1005 patients in Phase 3. The estimates from the models, including therapist effects, are presented in Table 7.7.

Table 7.7

Estimates from Multilevel Models with Trainee Therapists Removed

MLM Values	Phase 1			Phase 2			Phase 3		
	Value	*S.E.*	*p*	Value	*S.E.*	*p*	Value	*S.E.*	*p*
Intercept: Average therapist PHQ-9 Change	6.30	*0.34*	0.00	6.84	*0.29*	0.00	7.10	*0.23*	0.00
First PHQ-9	0.57	*0.04*	0.00	0.58	*0.04*	0.00	0.61	*0.04*	0.00
First Phobia	-0.12	*0.03*	0.00	-0.04	*0.03*	0.20	-0.06	*0.03*	0.04
First WSAS	-0.09	*0.03*	0.00	-0.10	*0.02*	0.00	-0.12	*0.02*	0.00
Interaction PHQ-9 / WSAS	-0.01	*0.00*	0.03	-0.01	*0.00*	0.03	-0.01	*0.00*	0.00
Level 2 (therapist) Variance	1.84	*0.79*		1.40	*0.59*		0.60	*0.34*	
Level 1 (patient) Variance	28.70	*1.44*		26.90	*1.19*		22.96	*1.04*	
Therapist Effect									
Trainees Removed	6.0%			4.9%			2.5%		
Trainees Included (original)	4.9%			4.7%			1.8%		
PHQ-9 Reliable Improvement Rate									
Trainees Removed	51.6%			54.7%			57.6%		
Trainees Included (original)	49.8%			54.7%			56.0%		

As presented in Table 7.7, the therapist effects show a similar pattern to the models using the Core Therapists samples with trainees included. As shown by the size of the standard error, the therapist variance in Phase 3 is not significant. This mirrors the reduction in variability between therapists during this Phase with that of the original dataset. Initial Phobia scores were retained in the models for consistency, though this was not a significant predictor variable in Phase 2 for the qualified therapists' sample, as with the original Core Therapists sample. Reliable improvement rates significantly improved (X^2 (2) = 7.34, *p* =.025) by a similar amount to the original dataset, albeit starting at a slightly higher rate.

Expanded Core Therapists Sample. This sensitivity assessment utilised a larger dataset of Core Therapists, including all therapists that had at least one patient in each

Phase (rather than 10 in the main analysis). This aimed to provide an assessment of the

reliability of the therapist effect phenomenon observed in the main analysis, as increasing

numbers at the therapist level will increase the reliability of the level 2 variance findings,

although potentially at the cost of maintaining the reliability of the estimates of the variables

in the model. By including therapists who had a minimum of 1 patient in each Phase, this

increased the number of Core Therapists to 53. The patient sample was 1104 in Phase 1,

1472 in Phase 2 and 1412 in Phase 3. A MLM was calculated for the new expanded Core

Therapists dataset for each Phase. Table 7.8 below shows the findings alongside the

therapist effects from the original Core Therapists sample for comparison.

Table 7.8

Values from Sensitivity Test Multilevel Models with Original Sample Therapist Effects

Estimates and Reliable Improvement Rates Shown

MLM Values	Phase 1			Phase 2			Phase 3		
	Value	*S.E.*	*p*	Value	*S.E.*	*p*	Value	*S.E.*	*p*
Average therapist PHQ-9 Change (Intercept)	6.15	*0.26*	*0.00*	6.39	*0.25*	*0.00*	6.44	*0.22*	*0.00*
First PHQ-9	0.55	*0.03*	*0.00*	0.57	*0.03*	*0.00*	0.59	*0.03*	*0.00*
First Phobia	-0.12	*0.03*	*0.00*	-0.04	*0.03*	*0.08*	-0.09	*0.03*	*0.00*
First WSAS	-0.08	*0.02*	*0.00*	-0.09	*0.02*	*0.00*	-0.10	*0.02*	*0.00*
Interaction PHQ-9 / WSAS	-0.01	*0.00*	*0.01*	-0.00	*0.00*	*0.03*	-0.01	*0.00*	*0.00*
Level 2 (therapist) Variance	1.38	*0.59*		1.75	*0.57*		0.88	*0.38*	
Level 1 (patient) Variance	30.71	*1.33*		28.13	*1.05*		26.45	*1.01*	
Therapist Effect									
>=3 Dataset	4.3%			5.9%			3.2%		
>=30 Dataset	4.9%			4.7%			1.8%		
PHQ-9 Reliable Improvement									
>=3 Dataset	49.8%			54.7%			56%		
>=30 Dataset	49.6%			52.6%			53.3%		

As shown in Table 7.8, a reduction in therapist effect from Phase 1 to Phase 3 is

evident, albeit a smaller reduction than in the original analysis, from 4.3% in Phase 1 to 3.2%

in Phase 3. In contrast to the original analysis, an increase in therapist variance is observed

in Phase 2, increasing from 4.3% (Phase 1) to 5.9% (Phase 2) of the total variance residing

at the therapist level. Accounting for the increase in therapist effect in Phase 2 in this sample compared to the original, and the larger therapist effect in Phase 3, this means that the reduction in therapist effect from Phase 2 to Phase 3 is in fact similar, with 2.9% points difference in the original dataset, and 2.7% points difference in the expanded sample. Changes in PHQ-9 reliable change were smaller than in the original sample, with an increase from 49.6% in Phase 1 to 53.3% in Phase 3, which was a non-significant increase (X^2 (2) = 3.61, p =.164). These models also demonstrate a consistent PHQ change across the three Phases.

As with the primary analysis, MCMC estimations and 95% PrI were used to assess the reliability of effects. As with the previous MCMC analysis the PrI's overlapped (see Appendix J) so the differences based on these estimations cannot be considered reliable and despite the increased sample size, the sample remains underpowered and is therefore provided for illustrative purposes. Nevertheless, this sensitivity analysis shows that the therapist effect in Phase 3 in this sample remained significant, which may be an indication that the Phase 3 effect in the original analysis may be an under-estimate.

Stepped Wedge Descriptive Analysis. The stepped wedge design in Phase 3 was largely implemented for practical purposes, as it allowed the implementation of DP-lite over a large therapist cohort populating a large geographical area (see Chapter 5 for further details). However, this design also allowed the Phase 3 datasets to be further split according to therapist cluster, and stepped wedge stage. Splitting the data in this way resulted in smaller numbers of therapists and patients in each data subset, making further MLM analysis unviable, and any other results from statistical tests should accordingly be interpreted tentatively. However, the PHQ-9 reliable improvement rate and mean PHQ change for each relevant therapist cluster to compare the two conditions (control and DP-lite) at each stepped wedge stage is shown in Table 7.10. In the table, 'control' refers to therapists in reflective skills practice groups only, and 'DP-lite' refers to therapists in reflective practice groups with the addition of DP training and support (see Chapter 5 for details). The initial stage of the stepped wedge implementation involved all therapist clusters

being in the control (reflective practice groups) condition. However, due to this period lasting

only 1-month, insufficient cases (patient N = 2) were available for comparison, so this period

is excluded. As shown in Table 7.9, the PHQ-9 reliable improvement rate and average PHQ-

9 change score appear higher in the DP-lite conditions in the two stepped wedge stages

where both control and DP-lite groups were running concurrently. However, no significant

effect of DP condition on outcomes was identified when tested, on either reliable

improvement rates (Stage 1, X^2 (1) = .061, p =.805; Stage 2, X^2 (1) = .332, p =.564) or

mean PHQ-9 change (Stage 1, t(86) = -.235, p = .087; Stage 2, t(66) = -1.328, p = .620) . In

the final stage when all therapist clusters are in the DP-lite condition, reliable change rates

and average PHQ-9 appear to reduce overall, though there is no comparator control group in

this stage.

Table 7.9

Combined Clusters Under Control and DP Conditions Over the Stepped Wedge Stages

Stepped Wedge Stage	DP Condition (N patients, N therapists)	PHQ-9 Reliable Improvement Rate (%)	Mean PHQ Change (SD)
1	Control (61, 24)	63.9	6.92 (4.4)
	DP Lite (27, 7)	66.7	7.19 (6.0)
2	Control (28, 11)	60.7	6.61 (6.5)
	DP Lite (40, 11)	67.5	8.65 (6.0)
3	DP Lite (405,34)	53.1	6.25 (6.0)

Summary

Other than small differences in some features of the patient level clinical data, the

variables associated with outcome were consistent across datasets. The development of

MLMs saw three predictor variables utilised in all models (initial PHQ-9 score, initial phobia

score and initial WSAS score) with a significant interaction between initial PHQ-9 score and

initial WSAS score in all models. Comparing the MLMs across the three Phases identified a reduction in the therapist effect between Phase 2 (4.7%) and Phase 3 (1.8%), with the largest therapist effect in Phase 1 (4.9%). In Phase 3 the variability between therapists was minimal, there was no therapist who significantly differed from the average therapist. In practice this meant that the range of average outcomes that patients might experience reduced from a range of 4 points on the PHQ, to a range of just under 2 points. In Phase 1 the higher effectiveness therapist was on average producing double the amount of clinical change on PHQ-9 than the lowest effectiveness therapist. In Phase 3 this difference was 1.8 points of change on the PHQ-9. This meant that the therapist a patient saw in Phase 3 mattered less than in Phase 1 and Phase 2. The reduction in therapist variability appeared to reside at the lower end of the effectiveness ranking, meaning that therapists with lower effectiveness in Phase 1 were getting better outcomes in Phase 3. It was observed that therapists moved between rankings across the Phases. Comparisons of PHQ-9 reliable improvement rates demonstrated significant improvement in this clinical outcome over the Phases, despite the difference in overall average outcome on PHQ-9 change being non-significant across the Phases. In sensitivity analyses with a larger sample of therapists, a similar pattern of reduction in therapist effect in Phase 3 was observed although a significant therapist effect was still present in Phase 3, and reliable improvement reductions were not significant. Due to being underpowered, MCMC 95% PrI were constructed around the effects to assess their reliability when comparing effects across Phases. These show that the effects found may not be reliably different from each other and that a larger sample of therapists is required to produce more accurate estimates. Comparisons of the therapist clusters in the stepped wedge stages indicated no significant difference between the DP-lite and reflective practice groups.

Chapter 8

Results 2 All Therapists

Overview

The previous chapter presented the results from a 'Core' population of therapists who had each treated a minimum of 10 patients in each Phase over the 4.5-year study period. This chapter presents the results from the 'All Therapists' dataset, providing a replication of the procedures of the Core Therapists analysis, to investigate the generalisability of the findings across the whole eligible therapist population in the service during the relevant Phases. Therapists included in this dataset must have treated a minimum of 10 patients during the specific Phase but may or may not have featured in all Phases across the 4.5-year period. This wider dataset therefore included the 35 Core Therapists, but also other therapists who met the eligibility criteria outlined in Chapter 6. This allowed exploration of whether the findings observed for therapists present in all three Phases, are also evident in a wider service context where therapists naturally enter and leave services.

Results

The following results begin with a descriptive analysis of the patient level data, including pre-post clinical scores and other relevant variables (the All Therapists Patient and Therapist Features are presented in Chapter 6). The identification of the significant variables for each Phase are then presented with each Phase Multilevel Model (MLM). Therapist residuals are presented graphically showing the variability across the therapist sample, and the Core Therapists are indicated to illustrate movement in this sample within the context of the larger sample. MCMC estimations are presented and compared with the Core Therapist sample for illustrative purposes as previously discussed. Finally, data collected by the service in relation to staff leavers and absence from all Phases, and the results of a staff questionnaire undertaken by the service during the intervention Phases are presented.

All Therapists: Patient Level Descriptive Analysis

Table 8.1 shows the main patient level variables for each Phase of the study.

Table 8.1

All Therapists Sample - Patient Level Descriptive Analysis

Study Phase	Pt N	Mean Patient Age in Years (SD)	Patient Gender (%)		Referral Source (%)				Patient IMD Quintile Group(%) (1 = highest deprivation,, 5= lowest deprivation group)				
			Female	Male	Self	GP	Mental Health Service	Other	1	2	3	4	5
1	1982	44.3 (15.80)	64.9	35.1	59.7	36.3	3.4	0.7	12.3	23.3	33.6	22.0	8.8
2	2227	39.7 (15.96)	67.0	33.0	69.5	26.8	3.2	0.5	11.7	21.7	33.7	20.7	12.2
3	2267	38.9 (15.31)	68.0	32.0	77.4	19.3	2.2	1.1	14.5	23.3	30.4	19.1	12.7

n.b. Referral Source 'other' includes: prison/probation services, employers, voluntary organisations

As seen in the Core Therapists datasets, the average age of the patients reduced significantly over the overall study period (F(2, 6473) = *71.91, p <*.001)*.* The difference was significant between Phase 1 and 2 (*t(*4207) = 9.35 *p <*.001) and between Phase 1 and 3 (*t*(4247) = 11.37, *p <*.001). There was no significant difference in ages between Phase 2 and 3 (*t*(4492) = 1.82, *p* = .069). Each Phase had a slightly lower average age than in the Core Therapists datasets. Again, as with the Core Therapists datasets, the majority of patients were female and this was the case in each Phase, between 64.9-68%, with no significant difference between Phases (X^2 (2) = 4.77, *p* = .092). The proportion of patients referred from particular sources was significantly different between the Phases (X^2 (4) = 165.89, *p* < .001). The differences were in line with the Core Therapists datasets, with the proportion of those who self-referred having significantly increased across all Phases (all p-values <.001), from 59.7% in Phase 1, 69.5% in Phase 2 to 77.4% in Phase 3, and a reduction in referrals originating from GPs, from 36.3% to 19.3% of patients. Deprivation scores were again consistent with the Core Therapists datasets, also showed a significant difference across the Phases, X^2 (8) = 34.97, *p* < .001, particularly between Phase 1 and Phase 3, X^2 (4) = 26.69, *p* < .001.

Table 8.2 reports the initial (pre) and final (post) scores and pre-post change scores on each of the clinical outcome measures, across the three Phases.

Table 8.2

All Therapists Sample – Pre/Post Clinical Scores

Study	PHQ-9			GAD-7			WSAS			Phobia		
Phase	Pre	Post	Change	Pre	Post	Change	Pre	Post	Change	Pre	Post	Change
1	15.0	9.06	5.9	13.2	7.9	5.3	20.3	14.3	6.3	9.1	6.5	2.6
	(6.15)	(6.81)	(6.39)	(5.20)	(5.82)	(5.80)	(9.29)	(10.36)	(9.62)	(6.68)	(6.54)	(5.97)
2	15.0	9.0	6.0	13.5	8.2	5.3	18.4	13.0	5.7	8.9	6.2	2.8
	(5.85)	(6.60)	(6.18)	(4.98)	(5.59)	(5.52)	(9.39)	(9.73)	(8.97)	(6.46)	(6.02)	(5.40)
3	14.8	8.8	6.0	13.5	8.0	5.5	19.0	13.0	5.8	9.7	6.5	3.2
	(5.85)	(6.27)	(6.00)	(4.87)	(5.48)	(5.51)	(8.84)	(9.72)	(8.44)	(6.38)	(6.06)	(5.35)

Differences between initial clinical measures mirrored those observed in the Core Therapists datasets. As such, these were small, though some significant differences were found on some measures; mean initial WSAS scores ($F(2,6473) = 22.74$, $p <.001$) with a significant decrease from Phase 1 to Phase 2 ($t(4207) = 6.50$, $p < .000$), increase from Phase 2 to 3 ($t(4492) = 2.14$, $p = .032$) and decrease from Phase 1 to Phase 3 ($t(4247) = 4.648$, $p = < .000$). As with the Core Therapists datasets, the initial Phobia scores were not normally distributed, and there was a significant difference in mean scores, K-W(2) = 22.45, $p < .001$, which contrasted with the Core Therapists datasets. Pairwise comparisons with adjusted p-values showed that the differences were significant between Phase 2 and Phase 3 ($p = .000$, $r = -.067$) and Phase 1 and Phase 3 ($p = .001$, $r = -.054$). As with the Core Therapists datasets there were no significant differences in mean change on any clinical measure in the All Therapists datasets, based on ANCOVA tests. Table 9.3 compared the three Phases on the PHQ-9 reliable improvement rate, whether therapy was completed and number of sessions attended.

Table 8.3

All Therapists Sample – Clinical Indicators: PHQ-9 Reliable Improvement Rate, Therapy

Completion and Sessions Attended

Phase	PHQ-9 Reliable Improvement Rate N (%)	Therapy Completion Rate N (%)	Session Number (SD)		
			Attended	Missed	Offered
1	974 (49.1)	1226 (61.9)	7.2 (3.95)	1.8 (2.02)	9.0 (4.73)
2	1149 (51.6)	1648 (74.0)	7.1 (3.94)	2.0 (2.05)	9.1 (4.86)
3	1157 (51.0)	1764 (77.8)	8.6 (4.58)	2.3 (2.25)	11.0 (5.55)

Other clinical indictors shown in Table 8.3 also reflected the observations made of the Core Therapists datasets, other than the reliable improvement rate. There was no significant change in PHQ-9 Reliable Improvement rate (X^2 (2) = 2.73, $p = .255$), contrasting with the significant increase in reliable improvement rate for Core Therapists. As with the Core Therapists, in Phase 3 the average number of attended sessions increased significantly (K-W(2) = 173.07, $p < .001$), between Phase 2 and 3 ($p < .001$, $r = -0.15$) and Phase 1 and 3 ($p < .001$, $r = -0.13$). Average missed sessions also increased significantly between all Phases (K-W(2) = 74.47, $p = < .001$). Average therapy completion rate increased significantly (X^2 (2) = 141.92, $p < .001$) between all Phases, from 61.9% to 77.8% of patients completing, similar to the increase from 64.6% to 76.5% in the Core Therapists Phase 1 to Phase 3 datasets. Contrasting with the Core Therapists sample however, the therapy completion rate in Phase 3 was higher by 3.8% points compared to Phase 2, whereas the Core Therapists sample saw therapy completion reduce in Phase 3 by 1.8% points, compared to Phase 2.

Identification of Common Predictor Variables

As with the Core Therapists analysis, the same variables were significant or non-significant across all datasets, with the exception of IMD level, which was a significant predictor variable in Phase 1 only. IMD level was non-significant in both other Phases and so was excluded from the comparison model, to ensure the model was controlling for the same significant variables across each Phase. The significant variables identified were consistent with those identified in the Core Therapists analysis, including an interaction between initial PHQ-9 score and initial WSAS score. In the All Therapists sample, Initial Phobia score was a significant predictor variable in all Phases, contrasting with the Core Therapists sample where it was a non-significant in Phase 2. The final MLM for each study Phase is shown in Table 8.4, including the therapist effect (full MLMs and tests of model assumptions are in Appendix K and Appendix L respectively).

Table 8.4

Estimates from Multilevel Models

MLM Values	Phase 1			Phase 2			Phase 3		
	Value	*S.E.*	*p*	Value	*S.E.*	*p*	Value	*S.E.*	*p*
Average therapist PHQ-9 Change (Intercept)	5.97	*0.20*	*0.00*	6.03	*0.21*	*0.00*	6.19	*0.17*	*0.00*
Initial PHQ-9	0.53	*0.03*	*0.00*	0.56	*0.03*	*0.00*	0.56	*0.02*	*0.00*
Initial Phobia	-0.09	*0.02*	*0.00*	-0.06	*0.02*	*0.01*	-0.07	*0.02*	*0.00*
Initial WSAS	-0.08	*0.02*	*0.00*	-0.10	*0.02*	*0.00*	-0.09	*0.02*	*0.00*
Interaction PHQ-9 / WSAS	-0.00	*0.00*	*0.04*	-0.01	*0.00*	*0.00*	-0.01	*0.00*	*0.00*
Level 2 (therapist) Variance	1.44	*0.45*		2.05	*0.51*		0.94	*0.31*	
Level 1 (patient) Variance	31.50	*1.02*		28.77	*0.88*		27.06	*0.82*	
Therapist Effect	4.4%			6.6%			3.4%		

The patient level variance consistently reduced over the study period in both the Core Therapist and All Therapists datasets. In contrast to the Core Therapists models however, the therapist level outcome variance increased, rather than decreased in the All Therapists

models, from 1.44 in Phase 1 to 2.05 in Phase 2. A decrease in the therapist level outcome variance, consistent with the Core Therapists, was observable in Phase 3 of the study (0.94). However, although this outcome variability reduced, in contrast to the Core Therapists analysis the variation between therapists in Phase 3 remained significant in the All Therapists model. The therapist effect was 4.4% in Phase 1, 6.6% in Phase 2 and 3.4% in Phase 3. This pattern is reminiscent of that seen in the expanded Core Therapists sample, which also observed the highest therapist effect in Phase 2. In terms of the difference in therapist variance comparing Phase 2 and Phase 3, however, the therapist variance reduced by 54.2% in Phase 3 in the All Therapists sample and by 64.4% in the Core Therapists sample, indicating a similar therapist effect reduction between Phase 2 and Phase 3 in both samples.

Table 8.5

Key Data from the MLM for Each Study Phase of Both Core and All Therapists Samples

Phase	Dataset	Therapist N	Patient N	Predictive Model – Severity & Functioning (Initial PHQ, WSAS & Phobia w. interaction PHQ/WSAS)				
				Therapist Effect (%)	Average Therapist PHQ Change (SE)	Total Variance	Patient Level Variance (SE)	Therapist Level Variance (SE)
1	**Core**	35	930	4.9	6.13 (0.30)	30.88	29.38 (1.39)	1.50 (0.66)
	All	81	1982	4.4	5.97 (0.20)	32.93	31.50 (1.02)	1.44 (0.45)
2	**Core**	35	1226	4.7	6.66 (0.26)	28.09	26.77 (1.10)	1.32 (0.52)
	All	80	2227	6.6	6.03 (0.21)	30.82	28.77 (0.88)	2.05 (0.51)
3	**Core**	35	1217	1.8	6.86 (0.21)	25.69	25.22 (1.04)	0.47 (0.29)
	All	74	2267	3.4	6.19 (0.17)	28.00	27.06 (0.82)	0.94 (0.31)

Table 8.5 shows the therapist effect, average PHQ-9 change (model intercept) and the patient and therapist level outcome variance for each model for both Core Therapists and All Therapists datasets. In contrast to the Core Therapists sample as mentioned above, the change in therapist outcome variance increased in Phase 2, where it decreased during this Phase in the Core Therapists model. In all Phases, the therapist effect was significant at the

0.001 level as indicated by improvements in model fit when therapist variability was introduced. For Phase 1 the -2*loglikelihood ratio of the model reduced by 31.73 ($p < .001$), for Phase 2 ($X^2(1) = 66.73$, $p < .001$), for Phase 3 ($X^2(1) = 20.21$, $p < .001$). The remaining pattern of findings from the Core Therapists analysis is replicated in this broader All Therapists dataset.

All Therapists Residual Charts

Figure 8.1 plots therapist residuals generated by the model with their 95% confidence intervals for the All Therapists datasets. The values have been transformed to show the impact on actual PHQ-9 change scores. In Phase 1 there are three therapists of significantly higher effectiveness than the average therapists, compared to two in the Core Therapists dataset. There are no lower effectiveness therapists in this dataset, compared to one in the Core Therapists Phase 1 dataset. In Phase 2 there are seven therapists of significantly higher and seven therapists of significantly lower effectiveness than the average therapist. This compares to two higher and one lower effectiveness therapists in the Core Therapists Phase 2 dataset. In Phase 3, there are two therapists of higher effectiveness and one of lower effectiveness, though the overall variability is less than either of the other Phases.

As with the Core Therapists models, in Phase 1 the average PHQ-9 change for the most effective therapist (8.1) is almost double that of the least effective therapist (4.1). This difference is even larger in Phase 2 in the All Therapists model, with the average patients PHQ-9 score change for the most effective therapist being 8.2, and the least effective therapist 3.4 average points of change. In Phase 3 this range has reduced from 7.9 average points of change for the most effective therapist, to 4.7 for the least effective therapist, so similar to, though with a slightly narrower range than Phase 1.

Figure 8.2 shows the plots without transforming the residuals to PHQ-9 change, with the 35 Core Therapists highlighted in red. In Phase 1, 18 of the Core Therapists average PHQ-9 change scores were residing above the residual line. As the variability across All Therapists increases in Phase 2, a clustering of Core Therapists around the mid and upper section is observable.

Figure 8.1

Ranked Therapists Showing Average PHQ-9 Change for Each All Therapists Caseload with 95% Confidence Intervals (CI)
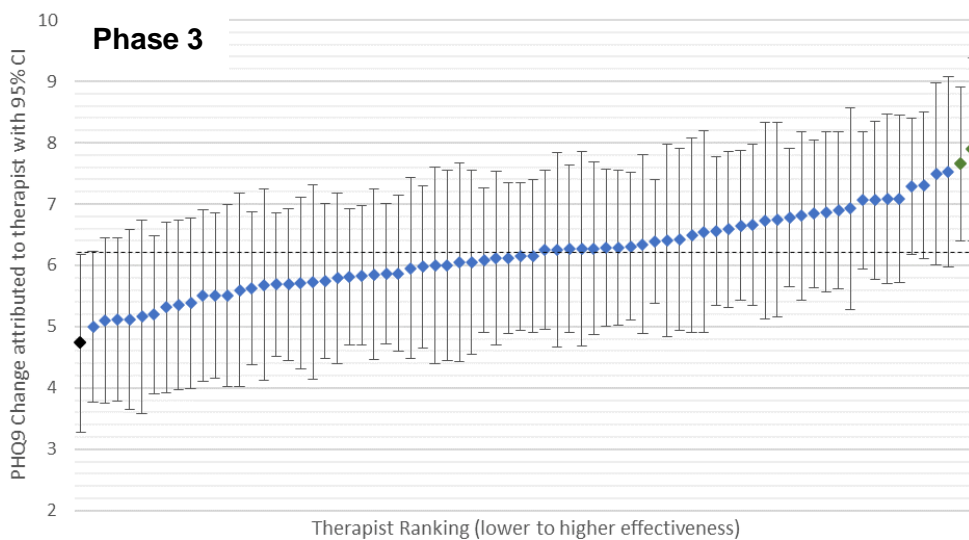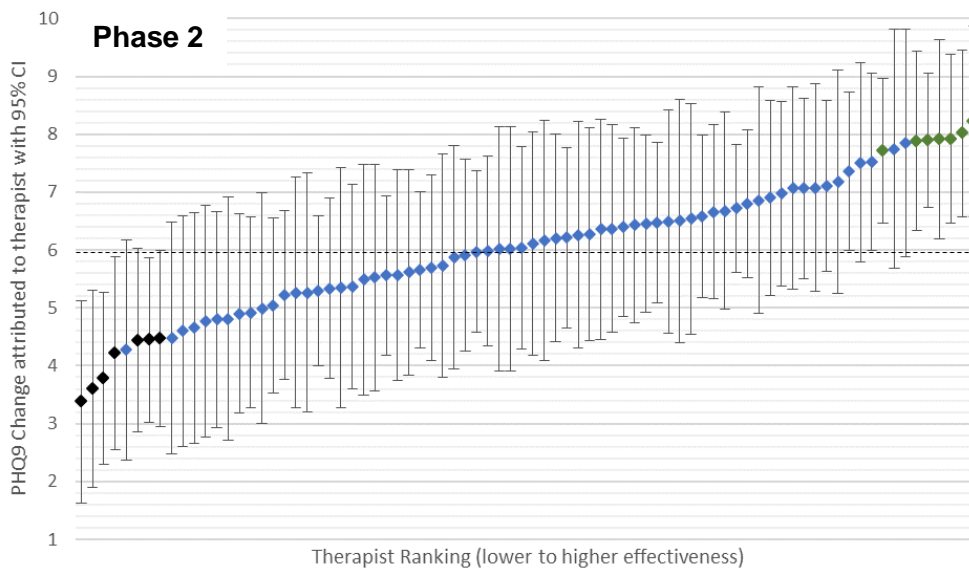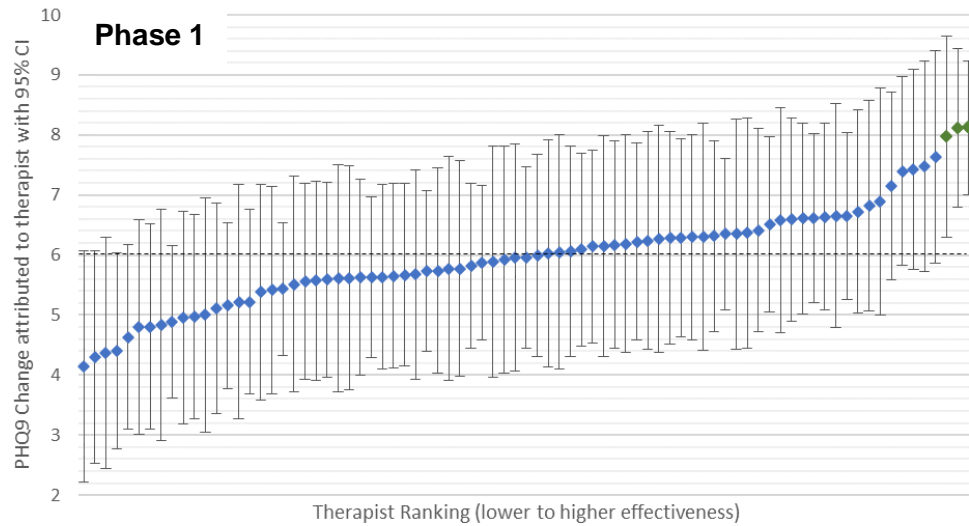
Legend

◆ Average effectiveness   ◆ Higher effectiveness   ◆ Lower effectiveness

Figure 8.2

Ranked Therapists Showing All Therapists Residuals with 95% Confidence Intervals (CI) with Core Therapists Highlighted

Phase 1

Therapist Residuals with 95%CI

0 = 5.975

Therapist Ranking (lower to higher effectiveness)

Phase 2

Therapist Residuals with 95% CI

0 = 6.035

Therapist Ranking (lower to higher effectiveness)

Phase 3

Therapist Residuals with 95%CI

0 = 6.195

Therapist Ranking (lower to higher effectiveness)
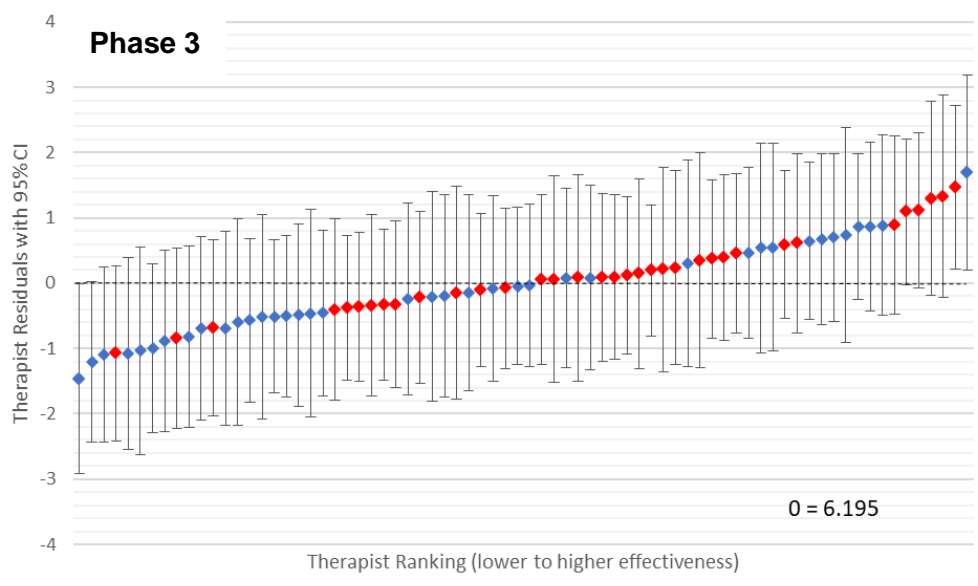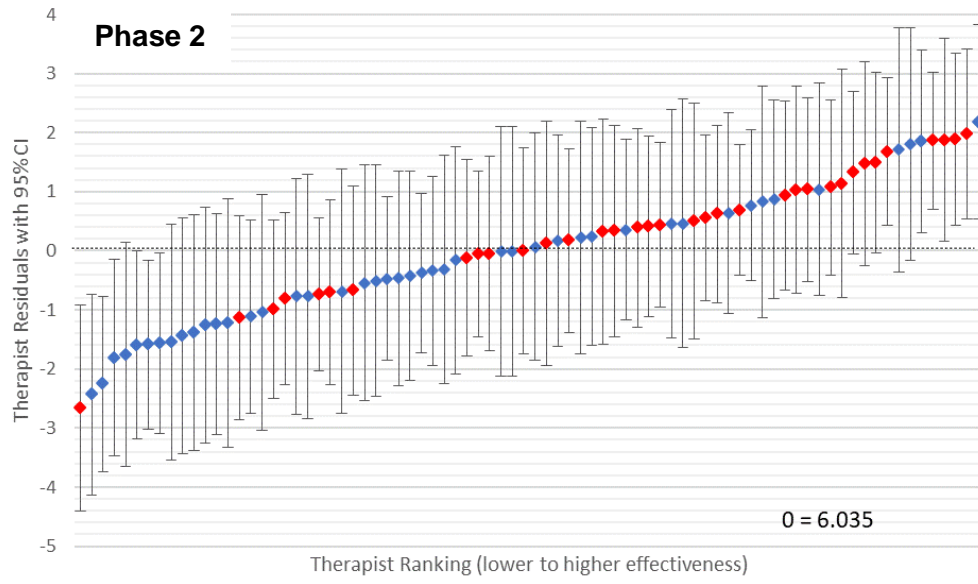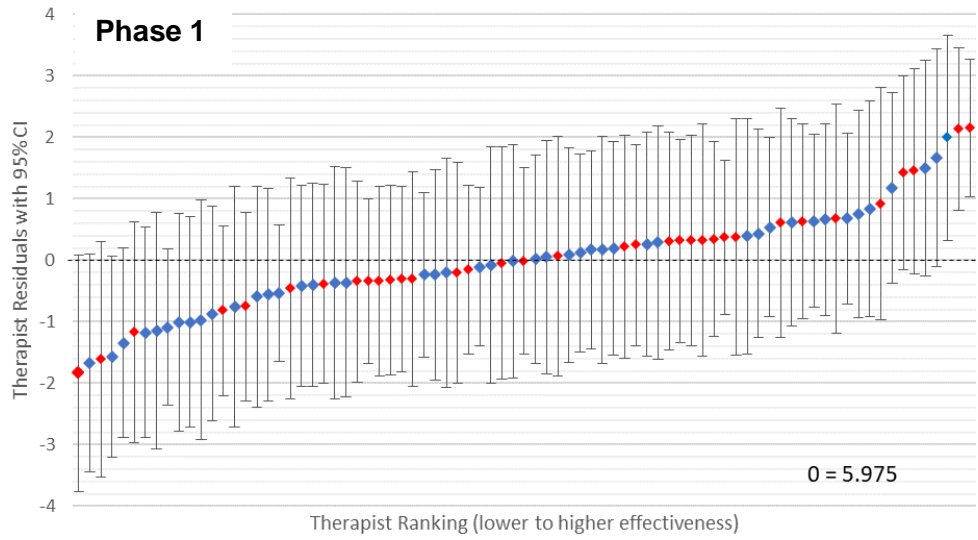
◆ In All Therapists sample only   ◆ Core Therapists

In this Phase there are 25 Core Therapists with average PHQ-9 change scores residing above the residual line. In Phase 3, as the therapist variability drops to its lowest point in the All Therapists dataset, there is more of a spread of Core Therapists around the mid point, with 22 Core Therapists residing above the residual line.

MCMC Estimations

As with the Core Therapists datasets, MCMC estimations of the therapist effects were calculated with 95% PrIs. These are plotted in Figure 8.3 alongside the Core Therapists estimates. In both datasets, the 95% Probability Intervals (PrIs) of each estimation overlap, so the effects found cannot be considered reliably different. Taken together, these plots show a reduction in the therapist effect in Phase 3 compared to Phase 2 and Phase 1 and illustrate the increase in therapist effect in Phase 2 compared to Phase 1 for the All Therapists model, in contrast to the slight reduction of the therapist effect in Phase 2 compared to Phase 1 for the Core Therapists model. It should be noted that the PrIs in Phase 3 in both samples are narrower, indicating increased reliability of the therapist effect in this Phase.

Figure 8.3

MCMC Therapist Effect with 95% Probability Intervals (PrI) for Each Study Phase for the Core and All Therapists Datasets
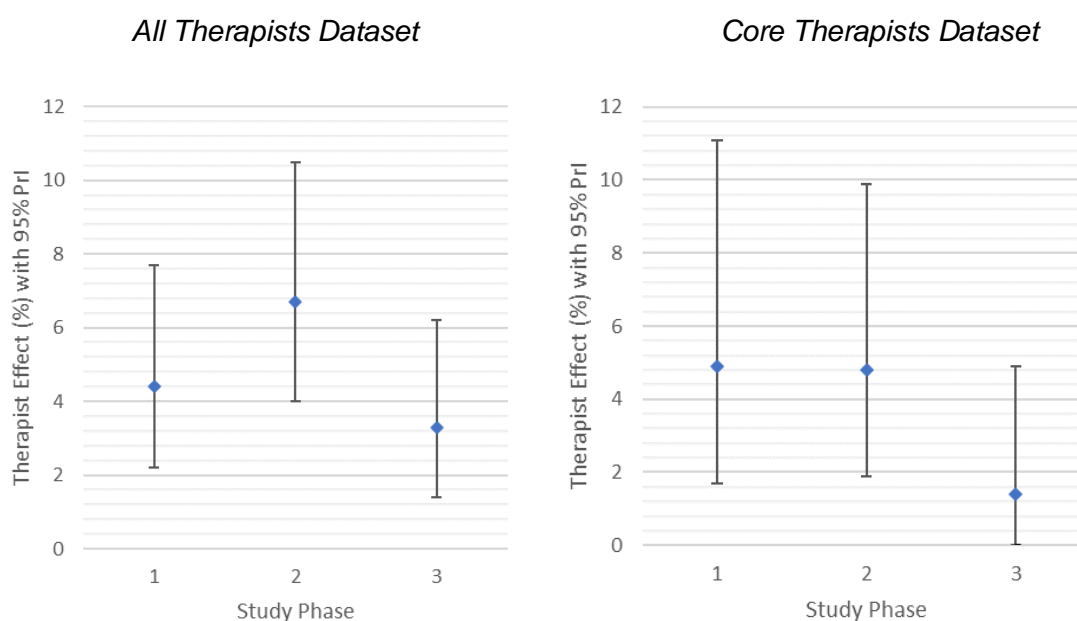
All Therapists: Additional Analysis

To consider further any effects on therapists across the Phases, sickness and retention rates were obtained from the service and compared across the Phases as shown in Table 8.6. This shows the sickness rate was reasonably consistent across the Phases although it was higher in Phase 3. There was an increase in staff leaving the service in Phase 2 and 3, however the service was growing in size (increasing staff numbers) during these Phases.

Table 8.6

Sickness Rates and Staff Leavers Over the Study Phases

	Phase 1	**Phase 2**	**Phase 3**
	Oct 15 - Mar 17	Apr 17 - Sept 18	Oct 18 - Mar 20
Days Lost	1895.51	2063.44	2540.46
Days available	57505.92	64705.67	72831.11
Sickness rate	3.3%	3.2%	3.5%
Staff Leavers	20	28	31

Table 8.7 and 8.8 below are data provided by the service, showing results from internal surveys carried out every three months for a two-year period, administered via anonymous survey to all staff (Step 2, Step 3, peer support work, admin, management). This two-year period spanned the entirety of Phase 2, and the first six months of Phase 3.

Table 8.7

Staff Responses to Internal Survey: Levels of Work-Related Stress Apr 2017 – Mar 2019

	During the last 12 months have you felt unwell as a result of work related stress?															
	PHASE TWO												**PHASE THREE**			
	Apr-Jun 2017		Jul – Sep 2017		Oct – Dec 2017		Jan- Mar 2018		Apr-Jun 2018		Jul-Sep 2018		Oct-Dec 2018		Jan – Mar 2019	
	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%
Yes	48	61	31	63	69	62	57	68	48	67	21	49	12	31	20	51
No	31	39	18	37	43	38	27	32	24	33	22	51	27	69	19	49

During the Phase 2 period between 49-68% of staff who answered reported having felt unwell as a result of work-related stress in the last 12 months. This had dropped to 49% in the last three-month period of Phase 2. In the first six months of Phase 3 this rating ranged between 31% and 51%. In relation to actions being taken by the service to improve staff wellbeing, in Phase 2 between 20-39% of staff who responded felt that the service did not take positive action in this regard. In the first 6 months of Phase 3 this ranged from 14% to 20%.

Table 8.8:

Staff Responses to Internal Survey: Perceived Service Investment in Staff Wellbeing Apr 2017 – Mar 2019

	Does your organisation/service take positive action on health and well-being?															
	PHASE TWO												**PHASE THREE**			
	Apr-Jun 2017		Jul – Sep 2017		Oct – Dec 2017		Jan- Mar 2018		Apr-Jun 2018		Jul-Sep 2018		Oct-Dec 2018		Jan – Mar 2019	
	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%
Yes, Definitely	11	14	6	11	6	5	10	12	9	12	3	5	13	21	8	14
Yes to some extent	42	54	33	60	80	72	39	49	43	58	44	75	36	59	40	72
No	25	32	16	29	25	23	31	39	22	30	12	20	12	20	8	14

Summary

This second results chapter presented the results of a broader dataset (All Therapists) incorporating the Core Therapists and other eligible therapists in the service over the study period, for the purpose of investigating the generalisability of the findings of Chapter 8 across a whole service cohort. The patient level descriptives of the All Therapists datasets were consistent with the Core Therapists datasets, showing that the two patient populations were similar and comparable. In addition, the key MLM predictive variables were also consistent across Core and All Therapists datasets. As with the Core Therapists datasets, a reduction in the therapist effect was also apparent in the All Therapists dataset in Phase 3 however, unlike the Core Therapists sample the therapist effect remained

significant, which may be due to the larger sample size producing more reliable estimates of effects. Alternatively, the additional therapists, some of whom had no intervention, while others had only one, may have introduced greater variability between therapists. The All Therapists dataset in Phase 2 had greater therapist variability than in Phase 1, which was in contrast to the Core Therapists datasets which showed a reduction in variability in this Phase. Overall, the reduction in the therapist effect between Phase 2 and Phase 3 was consistent in both datasets. Measures of clinical outcomes, PHQ-9 reliable improvement rates and mean PHQ-9 change, did not demonstrate significant improvements over the Phases and could therefore be considered consistent. Additional data provided by the service did not indicate any significant difference in sickness rates or staff retention across the Phases, however there was some limited indication that staff were less likely to feel unwell due to work stress in Phase 3 compared to Phase 2. In addition, staff who responded to the survey had a perception that the organisation/service was taking more positive action on health and wellbeing in Phase 3 compared to Phase 2. Although the additional analysis cannot determine a causal effect, the Phase 3 intervention did not appear to have any negative effects on staff wellbeing and during this Phase there appears to have been some improvement.

Chapter 9

Discussion

Overview

This chapter presents a summary of the findings from this study and considers the potential factors that may account for these results. An overall summary is followed by a focus on the primary finding which directly addresses the research questions of this thesis. Additional key findings are presented and considered similarly. Limitations of the study design are presented, followed by the implications of the findings of this study to psychological therapy research, individual therapists' practice and psychological therapy service management.

Summary of Key Findings

Using routinely collected outcome data, this longitudinal study comprising three distinct Phases – Baseline, Directive Intervention, and Collaborative Intervention – investigated the variability in therapist outcomes within a primary care psychological therapies service in England. As a proof of concept study, it aimed to determine whether therapist variability – the therapist effect – could be reduced and absolute patient outcomes improved in a practice setting following a Collaborative Intervention. The size of therapist effect and pre-post change in PHQ-9 scores were compared across Phases for a stable group of therapists (Core Therapists) who were present across all three Phases, as well as a sample termed All Therapists which comprised all therapists in each individual Phase meeting other inclusion criteria. Four significant co-variates, all measures of intake severity (PHQ-9, WSAS, Phobia Scale, and the interaction between PHQ-9 and WSAS) were found to be largely consistent across all models and were selected as control variables in all models.

Core Therapists

For the therapists who were present during all Phases and had experience of both interventions, PHQ-9 outcome results showed a significant improvement in PHQ-9 reliable improvement rates across Phases, but no significant change in mean PHQ-9 change. The therapist effect reduced across the Phases, notably between Phase 2 (4.7%) and Phase 3

(1.8%), a reduction in therapist variability of 64.4%. Overall, indications suggested a minimal therapist effect in Phase 3, with this being a much smaller therapist effect than in previous Phases. There was a smaller reduction in therapist effect between Phase 1 (4.9%) and Phase 2. However, the level of uncertainty around these estimates, as would be expected with a sample of this size, indicates that they may not be reliably different from each other.

The results also showed that in Phase 3 patients were offered, on average, one more session and also attended one more session compared with both Phase 1 and Phase 2. However, there was a very small reduction in therapy completion rates between Phase 2 (78.3%) and Phase 3 (76.5%) which followed a large increase from Phase 1 (64.6%). This means that although the therapist variability for PHQ-9 outcomes reduced by only a small amount between Phase 1 and Phase 2, the overall likelihood of patients completing therapy increased by 14%.

All Therapists

The pattern of results was broadly similar when therapists who may have experienced only one or neither of the interventions were included. Differences in pre-post change calculated by reliable improvement rates on PHQ-9 were non-significant, as were mean PHQ-9 change scores, while there was a reduction in therapist effect between Phase 2 and 3 (from 6.6% to 3.4%), a similar percentage reduction to the Core Therapists sample. However, unlike the Core Therapists sample, the therapist effect in Phase 3 remained statistically significant on all measures. Therefore, although the variability between the Core Therapists did not affect outcome in Phase 3, the variability between all therapists in the service did (i.e., there was greater therapist variability where therapists who had not experienced both interventions were included). In contrast to the Core Therapists sample, the therapist effect increased in Phase 2 to 6.6% from 4.4% in Phase 1. There was, therefore, an increase in the variability between therapists in Phase 2 that was present in the All Therapists sample but not in the Core Therapists sample. Like the Core Therapists sample, using MCMC estimate of therapists' effects with 95% PrIs indicated that there was insufficient reliability of the therapist effects to determine distinct differences between them,

but again this may be expected due to the size of the samples involved. In both the Core Therapist and All Therapist sample there was increased reliability of the therapist effect in Phase 3.

The increases in sessions offered and attended across Phases were similar to the Core Therapist sample. With regard completion rates, unlike the Core sample, the rate was highest in Phase 3, 77.8% compared with 74.0% in Phase 2 and 61.9% in Phase 1.

Answering the Research Questions

This study set out to answer the following questions:

1. Can a package of therapist-focused, bespoke, research-based actions be implemented to reduce therapist variability and improve patient outcomes in routine practice, when controlled for and analysed across a stable group of therapists over time?

2. Can such a package yield reductions in therapist variability and improved patient outcomes, when analysed at a service-wide level (i.e., accommodating the natural turnover of therapists within a service)?

The finding of reduced therapist variability in Phase 3 compared to Phase 2 and Phase 1 in both samples means that two patients receiving therapy from the service during Phase 3, with similar intake severity and complexity, would have more consistent outcomes irrespective of the individual therapist that provided their therapy than in the previous Phases. This suggests that therapist variability in outcomes can be reduced, both in a stable group of therapists, and at a service-wide level, across an entire service-wide therapist cohort. However, where the therapist effect was non-significant in the Core Therapists in Phase 3, this was not the case with the All Therapists cohort. So, although a reduction in variability may be possible, reducing the therapist effect to a non-significant level at a service-wide context was not demonstrated. In addition, although the therapist effects in Phase 3 were smaller, when making comparisons across Phases and including a measure of uncertainty around effects, the effects in each Phase may not be significantly different from each other. However, if taking the best estimates of effect, as usually reported in other

therapist effect studies, a reduction in therapist variability, such as seen in this study, has considerable impact in practice, and is likely to be of interest to service managers and commissioners. Therapists in Phases 1 and 2 obtained average outcome improvements with their patients ranging from 4 to 8 PHQ-9 points of change, compared to the narrower range from 6 to 7.8 PHQ-9 points of change in Phase 3. The reduction in therapist variability appeared to reside at the lower end of the effectiveness ranking, meaning that therapists with lower effectiveness in earlier Phases were getting better outcomes in Phase 3. This amounted to a 'levelling out' of outcomes, thus reducing any therapist lottery that existed in previous Phases.

With reference to the first research question, in a stable group of therapists, the variability between therapists reduced to a minimal level, and there was no significant difference in outcomes between the therapists in this cohort in Phase 3. In relation to the second research question, whilst a significant therapist effect remained in Phase 3 in the broader therapist sample due to the initial therapist variability being larger, a similar reduction in the therapist effect was observed as was seen in the Core Therapists sample. In relation to an improvement in patient outcomes, in a stable group of therapists, reliable improvement rates improved significantly on the primary measure from Phase 1 to Phase 3, with under 50% of patients reliably improving in Phase 1 increasing to 56% in Phase 3. However, the increase in reliable improvement rates in the broader All Therapists sample was non-significant. There was no significant change in mean improvement on the primary measure between Phases in either sample. The results suggest that there is sufficient indication that it may be possible to reduce therapist outcome variability whilst maintaining patient outcomes, both in a stable therapist sample, and in a broader service-wide sample, to warrant further investigation.

Based on the premise that these results, whilst not possible to definitively confirm without larger cohort samples, are indicative of some positive practical improvement in therapist variability in Phase 3 for a stable group of therapists, and hence are potentially

important for service managers working in clinical settings, the following sections will discuss potential explanations for the main findings.

Explanations for Findings

As the study Phases were designed to test the impact of a series of changes, or interventions within the context of a live and dynamic service, it is not possible to attribute causality to the research findings. However, given the summary above, there are several accounts that could provide some explanation for the findings. The design of the study was such that it set out to compare therapist outcomes following an externally driven service improvement programme (Phase 2), with the research intervention (Phase 3) that was implemented after a period of service stabilisation. In this way, the primary comparator was between these two Phases, as both Phases involved service changes aiming to improve outcomes. This meant that change in outcomes in Phase 3 could more reliably be attributed to the addition of the specific Phase 3 improvement components, rather than the effect of *any* service improvement.

Following the Phase 1 Baseline period, there were two 18-month long intervention Phases: the *Directive Intervention* and the *Collaborative Intervention*. The Directive Intervention Phase in the service was characterised by rapid, externally driven, top-down (directive) change and acted as a primary comparator for the third Phase, as a quasi-control for the effect of any service intervention on the baseline outcomes. The changes in the Directive Intervention Phase included improvements to the patient assessment process and reduction in workshop-based interventions, with the purpose of improving NICE treatment compliance (Clark, 2011, 2018; Clark et al., 2009; DH, 2012b; Gyani et al., 2013; Parry et al., 2011), reducing waiting times for treatment and changes to clinical supervision ensuring an IAPT compliant supervision model including use of patient outcome feedback (Clark, 2011; Parry et al., 2011).

The Collaborative Intervention Phase involved the implementation of a specific intervention, developed collaboratively between the staff and the research team, supported by the management team. This intervention applied ideas from the deliberate practice (DP)

literature and involved the implementation of monthly small (3 to 4 therapists) DP (Ericsson et al., 1993) peer-group skills sessions, where therapists practiced skills that had been identified by themselves and/or their clinical supervisor as being personal areas for development (referred to as DP-lite). The purpose of the intervention was to improve therapist confidence (Hill et al., 2020) as well to further develop therapeutic skills. In addition, a series of staff wellbeing events were undertaken across the service during this period, including training on resilience and opportunities to pursue self-care activities. The Collaborative Intervention Phase was characterised by a change to the management and service culture to incorporate collaborative, compassionate leadership models (Dethmer et al., 2014; Kofman, 2002; West et al., 2014, 2017) with a focus on staff wellbeing.

Primary Finding: Reduction in Therapist Effect in Phase 3 in Both Samples with Improved Clinical Outcomes

The finding of a reduction in therapist effect in Phase 3 suggests that therapist variability in practice may fluctuate over time but may be reduced by low-cost interventions focusing on all or a mixture of the following: micro-skills development/use of deliberate practice, peer-learning, developing a compassionate service culture which aims to build staff confidence, resilience and wellbeing. In addition to a reduction in therapist variability, a significant increase in reliable improvement rates on the primary measure was observed between Phase 1 and the subsequent Phases for those therapists present in all Phases. This increase in reliable improvement was non-significant in the broader therapists sample and was not reflected in improvement in mean change scores in either sample.

Use of Deliberate Practice (DP) or Deliberate Practice-Lite (DP-lite). As summarised above and discussed in more detail in Chapter 5, a main feature of the Collaborative Intervention in Phase 3 was the application of a 'lite' form of DP in small therapist peer groups. This was derived from combining feedback from therapist staff requesting more time to reflect on their practice and learn from each other in groups, and the findings from a systematic literature review (Chapter 2) indicating some potential for the use of DP within the field of psychotherapy. A DP approach was different to the supervision and

skills development otherwise provided by the service, which utilised more traditional and/or IAPT-compliant one-to-one supervision and case management (NHSE & NCCMH, 2021) involving the *discussion* of cases, rather than the *practice* of specific, transdiagnostic therapy micro-skills. However, as briefly mentioned in Chapter 5, given the aspects of DP that were arguably absent from the package, such as expert feedback and individual, solitary repeated practice (Miller et al., 2020, p.11; Ericsson & Lehmann, 1996), it may be more accurate to describe it as 'purposeful practice' rather than DP-lite. Purposeful practice involves engaging in a focussed practice activity to achieve a specific goal, getting immediate feedback, and repeating this until achieving the goal, which would correspond largely with the activities encouraged and undertaken in Phase 3 (Ericsson & Harwell, 2019; Ericsson & Poole, 2016).

Within the context of this study, DP-lite had the addition of the identification of the specific goal or skills to be developed through therapist self-evaluation, and supervisor evaluation, making it slightly more targeted than purposeful practice, though still falling short of pure DP for the reasons aforementioned. However, for both this intervention and for the implementation of pure DP, there is evidence that there is limited agreement on judgements of competence between supervisors/independent raters (Dennhag et al., 2012; Hill et al., 2017), making this aspect of the model problematic within psychotherapy. In summing up the limitations of using DP within psychological therapy practice, Ericsson and Poole (2016) argued that in a field that is not sufficiently well-developed, and without an absolute measure of performance and where performance is complex to measure or define, such as in the field of psychotherapy, 'true' DP may not be possible. Despite this, the findings of this study indicate that it is possible to introduce the ideas of DP into a service at low cost, and this may have a positive impact on therapist variability.

Due to the DP-lite intervention being implemented in a stepped wedge design, this allowed the comparison of therapist clusters under the same service conditions but with or without the DP-lite element of the peer group package. This analysis was limited by small numbers but overall did not demonstrate any significant impact of having DP-lite in addition to the peer group itself. This could suggest that the impact on therapist variability may be

simply the presence of a peer group skills session, not necessarily the DP element of the group. It should be noted, however, that there were peer supervision groups as part of the supervision model in Phase 1 where the largest therapist effect was reported for Core Therapists, and second largest for All Therapists, which may appear to contradict this as an explanation for the findings. It is possible, however, that both aspects together may have accounted for the potential impact; that is, collective DP, or DP in groups, impacts between-therapist variability through the learning of skill application from others, as well as the self-practice of such skills oneself.

Use of Feedback. In many ways it is not possible to disentangle the key elements of DP. However, there has been a growing interest in the provision of patient outcomes/feedback to therapists with a view to improving outcomes (this was explored in more depth in Chapters 2 and Appendix A). Brattland et al.'s (2018) naturalistic randomised trial discussed in the systematic literature review reported a positive impact on patient outcomes where therapists were provided with sessional feedback on patient ratings of their symptoms and experience of the session. These effects were observed even when therapist variability and initial severity of outcomes were controlled. However, the feedback condition also involved therapists deliberately practicing skills based on the feedback from their patients, so could not be considered a pure feedback condition. Despite this, in light of the findings of the present study it poses the question of whether the use of feedback from routine outcome measures, introduced in Phase 2 had any bearing on the findings, particularly the increase in reliable improvement rates demonstrated in the Core Therapists sample over Phases 2 and 3.

Brattland et al. (2018) reported there being little effect on patient outcomes initially but that the positive impact of using feedback increased over time, month-on-month in their four-year study. Firstly, the design of the length of the current study Phase periods, of 18 months duration, were such that they would reduce, though not eliminate, the risk of this possibility. In addition, it should be noted that, although the use of patient feedback in supervision was introduced in Phase 2, this was used as a measure of quality control or

performance primarily monitoring, and linked to, service targets. The use of feedback for the explicit purpose of quality control has been observed to potentially have a negative impact on patient outcomes (Wolpert, 2014), although this did not appear to be the case in the current study, where improved clinical outcomes were observed following the introduction of feedback into supervision (in Phase 2). It was noted that an important feature of the ROMs used in the Brattland et al. (2018) study was that the impetus was clinically driven rather than for use as a quality measure. In Phase 3 of the current study the use of patient feedback through sessional measures was recontextualised with a clinical and learning rationale, rather than as a therapist performance tool, and was implemented in a collaborative way with staff. If indeed the use of patient feedback to therapists does explain some of the indicative results in this research in relation to therapist variability, this would seem to support the argument made by Brattland et al. (2018). The authors imply that it is the way in which patient feedback is provided, and the explicit purpose for which it is being used, that is as or more important than the use of patient feedback per se. This may explain why the impact on the therapist effect appeared to be observed in Phase 3 but not Phase 2 despite ROMs being used in both Phases. However, the results of the current study suggest that use of patient feedback either as a quality measure *or* explicitly for clinical purposes, in a stable group of therapists, may contribute to improved clinical outcomes.

How feedback is used was also a focus for a follow-up to the 2018 study (Brattland et al., 2019) reporting on the mediating impact of the working alliance in therapy on the effects observed in the original study. In this follow up study a much larger therapist effect was reported and working alliance did appear to be improved in the feedback condition. Within the conclusions drawn from the study, the authors indicated that improving therapists' practice may involve supporting therapists in how to respond to the feedback they receive from their patients, rather than simply receiving the feedback itself, again a similar conclusion to that of the original study. It is also important to note that the improvements reported by Brattland et al. (2018) were in relation to patient outcomes and no reduction in the therapist effect was found in their study. This contrasts with the current study which, for a

stable group of therapists found a significant improvement in clinical outcomes following the introduction of a range of service improvements, including the use of feedback, and a reduction in therapist variability only appearing in Phase 3. A strength of the current study in providing indicative support for the conclusions made by Brattland et al. (2019) may be that the Phase 3 intervention focussed on using patient feedback *for the purpose of individual therapist skills development within a supportive, reflective practice environment*. Given the reduction in variability observed during this Phase, it may suggest that to provide both better, *and* more consistent clinical outcomes across a service, the use of feedback and consideration of how the feedback is presented contextually to therapists, may be important. This may have implications for the use of sessional patient feedback and support structures around this feedback for therapists within psychological therapies services. This point is discussed further below.

Collaborative Approach. During Phase 3 there was a focus and training on conscious/compassionate leadership (Dethmer et al., 2014; Kofman, 2002; West et al., 2014, 2017) across the senior leadership team in the service. Although not a direct part of the Collaborative Intervention, this was a deliberate attempt within the management team to address the low staff morale that had developed during the Phase 2 directive work by developing an increasingly staff-focused, compassionate environment within the service. This was in some ways linked and grown as part of the staff wellbeing initiative and events, which all contributed to a generally more compassionate approach to staff across the service during Phase 3. In addition, the focus on collaborative working within specified areas of expertise (i.e. the delineation of distinct Step 2 and Step 3 teams) promoted through the management structure to the therapist teams may have provided an increased sense of shared knowledge and expertise (Friedrich, 2009). This collective expertise in turn promotes a shared purpose and value, both areas deemed important in achieving positive organisational outcomes (Carmeli & Schaubroeck, 2006; West et al., 2003, 2017). The changes to the management structure and culture are unlikely to have had a direct impact on patient outcomes, but it is worth noting in terms of a) ease of implementation of the

collaborative package due to increased support from management; b) involving staff in the development of the service appeared to increase positive engagement in the interventions, compared to the disengagement and low staff morale of the Directive Intervention Phase; and c) allowed the development of a supportive environment for staff where their wellbeing was valued by management and considered within decision making and change management. This was in addition to the more explicit focus on staff wellbeing within the wellbeing events. Brattland et al. (2018) note that support from management for a therapist intervention, such as the one they implemented, is a key component for success. Alongside the evidence of a lack of staff engagement and work satisfaction relating to poorer outcomes within the Delgadillo, Saxon et al. (2018) study, this suggests that the change of culture in Phase 3 may well have had an indirect impact on the outcomes observed, or at the least on the practicability of introducing the specific interventions into the service.

Additional Findings

Within-Therapist Variability Reduction in Phase 3 in Both Samples. In addition to the indications in both Core Therapists and All Therapists groups that the therapist effect (or what could be called the between-therapist effect, as it describes between-therapist variability) reduced, the within-therapist effect also reduced in Phase 3. That is, the confidence interval range for each therapist appeared to reduce in Phase 3 (compared to both previous Phases), suggesting that the outcomes of the patients treated by each individual therapist were more similar to each other than in previous Phases. Or, expressed in another way, the therapists appeared to achieve more consistent outcomes across their individual patient caseload in Phase 3. Viewed as a potential side effect of the DP-lite aspect of the Collaborative Intervention, this may be consistent with the conclusions of Miller et al. (2020) who described DP as a "*within-person effect*" (p. 11). This might be extended to hypothesise that not only will therapists perform better the more they practice, but that they will perform more consistently the more they practice. With the impact and use of DP in psychotherapy still in its infancy, this may be a hypothesis that provides a focus for future DP

studies, that have so far measured improvements in overall patient mean outcomes rather than the consistency of outcomes by therapists.

Strengths and Limitations of Study Design

As previously mentioned, many of the strengths of the design of the current study also limit the certainty with which conclusions can be drawn. It is by its very nature that practice based research seeks to understand psychotherapy in practice, granting it high external validity, but at the cost of specificity of causality for the results it generates. Nevertheless, the current study provides a unique opportunity to understand the impact of a series of service improvements, in a longitudinal design, with a primary researcher providing the perspective of being embedded within the service leadership. Measured against the recently refined characteristics of higher-quality practice-oriented studies outlined by Castonguay et al. (2021), the study aligns with the three key features of: clinical helpfulness, feasibility, and methodological and statistical sophistication, as follows:

1) Clinical Helpfulness - this study provides evidence for ways of improving the impact of therapy and information to better understand patient and therapist characteristics, as well as contextual variables impacting psychotherapy.

2) Feasibility – this study meets all four of the feasibility criteria as follows; it did not impose drastic changes in routine clinical practice; minimal burden was placed on therapists in terms of time or additional tasks; research tasks (Collaborative Intervention package) were easily integrated into practice; and research tasks were able to be retained following the completion of the research.

3) Methodological and statistical sophistication – of the eight core areas under this feature, the present study meets five: repeated assessments of outcome; use of statistical analysis accounting for the nesting of data at multiple levels; use of within and between patient/therapist analyses; investigation of moderator variables; and investigation of mediator variables. Although overall and compared with much of the research in this area, the therapist and patient sample was large, given the refinement and use of the samples over the time period of the

three Phases, each Phase sample needed to be larger, particularly the therapist

sample, to be able to reliably compare the therapist effects with statistical

confidence. The final core criteria that this study did not meet was that it did not

include a large number of sites, as it was a single-site study.

There were some specific aspects of the current study design that provided both strengths,

and limitations to the overall thesis. These are presented and discussed in the following

sections.

Methodological Limitations

Therapist effects are derived from two variances; therefore, it was not possible to

consider the statistical significance of different therapist effects across Phases. However, as

a means of comparing effects with some measure of their reliability, MCMC methods are

recommended and were used (Browne, 2015). This allows an assessment of whether the

effects were reliably different. A further limitation is that the sample size of therapists did not

meet the recommended levels for the most reliable estimates of effects and the 95%

probability intervals derived from the MCMC analysis are likely to be wider than might be

expected with a larger sample.

The sample size limits any firm conclusions regarding the size of therapist effects

and, in particular, any reliable differences between effects across Phases. However, as part

of the proof of concept design, a range of comparisons were made between Phase models:

the size and significance of the therapist level variance and therapist effect in each Phase,

the improvements in model fit, the levelling-out of the caterpillar plots and the numbers of

therapists found to be non-average in each Phase. Collectively, and in conjunction with

clinical outcomes, these were used to assess whether service delivery had improved in

Phase 3. The results suggest that it had, but a further multisite randomised study with a

sample of over 100 therapists is required to make more reliable conclusions about whether

the effects had reduced and whether this was a result of the interventions.

Embedded Primary Researcher

The primary researcher was a senior member of the clinical leadership team throughout the study Phases. The study itself was initiated at a time of poor performance in the service, prior to a series of service changes. The primary researcher was committed to ensuring that any impact of the changes that might be implemented as a result of the poor service performance would be more rigorously investigated than would otherwise be the case in a service context. In so doing, the context of the primary researcher within the study design provided the possibility of the inclusion of the service leadership context that both generated and permitted the service changes at Phase 2 and Phase 3. Details of this context are rarely possible even within practice-based studies, meaning that the application of interventions or study conditions can be unclear, and any future replication therefore challenging. A lack of detail regarding the broader context of research setting limits both corroborating research endeavours but also the application of findings into real world contexts, as information about the supporting structure around the studies is either missing, or support is externally driven by an academic or research team (Castonguay et al., 2021).

By having the primary researcher situated within the leadership and management team, this allowed the Phase 3 intervention to be invested in and engaged with by management, supervisors and clinicians across an entire service. In the current study it is possible to understand both the specific interventions, and the nature and context of their application which, as argued above, may be a defining feature in generating the findings observed. The nature of the embedded researcher in the current study may be compared with a similar design feature in the Goldberg, Babins-Wagner et al. (2016) Calgary-based study. In a similar way, by having an embedded researcher within the service setting, the Calgary study was able to present not only details of policy changes implemented across the 7-year study period, but also the nature of how these policies were implemented and adapted over time. By providing this level of detail, it was possible to use information from the Calgary study to inform some of the changes made in the current study, such as the change in emphasis of the use of ROMs from performance target to feedback for learning

(discussed earlier), for example. Without an embedded researcher also involved in the day-to-day context of the service and research setting, this information would have been unlikely to have been available to this depth.

Nevertheless, the primary researcher being embedded in the service did also pose some challenges, both in practical terms and in terms of objectivity. The controversial nature of the changes and type of directive implementation style in Phase 2 meant that any association with these changes and the primary researcher may impact the ability to develop a collaborative package of interventions for Phase 3. This was overcome partly through the method of engaging staff (i.e., through focus groups and involvement from the University research team), but also through incorporating a therapist member of staff into the research site group as a collaborator, supporting the dissemination of DP ideas to the wider staff group (see Chapter 3 for details). A potential risk of a loss of objectivity is also inherent in a design with an embedded primary researcher working in the study setting itself. This was an aspect of the work that was often reflected on and discussed within the context of the primary researcher's supervision sessions.

The nature of the analysis of the primary datasets only being possible at the conclusion of all Phases (as it was not possible to determine the Core Therapist sample until the conclusion of the final Phase), meant that the design had an inherent control for any potential bias in the interpretation of analysis impacting the Intervention Phase. The additional therapist staff member in the on-site research team also provided an additional control for any potential lack of objectivity of the primary researcher. Their input was independent, primarily providing the additional voice of the therapy staff within research discussions on-site, with no particular attachment to a specific outcome. The use of a research diary was also a simple, yet useful way of monitoring potential biases, such as reviewing decision making processes and rationales retrospectively to identify any inconsistencies or loss of objectivity. It was therefore of benefit to both the study and the service improvements themselves that this work be carried out within the rigours of a PhD programme where appropriate levels of research governance were in place.

Longitudinal Study and Length of Phases

As presented above (and discussed in Chapter 6), a deliberate design feature of the study was to ensure that each Phase was of sufficient length to include both an implementation and stabilisation period. This aimed to allow the results of the changes to be discernible, if at all, within the same Phase in which the changes took place. However, the intervention period could not be too long as it is in the nature of a live service to constantly change and develop. So, if the Phase periods were too long this would weaken the link between the specific changes and any observable impact in the service data. An 18-month period was therefore agreed, which was to some extent mapped on to the implementation and stabilisation of changes in Phase 2, as these changes were largely externally driven. With the Phase 3 package it was possible to moderate the timeframe more deliberately to match the needs of the research design and the service. Despite this, given that the DP-lite peer groups were monthly, and given the phased, stepped wedge approach to implementation, this allowed only 6 months of DP-lite practice for the final therapist cluster, potentially reducing the impact of DP-lite within the study. It also meant that the number of patients in each stage of the stepped wedge implementation was small, reducing the reliability of any findings when comparing clusters. It would therefore have been preferable to increase the time in which therapists were in each stage of the DP-lite implementation, and the period after which all therapists were in the DP-lite condition. However, in increasing the Phase length in this way it would have risked the impact of further subsequent service changes potentially confounding the results. On balance and in hindsight, extending each Phase to a 2-year time period would have improved the ability to detect any effects more specifically related to the DP-lite aspect of the Collaborative Intervention. The Calgary study (Goldberg, Babins-Wagner et al., 2016) spanned a 7-year period, however may not have been subject to the external interventions and rapid changes experienced in English NHS settings, such as in the current study.

Clinic Effects and Other Therapist Variables

The findings from Firth et al. (2019), discussed in Appendix A, poses questions for the current study findings and design. Firth et al. reported a significant clinic effect in addition to the therapist effect. The therapist effect was generally consistent despite the addition of other variables, whereas the clinic effect reduced significantly when patient variables were added to the model. Despite this reduction, the clinic effect remained significant. That is, there were factors contributing to patient outcome at the clinic level that remained unexplained by any of the variables measured. It was not within the scope of this study to measure service effects, which would have involved multiple service sites, therefore it was not possible to disentangle any therapist effect from service effects. However, the importance of the answer to this question may be less significant in a practice context. Instead, knowing that we can intervene at the service level and have an impact across the therapist level may be sufficient in terms of improving outcome consistency for patients.

As mentioned earlier in this chapter, the lack of information about therapists themselves, such as levels of personal resilience, meant that these variables could not be factored into the study analysis, which was a limitation of the study. The decision to exclude additional therapist variables was primarily based on the perceived burden of requiring therapists to complete additional measures, in the context of a service with significant performance challenges and low staff morale. However, having additional measures of key therapist variables from the existing literature, such as resilience and mindfulness (Green et al., 2014; Pereira et al., 2017), burnout (Delgadillo, Saxon et al., 2018; McCarthy & Freize, 1999; Saxon & Barkham, 2012), professional self-doubt (Nissen-Lie, Monsen et al., 2013; Nissen-Lie et al., 2017; Odyniec et al., 2017) or personality traits (Delgadillo et al., 2020) would have further informed the impact of the two intervention Phases, or identified other variables impacting the therapist effect.

Implementation of DP-Lite

Introducing the concepts and practice of DP can be a challenge to therapists who may be used to more familiar ways of reflecting on clinical work and used to more traditional

supervision tasks. This cannot be underestimated, and more time spent on introducing these concepts would have been valuable (i.e., a half or full-day session for the introductory sessions, rather than just 2 hours) within the current study. With the benefit of hindsight, a clearer emphasis on what DP may look like in practice, with examples of its use should have been included as a main part of the introductory session, based on feedback from the early implementers. Clearer guidelines on how to run a session (as outlined in the 'hints and tips' guide, see Appendix G) should have been provided earlier/at roll out. When introducing ROMs and DP into their research setting, Goldberg, Babins-Wagner et al. (2016) found that, prior to the use being mandatory, 60% of therapists did not use them (though this was primarily in reference to the use of ROMs). Therefore, making use of DP mandatory within the groups was considered within the research group during implementation. However, on balance it was felt that this would conflict with the collaborative approach being taken during this Phase.

Whilst balancing the impact of 'over-measuring' therapists given the importance of the collaborative nature of the Phase 3 intervention, including a way of recording how much therapists engaged in DP both within the peer groups and in solitary practice would have helped to validate the overall success of the implementation of DP-lite. This may have added to evidence reported by Chow et al. (2015) that found it was the amount of time engaged in solitary activities to improve practice that was the predictor of outcomes. The current study focussed on peer group practice and did not measure whether this resulted in an increase in therapists solitary DP. However, it is possible that it might have done so simply due to exposure to the ideas, or possibly the reverse. There may have been a belief that therapists had 'already done' their practice in the group so didn't need to do anything additional to improve their therapy practice.

As discussed previously, both Brattland et al. (2019) and Delgadillo, de Jong et al. (2018) report the nature of how feedback is used as being a significant feature and both studies involved elements of learning and skills practice as part of the feedback process. The idea that it is *how* the feedback is used, rather than simply receiving it that makes the

impact, supports the focus within the Phase 3 intervention on individual skills development based on supervisor feedback and therapists' own sense of their areas for development. However, given the evidence that therapists are not the best assessors of their own deficits or strengths (Chui & Hill, 2020; Nissen-Lie et al., 2017; Ziem & Hoyer, 2020), there is an argument that the current study may have benefitted from the use of more objective feedback that could shed more accurate light on a therapist's individual areas for learning. Potentially this may have led to an improvement in clinical outcomes between Phase 2 and 3, in addition to the flattening of variability that was indicated. Miller et al. (2015) argued that limitations in the implementation and rewards envisaged from the use of ROMs in psychotherapy are due to the fact that use of ROMs constitutes only the first two steps of the recipe for expertise development – that is, determining a baseline level of effectiveness, and obtaining systematic ongoing feedback on practice. The third step requires engaging in DP (Miller et al., 2013). It could be argued that the current study focused on the third step (albeit in a 'lite' way) but did not have the robust system for feedback that might be more impactful. In the Core Therapist sample, the observed improvement in clinical outcomes (reliable change rate) following the introduction of feedback from ROMs, and the subsequent reduction in variability following the introduction of DP, may suggest that these two elements are targeting distinct though interrelated aspects of therapists clinical outcomes. Future research may be warranted in disentangling the potential hypothesis that DP in groups within a service reduces between and within therapist variability, but that the tracking of off-track cases through ROMs is necessary to improve outcomes overall (Delgadillo, de Jong et al., 2018).

Implications for Therapists and Services

In setting out to answer whether therapist variability can be reduced and outcomes improved across a service setting, the results have provided an indication, or proof of concept, that it may be possible to reduce therapist variability, whilst having no negative impact on clinical outcomes or possibly improving clinical outcomes. The interventions involved in the Collaborative Intervention Phase where the impact on therapist variability was

observed were largely delivered at a service level, however there are potential implications
for the individual practice of therapists. Although causality is not assumed, given the
observations made in this study alongside the other literature discussed, it could be
suggested that there may be a number of activities that a therapist may want to pursue in
order to bring their practice in line with others. These activities are: learning from peers
through deliberate skills practice (Firth et al., 2020; Goldberg, Babins-Wagner et al., 2016;
Miller et al., 2013); reflecting on one's own practice to identify development areas (Chow et
al., 2015; Goldberg, Babins-Wagner et al., 2016); seeking feedback from others i.e.
supervisors, peers, and patients through the use of ROMs, to identify potential development
areas (Brattland et al., 2018, 2019; Delgadillo, de Jong et al., 2018; Goldberg, Babins-
Wagner et al., 2016), In addition to these, taking action to sustain individual resilience and
wellbeing (Delgadillo, Saxon et al., 2018; Green et al., 2014; Pereira et al., 2017) may be a
factor that therapists may want to cultivate. The focus on individual case supervision which
has traditionally been the key forum for professional learning and development for
psychological therapists, should be supplemented with small peer group opportunities to
learn and practice skills with other therapists, and should include the use of ROMs for the
purpose of feedback and learning. Whilst services and supervisors have a responsibility for
how this feedback is delivered, therapists also have a responsibility: to seek and learn from
such information in an open and curious way for the objective of continuous improvement of
the therapy they provide to their patients (Boswell et al., 2015; Goldberg, Rousmaniere et al.,
2016).

The implications for services have largely been covered in the previous sections.
However, it is important to clearly define the two interconnected elements of service design
and management that may be implicated by the results of the current study. Firstly, the
content of the service intervention or model, and secondly the way in which a service
intervention or model is implemented and sustained by service leaders and staff. In Phase 2,
where improved clinical outcomes but higher levels of therapist variability were observed, the
content of the service intervention was primarily focused on the achievement of service

targets, with a commitment to the consistent adherence to national IAPT guidelines (NHSE &

NCCMH, 2021) including supervision models and the use of ROMs. Phase 2 was delivered

in a directive, externally-driven style, with limited staff involvement in decision making or idea

generation. The implementation of ROMs was explicitly linked to service target monitoring,

and supervision and learning opportunities were limited to individual case supervision.

Clearly there is sufficient evidence to suggest that the IAPT model of service delivery

generates strong clinical outcomes (see Chapter 4 for full details). However, the results of

the current study suggest that either the content or method of implementation may not be

sufficient to reduce the therapist effect, and in some cases may increase therapist variability.

In Phase 3, the service model remained consistent with Phase 2, with a continued

adherence to IAPT guidance (NHSE & NCCMH, 2021). However, there were additional

changes made with a refinement to the supervision model and the introduction of small peer

skills practice groups, as well as a redesign of the service management structure and a

focus on staff wellbeing. In relation to the model of implementation and management of

these changes, in Phase 3, changes were guided by staff consultation and preferences, with

the involvement of staff in the implementation of changes that impacted on them. The

introduction of small skills-based groups was initially suggested by the staff during focus

groups, and the implementation of DP-lite was undertaken collaboratively, with staff

consulted on how the groups might run, and providing researchers and managers with

feedback on how they could be improved and developed. ROMs were explicitly linked to

individual learning, and whilst the link to performance targets was not completely removed,

there was an emphasis on the personal ownership of the therapists' data by the individual

therapists and supervisors themselves. Ideas for staff wellbeing events were also generated

by the staff group. The management restructure involved a commitment to compassionate,

conscious leadership (Dethmer et al., 2014; Kofman, 2002; West et al., 2014; West et al.,

2017) supported by internal and external training events. There was also the deliberate

inclusion of horizontal overlays to the structural design of the management hierarchy to

encourage collaboration between clinical and operational managers at multiple levels (see Chapter 5).

Given the indicative reduction in the therapist effect in Phase 3, compared to Phase 2 in both samples, and the maintenance of improved clinical outcomes, it could be recommended that services consider not just the focus on professional development and skills practice in the ways described, but also consider the management culture in which therapists work. If indeed this focus on collaborative working did account for some of the reduction in the therapist effect, it should be further support for the development of increased reflection and innovation in the realm of change management and health care leadership. Service managers leading services based on models that mandate the consistent delivery of evidence-based practice, such as the IAPT model, may want to further consider how they apply this approach in a way that incorporates, rather than alienates, the ideas and learning of the therapists that work within it. In a parallel way that evidence-based practice and practice-based evidence need to be used to complement each other to provide benefits that are greater than the sum of their parts, so do service managers, healthcare leaders, and therapists. The current study presents the potential benefits of combining the collective intelligence of therapists, managers and researchers, in the application of health policy enhanced by other research literature in such a way that ultimately improves the consistency of beneficial clinical outcomes for patients.

Summary of Areas for Future Research

The findings from the current study highlight the importance of continued investigation into therapist effects, and the impact that service level and therapist level interventions can have on this significant factor influencing patient outcomes (Wampold & Owen, 2021). Few studies have investigated the therapist effect over time (Goldberg, Rousmaniere et al., 2016), and fewer studies have examined the moderating factors that might be used to limit the therapist effect in practice (Wampold & Owen, 2021). Given the findings presented in this thesis, that it may be possible to reduce the therapist effect, further studies focussing on the aspects of the current study's Phase 3 intervention should be

undertaken. These should include larger therapist cohorts (across multiple sites) to establish if the improvement in therapist effect is replicated and to be able to test the change for reliable differences using the methods presented in this thesis. This may also allow a stronger causal link between factors that may be directly responsible for any reduction, if indeed it is not a multi-factorial phenomenon. The growing use of DP with psychological therapy should be further researched, particularly given the potential for 'lite' versions that still may have positive effects, but with less burden on services and therapists. The hypothesis that a combined, sensitive use of ROMs and DP may target both therapist variability and an improvement in outcomes overall (Brattland et al., 2018, 2019; Delgadillo, de Jong et al., 2018; Goldberg, Babins-Wagner et al., 2016; Goodyear et al., 2017), should be further explored. This would include further testing of the hypothesis that DP may function at a within-therapist level (Miller et al., 2013) - that is, therapists' outcomes within their own caseload become more consistent. The use of peer groups to deliberately practice skills and obtain immediate feedback may be the factor that establishes the between-therapist level of variability, and the current findings would support the continuation of research into this position.

As an example of practice-based evidence, this research met the three defining characteristics; using data collected in routinely clinical practice, measuring aspects of everyday practice, and involving the collaborative development of specific practices (in this case the Collaborative Intervention package) rather than research-imposed restrictions (Castonguay et al., 2021). This, combined with the embedded nature of the primary researcher and the involvement of clinicians within the development of the research Phase, resulted in members of the service at multiple levels being empowered to own and use their practice data. Critiques of the 'top-down' approach to evidence-based practice, where therapists and services are provided with information about the most effective way of practicing based on a narrow range of scientific studies (in the UK, these are generally RCTs), look to the increased use and application of practice-based evidence in both practice and policy (Barkham & Margison, 2007; Castonguay et al., 2021). Indeed, the research

described in this thesis, involving a service-wide approach to learning and the more

sophisticated use of data analysis in a service than would be standard practice, provides an

example of how services and individual clinicians can be empowered to understand the

evidence that they produce within their own day-to-day work. This can thereby contribute to

the amalgamation of both PBE and evidence-based practice in a way that expands rather

than narrows our understanding of the interplay between patient, therapist, and therapy in

routine practice (Castonguay et al., 2021).

References

Ali, S., Littlewood, E., McMillan, D., Delgadillo, J., Miranda, A., Croudace, T., & Gilbody, S. (2014). Heterogeneity in patient-reported outcomes following Low-Intensity mental health interventions: A multilevel analysis. *PLoS ONE,* 9(9), e99658. https://doi.org/10.1371/journal.pone.0099658

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®).* American Psychiatric Pub.

Andrews, J. D. W. (1990). Interpersonal self-confirmation and challenge in psychotherapy. *Psychotherapy: Theory, Research, Practice, Training, 27*(4), 485–504. https://doi.org/10.1037/0033-3204.27.4.485

Antonuccio, D. O., Davis, C., Lewinsohn, P. M., & Breckenridge, J. S. (1987). Therapist variables related to cohesiveness in a group treatment for depression. *Small Group Behaviour, 18,* 557-564. https://doi.org/10.1177/104649648701800409

Atkins, P. W., & Parker, S. K. (2012). Understanding individual compassion in organizations: The role of appraisals and psychological flexibility. *Academy of Management Review, 37*(4), 524-546. https://doi.org/10.5465/amr.2010.0490

Avolio, B. J., Walumbwa, F. O., & Weber, T. J. (2009). Leadership: Current theories, research, and future directions. *Annual Review of Psychology*, *60*, 421-449. https://doi.org/10.1146/annurev.psych.60.110707.163621

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, pp. 258-297).* Wiley.

Barber, J. P., & Muenz, L. R. (1996). The role of avoidance and obsessiveness in matching patients to cognitive and interpersonal psychotherapy: Empirical findings from the

Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 64*(5), 951–958. https://doi.org/10.1037/0022-006X.64.5.951

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

Barkham, M., & Mellor-Clark, J. (2003). Bridging evidence-based practice and practice-based evidence: Developing a rigorous and relevant knowledge for the psychological therapies. *Clinical Psychology & Psychotherapy, 10*(6), 319-327. https://doi.org/10.1002/cpp.379

Barkham, M., Lutz, W., & Castonguay, L. G. (Eds.). (2021). *Bergin and Garfield's handbook of psychotherapy and behavior change*. Wiley.

Barkham, M., Saxon, D., Hardy, G. E., Bradburn, M., Galloway, D., Wickramasekera, N., Keetharuth, A. D., Bower, P., King, M., Elliott, R., Gabriel, L., Kellett, S., Shaw, S., Wilkinson, T., Connell, J., Harrison, P., Ardern, K., Bishop-Edwards, L., Ashley, K… & Brazier, J. E. (2021). Person-centred experiential therapy versus cognitive behavioural therapy delivered in the English Improving Access to Psychological Therapies service for the treatment of moderate or severe depression (PRaCTICED): a pragmatic, randomised, non-inferiority trial. *The Lancet Psychiatry, 8*(6), 487–499. https://doi.org/10.1016/S2215-0366(21)00083-3

Barkham, M., Shapiro, D. A., Hardy, G. E., & Rees, A. (1999). Psychotherapy in two-plus-one sessions: Outcomes of a randomized controlled trial of cognitive–behavioural and psychodynamic–interpersonal therapy for subsyndromal depression. *Journal of Consulting and Clinical Psychology*, *67*(2), 201-211. https://doi.org/10.1037/0022-006X.67.2.201

Barrett, M. S., & Berman, J. (2001). Is psychotherapy more effective when therapists disclose information about themselves? *Journal of Consulting and Clinical Psychology, 69,* 597–603. https://doi.org/10.1037/0022-006X.69.4.597

Basset, T., Faulkner, A., Repper, J., & Stamou, E. (2010). *Lived experience leading the way: Peer support in mental health.* London: Together UK. https://www.slamrecoverycollege.co.uk/uploads/2/6/5/2/26525995/lived_experience_peer_support_in_mental_health.pdf

Beck, A., Burdett, M., & Lewis, H. (2015). The association between waiting for psychological therapy and therapy outcomes as measured by the CORE-OM. *British Journal of Clinical Psychology*, *54*(2), 233-248. https://doi.org/10.1111/bjc.12072

Bennett-Levy, J., Richards, D. A., Farrand, P., Christensen, H., Griffiths, K. M., Kavanaugh, D. J., Klein, B., Lau, M. A., Proudfoot, J., Ritterband, L., White, J., & Williams, C. (Eds.). (2010). *Oxford guide to low intensity CBT interventions.* Oxford University Press. https://doi.org/10.1093/med/9780199590117.001.0001

Beutler, L. E., Clarkin, J. F., & Bongar, B. (2000). *Guidelines for the systematic treatment of the depressed patient.* New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195105308.001.0001

Beutler, L. E., Kim, E. J., Davison, E., Karno, M., & Fisher, D. (1996). Research contributions to improving managed health care outcomes. *Psychotherapy: Theory, Research, Practice, Training, 33*(2), 197–206. https://doi.org/10.1037/0033-3204.33.2.197

Beutler, L. E., Machado, P. P. P., & Neufeldt, S. (1994). Therapist variables. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (4th Edition, 259–269). New York: Wiley.

Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., & Wong, E.,
(2003). Therapist variables. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook
of psychotherapy and behavior change* (5th Edition, 227–306). Hoboken, N.J:. Wiley.

Beyebach, M., & Carranza, V. E. (1997). Therapeutic interaction and dropout: Measuring
relational communication in solution-focused therapy. *Journal of Family Therapy, 19,*
173–212. https://doi.org/10.1111/1467-6427.00047

Bird, A. (2006). *We need to talk: the case for psychological therapy on the NHS*. Mental
Health Foundation.

Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine
outcome monitoring in clinical practice: Benefits, challenges, and solutions.
Psychotherapy Research, 25, 6–19.
https://dx.doi.org/10.1080/10503307.2013.817696

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: access,
effectiveness and efficiency: narrative literature review. *The British Journal of
Psychiatry*, *186*(1), 11–17. https://doi.org/10.1192/bjp.186.1.11

Brattland, H., Koksvik, J. J., Burkeland, O., Gråwe, R. W., Klöckner, C., Linaker, O. M.,
Ryum, T., Wampold, B., Lara-Cabrera, M. L., & Iversen, V. C. (2018). The effects of
routine outcome monitoring (ROM) on therapy outcomes in the course of an
implementation process: A randomized clinical trial. *Journal of Counseling
Psychology, 65*(5), 641–652. http://dx.doi.org/10.1037/cou0000286

Brattland, H., Koksvik, J. J., Burkeland, O., Klöckner, C., Lara-Cabrera, M. L., Miller, S. D.,
Wampold, B., Ryum, T., & Iversen, V. C. (2019). Does the working alliance mediate
the effect of routine outcome monitoring (ROM) and alliance feedback on
psychotherapy outcomes? A secondary analysis from a randomized clinical trial.
Journal of Counseling Psychology, 66(2), 234–246.
https://doi.org/10.1037/cou0000320

Browne, W. J. (2015). *MCMC estimation in MLwiN, v2. 33.* Centre for Multilevel Modelling University of Bristol.

Browne, W. J., & Rasbash, J. (2009). *MCMC estimation in MLwiN*. Bristol: Centre of Multilevel Modelling, University of Bristol.

Carmeli, A., & Schaubroeck, J. (2006). Top management team behavioral integration, decision quality, and organizational decline. *The Leadership Quarterly*, *17*(5), 441-453. https://doi.org/10.1016/j.leaqua.2006.06.001

Castonguay, L. G., Barkham, M., Lutz, W., & McAleavey, A. A. (2013). Practice-oriented research: Approaches and applications. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 85-133). John Wiley & Sons.

Castonguay, L. G., Barkham, M., Youn, S. J., & Page, A. C. (2021). Practice-based evidence - Findings from routine clinical settings. In M. Barkham, W. Lutz & L.G. Castonguay (Eds.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (7th Ed., pp. 191–222). Wiley.

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*(3), 337-345. https://doi.org/10.1037/pst0000015

Chui, H., & Hill, C. E. (2020). Keep calm and alert and carry on: therapist calmness and fatigue in relation to session process. *Counselling Psychology Quarterly, 33*(1), 66-78. https://doi.org/10.1080/09515070.2018.1502160

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. *International Review of Psychiatry*, *23*(4), 318-327. https://doi.org/10.3109/09540261.2011.606803

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

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*(11), 910-920. https://doi.org/10.1016/j.brat.2009.07.010

Clements-Hickman, A. L., & Reese, R. J. (2020). Improving therapists' effectiveness: Can deliberate practice help? *Professional Psychology: Research and Practice, 51*(6), 606–612. https://doi.org/10.1037/pro0000318

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*(5), 614-625. https://doi.org/10.1007/s10488-017-0790-5

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*(2), 178-181. https://doi.org/10.1080/10503300500265025

Crits-Christoph, P., & Gibbons, M. B. C. (2021) Psychotherapy process-outcome research: Advances in understanding causal connections. In M. Barkham, W. Lutz & L.G. Castonguay (Eds.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (7th Ed., pp.263-296). Wiley.

Crits-Christoph, P., & Mintz, J. (1991). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *Journal of Consulting and Clinical Psychology, 59*(1), 20-26. https://doi.org/10.1037/0022-006X.59.1.20

Darzi, L. (2018). *The Lord Darzi review of health and care.* London: Institute for Public Policy Research.

Delgadillo, J., Branson, A., Kellett, S., Myles-Hooton, P., Hardy, G. E., & Shafran, R. (2020). Therapist personality traits as predictors of psychological treatment outcomes. *Psychotherapy Research, 30*(7)*,* 857-870. https://doi.org/10.1080/10503307.2020.1731927

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

Delgadillo, J., Kellett, S., Ali, S., McMillan, D., Barkham, M., Saxon, D., Donohoe,G., Stonebank, H., Mullaney, S., Eschoe, P., Thwaites, R., & Lucock, M. (2016). A multi-service practice research network study of large group psychoeducational cognitive behavioural therapy. *Behaviour Research and Therapy*, *87*, 155-161. https://doi.org/10.1016/j.brat.2016.09.010

Delgadillo, J., Saxon, D., & Barkham, M. (2018). Associations between therapists' occupational burnout and their patients' depression and anxiety treatment outcomes. *Depression and Anxiety*, *35*(9), 844-850. https://doi.org/10.1002/da.22766

Dennhag, I., Gibbons, M. B. C., Barber, J. P., Gallop, R., & Crits-Christoph, P. (2012). Do supervisors and independent judges agree on evaluations of therapist adherence and competence in the treatment of cocaine dependence? *Psychotherapy Research*, *22*(6), 720-730. https://doi.org/10.1080/10503307.2012.716528

Department of Health. (2008a). *Improving access to psychological therapies implementation plan: national guidelines for regional delivery.* HMSO.

https://webarchive.nationalarchives.gov.uk/ukgwa/20100604110854/http://www.dh.go

v.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_0

83150

Department of Health. (2008b). *Improving Access to Psychological Therapies (IAPT)*

outcomes toolkit 2008/9. HMSO. http://ipnosis.postle.net/PDFS/iapt-outcomes-toolkit-

2008-november(2).pdf

Department of Health. (2008c). *High quality care for all.* HMSO.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachm

ent_data/file/228836/7432.pdf

Department of Health. (2011a). *The IAPT data handbook.* (Version 2). HMSO.

https://webarchive.nationalarchives.gov.uk/ukgwa/20160302160058/http:/www.iapt.n

hs.uk/silo/files/iapt-data-handbook-v2.pdf

Department of Health. (2011b). *No health without mental health: a cross-government mental*

health outcomes strategy for people of all ages. HMSO.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachm

ent_data/file/213761/dh_124058.pdf

Department of Health. (2012a). *IAPT Key Performance Indicator (KPI) Technical Guidance*

for Adult IAPT Services 2012/13 v.2. HMSO.

https://webarchive.nationalarchives.gov.uk/ukgwa/20160302155857mp_/http://www.i

apt.nhs.uk/silo/files/iapt-kpi-technical-guidance-201213-v20-.pdf

Department of Health. (2012b). *IAPT three year report – The first million patients.* HMSO.

https://webarchive.nationalarchives.gov.uk/ukgwa/20140505102619mp_/http://www.i

apt.nhs.uk/silo/files/iapt-3-year-report.pdf

Department of Health and University College London. (2015). *National curriculum for the education of psychological wellbeing practitioners*. (3rd edition). University College London. https://www.ucl.ac.uk/pals/sites/pals/files/pwp_review_-_final_report.pdf

Dethmer, J., Chapman, D., & Klemp, K. W. (2014). *The 15 commitments of conscious leadership: A new paradigm for sustainable success*. Conscious Leadership Group.

Elkin, I., Falconnier, L., Martinovich, Z., & Mahoney, C. (2006). Therapist effects in the national institute of mental health treatment of depression collaborative research program. *Psychotherapy Research, 16,* 144–160. https://doi.org/10.1080/10503300500268540

Erekson, D. M., Janis, R., Bailey, R. J., Cattani, K., & Pedersen, T. R. (2017). A longitudinal investigation of the impact of psychotherapist training: Does training improve client outcomes? *Journal of Counseling Psychology, 64*(5), 514–524. https://doi.org/10.1037/cou0000252

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–703). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.038

Ericsson, K. A., & Harwell, K. W. (2019). Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: Why the original definition matters and recommendations for future research. *Frontiers in Psychology*, *10*, 2396. https://doi.org/10.3389/fpsyg.2019.02396

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363–406. https://doi.org/10.1037/0033-295X.100.3.363

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*(1), 273–305. https://doi.org/10.1146/annurev.psych.47.1.273

Ericsson, K. A., & Pool, R. (2016). *Peak: Secrets from the new science of expertise*. Random House.

Errázuriz, P., & Zilcha-Mano, S. (2018). In psychotherapy with severe patients discouraging news may be worse than no news: The impact of providing feedback to therapists on psychotherapy outcome, session attendance, and the alliance. *Journal of Consulting and Clinical Psychology, 86*(2), 125–139. https://doi.org/10.1037/ccp0000277

Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence based Mental Health, 1*(3), 70-72. https://doi.org/10.1136/ebmh.1.3.70

Firth, N., Barkham, M., Kellett, S., & Saxon, D. (2015). Therapist effects and moderators of effectiveness and efficiency in psychological wellbeing practitioners: A multilevel modelling analysis. *Behavioural Research and Therapy, 69,* 54-62. https://doi.org/10.1016/j.brat.2015.04.001

Firth, N., Saxon, D., Stiles, W., & Barkham, M. (2019). Therapist and clinic effects in psychotherapy: A three-level model of outcome variability. *Journal of Consulting and Clinical Psychology, 87*(4), 345–356. https://doi.org/10.1037/ccp0000388

Firth, N., Saxon, D., Stiles, W., & Barkham, M. (2020). Therapist effects vary significantly across psychological treatment care sectors. *Clinical Psychology and Psychotherapy, 27,* 770-778. https://doi.org/10.1002/cpp.2461

Friedrich, T. L., Vessey, W. B., Schuelke, M. J., Ruark, G. A., & Mumford, M. D. (2009). A framework for understanding collective leadership: The selective utilization of leader

and team expertise within networks. *The Leadership Quarterly*, *20*(6), 933-958. https://doi.org/10.1016/j.leaqua.2009.09.008

Gambia Hepatitis Study Group. (1987) The Gambia hepatitis intervention study. *Cancer Research, 47,* 5782-5787.

Gillard, S. (2019). Peer support in mental health services: where is the research taking us, and do we want to go there? *Journal of Mental Health, 28*(4), 341-344. https://doi.org/10.1080/09638237.2019.1608935

Goldberg, S. B., Babins-Wagner, R., Rousmaniere, T., Berzins, S., Hoyt, W. T., Whipple, J. L., Miller, S. D., & Wampold, B. E. (2016). Creating a climate for therapist improvement: A case study of an agency focused on outcomes and deliberate practice. *Psychotherapy, 53,* 367–375. https://doi.org/10.1037/pst0000060

Goldberg, S. B., Hoyt, W. T., Nissen-Lie, H. A., Nielsen, S. L., & Wampold, B. E. (2016). Unpacking the therapist effect: Impact of treatment length differs for high-and low-performing therapists. *Psychotherapy Research*. https://doi.org/10.1080/10503307.2016.1216625

Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., & Wampold, B. E. (2016). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology, 63*(1), 1-11. https://doi.org/10.1037/cou0000131

Goldberg, S. B., Thompson, M. N., & Nielsen, S. L. (2018). Patient financial distress and treatment outcomes in naturalistic psychotherapy. *Journal of Counseling Psychology, 65*(4), 523-530. https://doi.org/10.1037/cou0000264

Goldstein, H., & Healy, M. J. R. (1995). The graphical presentation of a collection of means. *Journal of the Royal Statistical Society, 158*, 175–177. https://doi.org/10.2307/2983411

Goldstein, H., & Spiegelhalter, D. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society, 159*, 385–443. https://doi.org/10.2307/2983325

Goodyear, R. K., Wampold, B. E., Tracey, T. J., & Lichtenberg, J. W. (2017). Psychotherapy expertise should mean superior outcomes and demonstrable improvement over time. *The Counseling Psychologist*, *45*(1), 54-65. https://doi.org/10.1177/0011000016652691.

Green, H., Barkham, M., Kellett, S., & Saxon, D. (2014). Therapist effects and IAPT psychological wellbeing practitioners (PWPs): a multilevel modelling and mixed methods analysis. *Behaviour Research and Therapy, 63,* 43-54. https://doi.org/10.1016/j.brat.2014.08.009

Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: lessons from year one of IAPT. *Behaviour Research and Therapy*, *51*(9), 597-606. https://doi.org/10.1016/j.brat.2013.06.004

Ham, C. (2014). *Reforming the NHS from within: beyond hierarchy, inspection and markets*. London: The King's Fund. https://www.kingsfund.org.uk/time-to-think-differently/publications/reforming-nhs-within

Ham, C., Baker, G. R., Docherty, J., Hockey, P., Lobley, K., Tugendhat, L., & Walshe, K. (2011). *The future of leadership and management in the NHS: no more heroes.* London: The King's Fund. https://www.kingsfund.org.uk/publications/future-leadership-and-management-nhs

Hawton, K. E., Salkovskis, P. M., Kirk, J. E., & Clark, D. M. (1989). *Cognitive behaviour therapy for psychiatric problems: A practical guide*. Oxford University Press.

Hayes, J. A., Owen, J., & Bieschke, K. J. (2015). Therapist differences in symptom change with racial/ethnic minority clients. *Psychotherapy, 52*(3), 308-314. https://doi.org/10.1037/a0037957

Heck, R. H., & Thomas, S. L. (2020). *An introduction to multilevel modelling techniques: MLM and SEM approaches*. Routledge.

Heinonen, E., & Nissen-Lie, H. A. (2020). The professional and personal characteristics of effective psychotherapists: a systematic review. *Psychotherapy Research, 30*(4), 417-432. https://doi.org/10.1080/10503307.2019.1620366

Hemming, K., Haines, T P., Chilton, P. J., Girling, A. J., & Lilford, R. J. (2015). The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *British Medical Journal, 350,* h391*.* https://doi.org/10.1136/bmj.h391

Henry, W. P., Schacht, T. E., & Strupp, H. H. (1990). Patient and therapist introject, interpersonal process, and differential psychotherapy outcome. *Journal of Consulting and Clinical Psychology, 58,* 768-774. https://doi.org/10.1037/0022-006X.58.6.768

Heppner, P. P., & Heesacker, M. (1983). Perceived counsellor characteristics, client expectations, and client satisfaction with counselling. *Journal of Counselling Psychology, 30,* 31-39. https://doi.org/10.1037/0022-0167.30.1.31

Higgins, J. P. T., Savović, J., Page, M. J., Elbers, R. G., & Sterne, J. A. C. (2011). Chapter 8: Assessing risk of bias in a randomized trial. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.). *Cochrane handbook for systematic reviews of interventions version 6.2* (updated February 2021). Cochrane.

Hill, C. E., Kivlighan, D. M. III, Rousmaniere, T., Kivlighan, D. M., Jr., Gerstenblith, J. A., & Hillman, J. W. (2020). Deliberate practice for the skill of immediacy: A multiple case

study of doctoral student therapists and clients. *Psychotherapy, 57*(4), 587–597. https://doi.org/10.1037/pst0000247

Hill, C. E., Spiegel, S. B., Hoffman, M. A., Kivlghan, D. M., & Gelso, C. J. (2017). Therapist expertise in psychotherapy revisited. *The Counseling Psychologist, 45*(1), 7-53. https://doi.org/10.1177/0011000016641192

Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioural therapy: A review of meta-analyses. *Cognitive Therapy and Research*, *36*(5), 427-440. https://doi.org/10.1007/s10608-012-9476-1

Hosch, H. M., Barrientos, G. A., Fierro, C., Ramirez, J. I., Pelaez, M. P., Cedillos, A. M., Meyer, L. D., & Perez, Y. (1995). Predicting adherence to medications by Hispanics with schizophrenia. *Hispanic Journal of Behavioural Sciences, 17,* 320-333. https://doi.org/10.1177/07399863950173003

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

Hunt, D. D., Carr, J. E., Dagadakis, C. S., & Walker, E. A. (1985). Cognitive match as a predictor of psychotherapy outcome. *Psychotherapy: Theory, Research, Practice, Training*, *22*(4), 718. https://doi.org/10.1037/h0085558

IBM Corp., (Released 2017). IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.

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*(1), 12-19. https://doi.org/10.1037/0022-006X.59.1.12

Johns, R. G., Barkham, M., Kellett, S., & Saxon, D. (2019). A systematic review of therapist effects: A critical narrative update and refinement to Baldwin & Imel's (2013) review. *Clinical Psychology Review, 67,* 78-93. https://doi.org/10.1016/j.cpr.2018.08.004

Kings Fund. (2014). *Future Organisational Models for the NHS: Perspectives for the Dalton Review.*

https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/future-organisational-models-for-the-nhs-kingsfund-jul14.pdf

Kofman, F. (2002). *Conscious business: How to build value through values.* Boulder: Sounds True

Kraus, D. R., Bentley, J. H., Alexander, P. C., Boswell, J. F., Constantino, M. J., Baxter, E. E., & Castonguay, L. G. (2016). Predicting therapist effectiveness from their own practice-based evidence. *Journal of Consulting and Clinical Psychology, 84*(6), 473-483. https://doi.org/10.1037/ccp0000083

Krippner, K. M., & Hutchinson, R. L. (1990). Effects of a brief intake interview on clients' anxiety and depression: Follow-up. *International Journal of Short-term Psychotherapy, 5,* 121-130.

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

Kubátová, J. (2018). The phenomenon of conscious business–which is ahead–practice or academia? A literature review. In *21st Annual International Conference Enterprise and Competitive Environment. Conference Proceedings, Mendel University, Brno* (pp. 361-369).

Lambert, M. J. (Ed.). (2003). *Bergin and Garfield's handbook of psychotherapy and behavior change.* (5th Edition) John Wiley & Sons.

Lambert, M. J. (Ed.). (2013). *Bergin and Garfield's handbook of psychotherapy and behavior change.* (6th Edition) John Wiley & Sons.

Laska, K. M., Smith, T. L., Wislocki, A. P., Minami, T., & Wampold, B. E. (2013). Uniformity of evidence-based treatments in practice? Therapist effects in the delivery of cognitive processing therapy for PTSD. *Journal of Consulting Psychology, 60*(1), 31-41. https://doi.org/10.1037/a0031294

Layard, R. (2005). *Mental health: Britain's biggest social problem?* Presented at No. 10 Strategy Unit Seminar for Mental Health. London. https://www.researchgate.net/publication/238492767_Mental_Health_Britain's_Biggest_Social_Problem

Layard, R., & Clark, D. M. (2014). *Thrive: How better mental health care transforms lives and saves money*. Princeton University Press.

Layard, R., Clark, D., Knapp, M., & Mayraz, G. (2007). Cost-benefit analysis of psychological therapy. *National Institute Economic Review, 202*(1), 90-98. https://doi.org/10.1177/0027950107086171

Linley, P., Govindji, R., & West, M. (2007). Positive psychology approaches to public services leadership: An introduction to strengths-based leadership. *International Journal of Leadership in Public Services, 3*(4), 44-55. https://doi.org/10.1108/17479886200700029

Lord, R. G., & Brown, D. J. (2001). Leadership, values, and subordinate self-concepts. *The Leadership Quarterly*, *12*(2), 133-152. https://doi.org/10.1016/S1048-9843(01)00072-8

Lutz, W., Leon, S. C., 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*(1), 32. https://doi.org/10.1037/0022-0167.54.1.32

Maas, C. J. M., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58,* 127-137. https://doi.org/10.1046/j.0039-0402.2003.00252.x

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86-92. https://doi.org/10.1027/1614-2241.1.3.86

Margison, F. R., Barkham, M., Evans, C., McGrath, G., Clark, J. M., Audin, K., & Connell, J. (2000). Measurement and psychotherapy: Evidence-based practice and practice-based evidence. *The British Journal of Psychiatry*, *177*(2), 123-130. https//doi.org/10.1192/bjp.177.2.123

Marks, I. M. (1986). *Behavioural psychotherapy: Maudsley pocket book of clinical management.* Wright/IOP Publishing.

McCarthy, W. C., & Frieze, I. H. (1999). Negative aspects of therapy: Client perceptions of therapists' social influence, burnout, and quality of care. *Journal of Social Issues, 55,* 33-50. https://doi.org/10.1111/0022-4537.00103

McNeil, B. W., May, R. J., & Lee, V. E. (1987). Perceptions of counsellor course characteristics by premature and successful terminators. *Journal of Counselling Psychology, 34,* 86-89. https://doi.org/10.1037/0022-0167.34.1.86

McPherson, S., Evans, C., & Richardson, P. (2009). The NICE depression guidelines and the recovery model: Is there an evidence base for IAPT? *Journal of Mental Health*, *18*(5), 405-414. https://doi.org/10.3109/09638230902968258

Miller, S. D., Chow, D., Wampold, B. E., Hubble, M. A., Del Re, A. C., Maeschalck, M., & Bargmann, S. (2020). To be or not to be (an expert)? Revisiting the role of deliberate practice in improving performance. *High Ability Studies, 31*(1), 5-15. https://doi.org/10.1080/13598139.2018.1519410

Miller, S. D., Hubble, M. A., Chow, D. L., & Seidel, J. A. (2013). The outcome of psychotherapy: Yesterday, today, and tomorrow. *Psychotherapy, 50*(1), 88-97. https://doi.org/10.1037/a0031097

Miller, S. D., Hubble, M. A., Chow, D. L., & Seidel, J. A. (2015). Beyond measures and monitoring: Realizing the potential of feedback-informed treatment. *Psychotherapy, 52*(4), 449-457. https://doi.org/10.1037/pst0000031

Miller, S. D., Hubble, M. A., & Duncan, B. L. (2007). Supershrinks: what's the secret of their success? *Psychotherapy Networker*. https://www.psychotherapynetworker.org/blog/details/393/supershrinks

Miller, S. D., Hubble, M. A., & Wampold, B.E. (2017). Growing better therapists: A new opportunity for mental health administrators. *Administration and Policy in Mental Health and Mental Health Services Research, 44,* 732-734. https://doi.org/10.1007/s10488-017-0805-2

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *The British Journal of Psychiatry, 180*(5), 461-464. https://doi.org/10.1192/bjp.180.5.461

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

National Institute for Health and Care Excellence. (2004a). *Anxiety: management of anxiety (panic disorder, with and without agoraphobia, and generalised anxiety disorder) in adults in primary, secondary and community care. Clinical guideline No. 22.* https://www.nice.org.uk/guidance/cg22

National Institute for Health and Care Excellence. (2004b). *Depression: management of depression in primary and secondary care. Clinical guideline No. 23.* https://www.nice.org.uk/guidance/cg23

National Institute for Health and Care Excellence. (2005a). *Post-traumatic stress disorder: management. Clinical guideline No. 26.* https://www.nice.org.uk/guidance/cg26

National Institute for Health and Care Excellence. (2005b). *Obsessive-compulsive disorder and body dysmorphic disorder: treatment. Clinical guideline No. 31.* https://www.nice.org.uk/guidance/cg31

National Institute for Health and Care Excellence. (2006). *Computerised cognitive behaviour therapy for depression and anxiety. Technology appraisal guidance No. 97.* https://www.nice.org.uk/guidance/ta97

National Institute for Health and Care Excellence. (2009a). *Depression in adults: recognition and management. Clinical guideline No. 90.* https://www.nice.org.uk/guidance/cg90

National Institute for Health and Care Excellence. (2009b). *Depression in adults with a chronic physical health problem: recognition and management. Clinical guideline No. 91.* https://www.nice.org.uk/guidance/cg91

National Institute for Health and Care Excellence. (2011). *Common mental health problems: identification and pathways to care. Clinical guideline No. 123.* https://www.nice.org.uk/guidance/cg123

NHS Digital. (2014). *Psychological therapies, annual report on the use of IAPT services 2012-13.* Leeds, UK: Health and Social Care Information Centre. https://files.digital.nhs.uk/publicationimport/pub13xxx/pub13339/psyc-ther-ann-rep-2012-13.pdf

NHS Digital. (2020). *Psychological therapies, annual report on the use of IAPT services 2019-20.* Leeds, UK: Health and Social Care Information Centre. https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-annual-reports-on-the-use-of-iapt-services/annual-report-2019-20

NHS Digital. (2021). *Interactive visualisation of monthly data for Improving Access to Psychological Therapies (IAPT).* NHS Digital. Mental health data hub.

https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/mental-health-data-hub/psychological-therapies

NHS England. (2016). *The five year forward view for mental health.* London: Rep., NHS England Mental Health Taskforce.

https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf

NHS England. (2019). *The NHS long term plan.* London: NHS England.

https://www.longtermplan.nhs.uk/publication/nhs-long-term-plan/

NHS England and National Collaborating Centre for Mental Health. (2018). *The IAPT Manual v.1.* https://www.england.nhs.uk/publication/the-improving-access-to-psychological-therapies-manual/

NHS England and National Collaborating Centre for Mental Health. (2021). *The IAPT Manual v.5.* HMSO. https://www.england.nhs.uk/wp-content/uploads/2018/06/the-iapt-manual-v5.pdf

NHS Improvement. (2017). *IAPT service review - internal report.* Authors own copy

Newman, M. (2000). Recommendations for a cost-offset model of psychotherapy allocation using generalized anxiety disorder as an example. *Journal of Consulting and Clinical Psychology*, *68,* 549–555. https://doi.org/10.1037/0022-006X.68.4.549

Nissen-Lie, H. A., Goldberg, S. B., Hoyt, W. T., Falkenstrom, F., Holmqvist, R., Nielsen, S. L., & Wampold, B. E. (2016). Are therapists uniformly effective across patient outcome domains? A study on therapist effectiveness in two different treatment contexts. *Journal of Counseling Psychology, 63*(4), 367-378. https://doi.org/10.1037/cou0000151

Nissen-Lie, H. A., Havik, O. E., Høglend, P. A., Monsen, J. T., & Rønnestad, M. H. (2013). The contribution of the quality of therapists personal lives to the development of the

working alliance. *Journal of Counseling Psychology, 60*(4), 483-495. https://doi.org/10.1037/a0033643

Nissen-Lie, H. A., Monsen, J. T., Ulleberg, P., & Ronnestad, M. H. (2013). Psychotherapists' self-reports of their interpersonal functioning and difficulties in practice as predictors of patient outcome. *Psychotherapy Research, 23*(1), 86-104*.* https://doi.org/10.1080/10503307.2012.735775

Nissen-Lie, H. A., Ronnestad, M. H., Hoglend, 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 and Psychotherapy, 24,* 48-60. https://doi.org/10.1002/cpp.1977

Norcross, J. C., & Karpiak, C. P. (2017). Our best selves: Defining and actualizing expertise in psychotherapy. *The Counseling Psychologist, 45*(1), 66-75. https://doi.org/10.1177/0011000016655603

Odyniec, P., Probst, T., Margraf, J., & Willutzki, U. (2017). Psychotherapist trainees' professional self-doubt and negative personal reaction: Changes during cognitive behavioral therapy and association with patient progress. *Psychotherapy Research, 29*(1), 123-138. https://doi.org/10.1080/10503307.2017.1315464

Office for National Statistics. (2019). UK Government

https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2019estimates#ageing

Okiishi, J., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An empirical analysis of therapist effects. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, *10*(6), 361-373. https://doi.org/10.1002/cpp.383

Owen, J., Adelson, J., Budge, S., Kopta, S. M., & Reese, R. J. (2016). Good-enough level and dose-effect models: Variation among outcomes and therapists. *Psychotherapy Research, 26*(1), 22-30. https://doi.org/10.1080/10503307.2014.966346

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. National Institute for Health Research, Service Delivery and Organisation programme.

Pearce, C. L., & Conger, J. A. (2003). *Shared leadership: Reframing the hows and whys of leadership*. Sage.

Pereira, J., Barkham, M., Kellett, S., & Saxon, D. (2017). The role of practitioner resilience and mindfulness in effective practice: A practice-based feasibility study. *Administration and Policy in Mental Health and Mental Health Services Research, 44,* 691-704. https://doi.org/10.1007/s10488-016-0747-0

Piper, W. E., Joyce, A. S., McCallum, M., & Azim, H. F. (1998). Interpretive and supportive forms of psychotherapy and patient personality variables. *Journal of Consulting and Clinical Psychology, 66*(3), 558–567. https://doi.org/10.1037/0022-006X.66.3.558

Piper, W. E., McCallum, M., Joyce, A. S., Azim, H. F., & Ogrodniczuk, J. S. (1999). Follow-up findings for interpretive and supportive forms of psychotherapy and patient personality variables. *Journal of Consulting and Clinical Psychology, 67*(2), 267–273. https://doi.org/10.1037/0022-006X.67.2.267

Proudfoot, J., Ryden, C., Everitt, B., Shapiro, D. A., Goldberg, D., Mann, A., Tylee, A., Marks, I., & Gray, J. A. (2004). Clinical efficacy of computerised cognitive–behavioural therapy for anxiety and depression in primary care: randomised

controlled trial. *British Journal of Psychiatry*, *185*, 46–54.
https://doi.org/10.1192/bjp.185.1.46

Psychotherapy Expert Talks. (2017, April 7). *Tony Rousmaniere on deliberate practice and developing psychotherapy expertise* [Video]. YouTube.
https://www.youtube.com/watch?v=eR5OQ_mWros

Public Health England. (2019). *Public Health Dashboard.*
https://healthierlives.phe.org.uk/topic/public-health-dashboard/

Pybis, J., Saxon, D., Hill, A., & Barkham, M. (2017). The comparative effectiveness and efficiency of cognitive behaviour therapy and generic counselling in the treatment of depression: evidence from the 2nd UK National Audit of psychological therapies. *BMC Psychiatry*, *17*(1), 1-13. https://doi.org/10.1186/s12888-017-1370-7

Rasbash, J., Steele, F., Browne, W., Goldstein, H., (2017). *A user's guide to MLwiN version 3.00.* Centre for Multilevel Modelling: University of Bristol.
https://bristol.ac.uk/cmm/media/software/mlwin/downloads/manuals/3-00/manual-print.pdf

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

Richards, D., Richards, A., Barkham, M., Cahill, J., & Williams, C. (2002). PHASE: a 'health technology' approach to psychological treatment in primary mental health care. *Primary Health Care Research and Development*, *3,* 159-168.
https://doi.org/10.1191/1463423602pc103oa

Richards, D. A., & Suckling, R. (2008). Improving access to psychological therapy: The Doncaster demonstration site organisational model. *Clinical Psychology Forum, 181*, 9-16. https://hdl.handle.net/10036/42082

Richards, D. A., & Suckling, R. (2009). Improving access to psychological therapies: Phase IV prospective cohort study. *British Journal of Clinical Psychology, 48*(4), 377-396. https://doi.org/10.1348/014466509X405178

Richards, D. A., & Whyte, M. (2011). *Reach out: National programme student materials to support the delivery of training for psychological wellbeing practitioners delivering low intensity interventions (3rd edition).* London: Rethink Mental Illness. http://cedar.exeter.ac.uk/media/universityofexeter/schoolofpsychology/cedar/documents/Reach_Out_3rd_edition.pdf

Ricker, M., Nystul, M., & Waldo, M. (1999). Counselors' and clients' ethnic similarity and therapeutic alliance in time-limited outcomes of counselling. *Psychological Reports, 84,* 674-676. https://doi.org/10.2466/pr0.1999.84.2.674

Ricks, D. F. (1974). Supershrink: Methods of a therapist judged successful on the basis of adult outcomes of adolescent patients. In D. F. Ricks, A. Thomas, & M. Roff (Eds.), *Life history research in psychopathology: III.* University of Minnesota Press.

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

Robinson-Hickman, G. (2010). *Leading organizations: Perspectives for a new era*. Sage.

Roth, A. (1996). The relationship between outcome and therapist training, experience, and technique. In P. Fonagy & A. Roth (Eds.). *What works for whom? A critical review of psychotherapy research.* (1st ed., pp. 341-357). Guilford Press.

Roth, A., & Fonagy, P. (2005). *What works for whom: A critical review of psychotherapy research* (2nd ed.). Guilford Publications.

Roth, A. D., & Pilling, S. (2008). Using an evidence-based methodology to identify the competences required to deliver effective cognitive and behavioural therapy for

depression and anxiety disorders. *Behavioural and Cognitive Psychotherapy*, *36*(2), 129-147. https://doi.org/10.1017/S1352465808004141

Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry*, *6*(3), 412. https://doi.org/10.1111/j.1939-0025.1936.tb05248.x

Rousmaniere, T. G., Swift, J. K., Babins-Wagner, R., Whipple, J. L., & Berzins, S. (2016). Supervisor variance in psychotherapy outcome in routine practice. *Psychotherapy Research, 26*(2), 196-205. https://doi.org/10.1080/10503307.2014.963730

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*(4), 535–546. https://doi.org/10.1037/a0028898

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*(3), 575-588. https://doi.org/10.1002/cpp.2028

Saxon, D., Firth, N., & Barkham, M. (2017). The relationship between therapist effects and therapy delivery factors: Therapy modality, dosage, and non-completion. *Administration and Policy in Mental Health and Mental Health Services Research, 44,* 705-715. https://doi.org/10.1007/s10488-016-0750-5

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 Research, 44,* 598-613. https://doi.org/10.1007/s10488-016-0736-3

Schulberg, H. C., Pilkonis, P. A., & Houck, P. (1998). The severity of major depression and

    choice of treatment in primary care practice. *Journal of Consulting and Clinical*

    Psychology, 66(6), 932–938. https://doi.org/10.1037/0022-006X.66.6.932

Shapiro, D. A., Firth-Cozens, J., & Stiles, W. B. (1989). The question of therapists'

    differential effectiveness: A Sheffield Psychotherapy Project addendum. *The British*

    Journal of Psychiatry, *154*(3), 383-385. https://doi.org/10.1192/bjp.154.3.383

Slone, N. C., & Owen, J. (2015). Therapist alliance activity, therapist comfort, and systemic

    alliance on individual psychotherapy outcome. *Journal of Psychotherapy Integration,*

    25(4), 275-288. https://doi.org/10.1037/a0039562

Snape, C., Perren, S., Jones, L., & Rowland, N. (2003). Counselling - why not? A qualitative

    study of people's accounts of not taking up counselling appointments. *Counselling*

    and Psychotherapy Research, *3*(3), 239-245.

    https://doi.org/10.1080/14733140312331384412

Snijders, T. A., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and*

    advanced multilevel modelling. Sage.

Snowden, L. R., Hu, T.-w., & Jerrell, J. M. (1995). Emergency care avoidance: Ethnic

    matching and participation in minority-serving programs. *Community Mental Health*

    Journal, 31, 463-473. https://doi.org/10.1007/BF02188616

Spitzer, R. L., Kroenke, K., & Williams, J. B. (1999). Validation and utility of a self-report

    version of PRIME-MD: the PHQ primary care study. *JAMA*, *282*(18), 1737-1744.

    https://doi.org/10.1001/jama.282.18.1737

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing

    generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*(10),

    1092-1097. https://doi.org/10.1001/archinte.166.10.1092

Sterling, R. C., Gottheil, E., Weinstein, S. P., & Serota, R. (1998). Therapist/patient race and sex matching: Treatment retention and 9-month follow-up outcomes. *Addiction, 93,* 1043-1050. https://doi.org/10.1046/j.1360-0443.1998.93710439.x

Stiles, W. B., Honos-Webb, L., & Surko, M. (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice, 5*(4), 439–458. https://doi.org/10.1111/j.1468-2850.1998.tb00166.x

Sue, S., Fujino, D. C., Hu, L.-t., Takeuchi, D. T., & Zane, N. W. S. (1991). Community mental health services for ethnic minority groups: A test of the cultural responsiveness hypothesis. *Journal of Consulting and Clinical Psychology, 59*(4), 533–540. https://doi.org/10.1037/0022-006X.59.4.533

Talmon, M. (1990). *Single session therapy: Maximising the effect of the first (and often only) therapeutic encounter.* San Francisco: Jossey-Bass.

Tracey, T. J. (1986). Interactional correlates of premature termination. *Journal of Consulting and Clinical Psychology, 54*(6), 784–788. https://doi.org/10.1037/0022-006X.54.6.784

Tracey, T. J. G., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in psychotherapy: An elusive goal? *American Psychologist, 69*(3), 218–229. https://doi.org/10.1037/a0035099

Waller, G. (2009). Evidence-based treatment and therapist drift. *Behaviour research and therapy, 47*(2), 119-127. https://doi.org/10.1016/j.brat.2008.10.018

Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods and findings.* Lawrence Erlbaum Associates.

Wampold, B. E., & Bolt, D. M. (2006). Therapist effects: Clever ways to make them (and everything else) disappear. *Psychotherapy Research, 16*(2), 184-187. https://doi.org/10.1080/10503300500265181

Wampold, B. E., & Bolt, D. M. (2007). The consequences of "anchoring" in longitudinal

multilevel models: Bias in the estimation of patient variability and therapist effects.

Psychotherapy Research, 17, 509– 14. https://doi.org/10.1080/10503300701250339

Wampold, B. E., & Owen, J. (2021). Therapist effects: history, methods, magnitude. In M.

Barkham, W. Lutz & L.G. Castonguay (Eds.), *Bergin and Garfield's handbook of*

psychotherapy & behavior change (7th ed., pp. 297-326). Wiley.

Ward, E., King, M., Lloyd, M., Bower, P., Sibbald, B., Farrelly, S.,Gabbay, M., Tarrier, N., &

Addington-Hall, J. (2000). Randomised controlled trial of non-directive counselling,

cognitive-behaviour therapy and usual GP care for patients with depression. I:

Clinical effectiveness. *BMJ*, 321, 1383–1388.

https://doi.org/10.1136/bmj.321.7273.1383

West, M. A. (2002). Sparkling fountains or stagnant ponds: An integrative model of creativity

and innovation implementation in work groups. *Applied Psychology: An International*

Review, 51, 355 – 387.

West, M., Borrill, C. S., Dawson, J. F., Brodbeck, F., Shapiro, D. A., & Haward, B. (2003).

Leadership clarity and team innovation in health care. *The Leadership Quarterly, 14,*

393–410. https://doi.org/10.1016/S1048-9843(03)00044-4

West, M., Eckert, R., Collins, B., & Chowla, R. (2017). *Caring to change: How*

compassionate leadership can stimulate innovation in health care. The Kings Fund*.*

https://www.kingsfund.org.uk/publications/caring-change

West. M., Eckert, R., Steward, K., & Passmore, B. (2014). *Developing collective leadership*

for health care. The Kings Fund.

https://www.kingsfund.org.uk/publications/developing-collective-leadership-health-

care

Westwood, S., Morison, L., Allt, J., & Holmes, N. (2017). Predictors of emotional exhaustion, disengagement and burnout among improving access to psychological therapies (IAPT) practitioners. *Journal of Mental Health, 26*(2), 172-179. https://doi.org/10.1080/09638237.2016.1276540

Wolpert, M. (2014). Uses and abuses of patient reported outcome measures (PROMs): Potential iatrogenic impact of PROMs implementation and how it can be mitigated. *Administration and Policy in Mental Health and Mental Health Services Research*, *41,* 141-145. https://doi.org/10.1007/s10488-013-0509-1

Xiao, H., Hayes, J. A., Castonguay, L. G., McAleavey, A. A., & Locke, B. D. (2017). Therapist effects and the impacts of therapy nonattendance. *Psychotherapy, 54*(1), 58–65. https://doi.org/10.1037/pst0000103

Yeh, M., Eastman, K., & Cheung, M. K. (1994). Children and adolescents in community health centers: Does the ethnicity or the language of the therapist matter? *Journal of Community Psychology, 22,* 153-163. https://doi.org/10.1002/1520-6629(199404)22:2<153::AID-JCOP2290220210>3.0.CO;2-R

Yukl, G. (2009). Leading organizational learning: Reflections on theory and research. *The Leadership Quarterly*, *20*(1), 49-53. https://doi.org/10.1016/j.leaqua.2008.11.006

Zamosttny, K. P., Corrigan, J. D., & Eggart, M. A. (1981). Replication and extension of social influence process in counselling: A field study. *Journal of Counseling Psychology, 28,* 481-489.

Zeeck, A., Orlinsky, D. E., Hermann, S., Joos, A., Wirsching, M., Weidmann, W., & Hartmann, A. (2012). Stressful involvement in psychotherapeutic work: Therapist, client and process correlates. *Psychotherapy Research, 22*(5), 543-555. https://doi.org/10.1080/10503307.2012.683345

Ziem, M., & Hoyer, J. (2020). Modest, yet progressive: Effective therapists tend to rate therapeutic change less positively than their patients. *Psychotherapy Research, 30*(4), 433-446. https://doi.org/10.1080/10503307.2019.1631502

Zilcha-Mano, S., & Errázuriz, P. (2015). One size does not fit all: Examining heterogeneity and identifying moderators of the alliance–outcome association. *Journal of Counseling Psychology, 62*(4)*,* 579–591. https://doi.org/10.1037/cou0000103

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

Table of Appendices

Appendix A   Validation Literature Review and Methods

Appendix B   Primary Literature Review Method

Appendix C   Ethics Approval Phase 1 and 2 July 2017

Appendix D   Full List of IST Service Recommendations in Phase 2

Appendix E   Ethics Approval Phase 3 September 2018

Appendix F   Focus Group Resources and Questions

Appendix G   Deliberate Practice Hints and Tips and Recording Form

Appendix H   Core Therapists Full MLM

Appendix I   Core Therapists Tests of Model Assumptions

Appendix J   Core Therapists Sensitivity Test

Appendix K   All Therapists Full MLM

Appendix L   All Therapists Tests of Model Assumptions

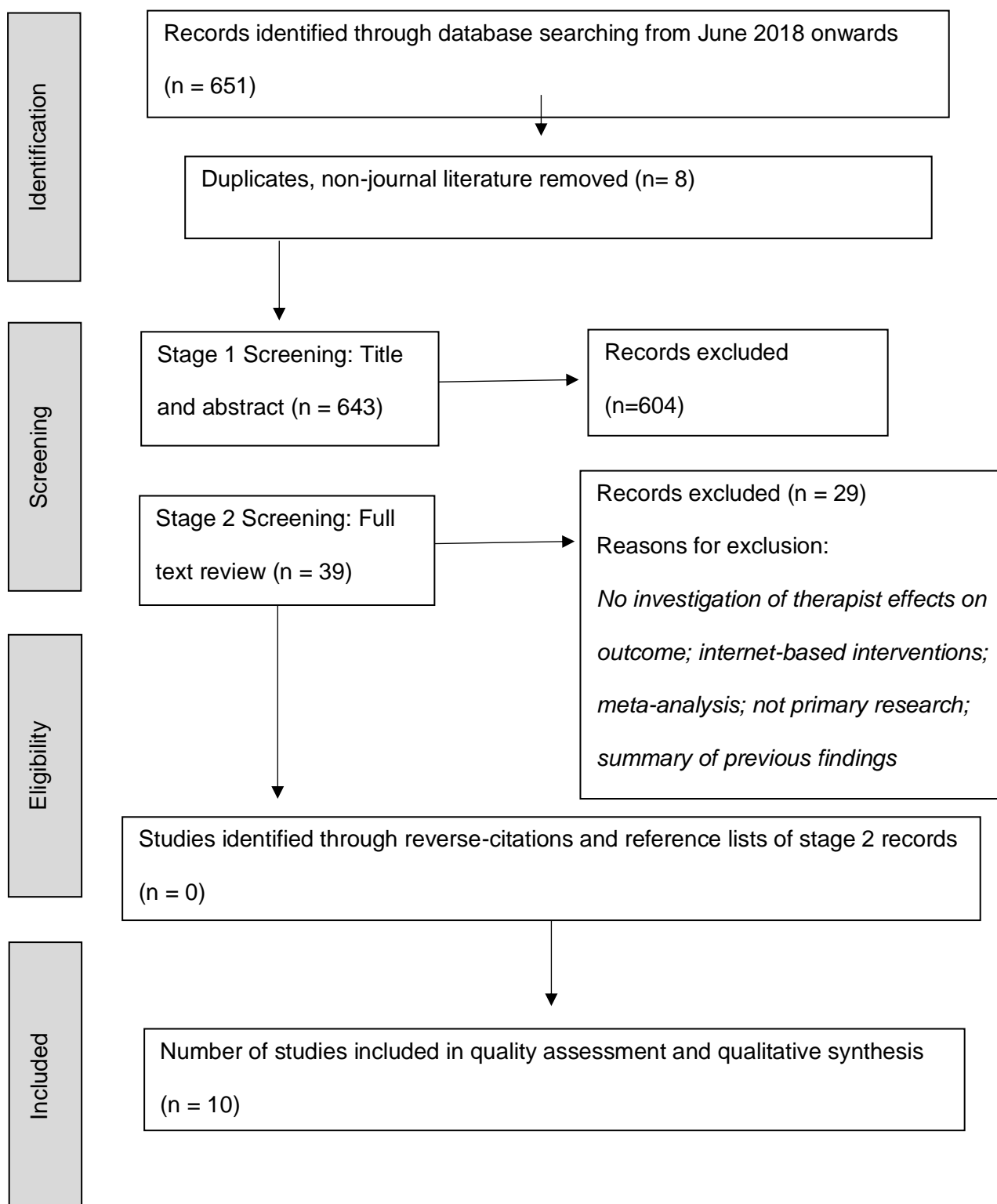**Appendix A: Validation Literature Review and Methods**

Overview

This chapter presented a systematic review originally undertaken to generate ideas for an intervention aimed at improving the outcomes of therapists in routine practice and reducing therapist variability. Following the development and implementation of this intervention as set out in subsequent chapters of this thesis, a refresh of the initial review was undertaken in 2021 to provide validation of the selected interventions – in effect, to summarise any additions to the research literature that might confirm or contradict the conclusions from the initial literature review documented in Chapter 2.

This second literature review focused on studies that investigated the therapist effect, including any explanatory or moderating factors identified or proposed, over the period June 2018 – Jan 2021. A systematic review was conducted, including studies found in Medline, PsychINFO, Web of Science and SCOPUS and through reverse citations. Ten eligible studies were identified, and a narrative synthesis was provided in the context of specific domains of therapist effects themed from the previous review. Particular attention was paid to any potential explanatory or moderating factors identified or missing from the previous review, that may validate or otherwise, an intervention aiming to reduce variability and increase therapist effectiveness. Most studies were observational and limitations of the studies mirrored that of the previous review. Studies in this period were consistent with the previous review and therefore validated the intervention areas explored within the research study. A PRISMA diagram covering this secondary search is shown in Figure 10.1

Figure 10.1

PRISMA Diagram of Study Selection for Search June 2018 – January 2021

```
┌──────────────┐   ┌─────────────────────────────────────────────────────────────┐
│              │   │ Records identified through database searching from June 2018  │
│ Identification│  │ onwards                                                       │
│              │   │ (n = 651)                                                     │
└──────────────┘   └─────────────────────────────────────────────────────────────┘
                                              │
                                              ▼
                   ┌─────────────────────────────────────────────────────────────┐
                   │ Duplicates, non-journal literature removed (n= 8)            │
                   └─────────────────────────────────────────────────────────────┘
                                              │
                                              ▼
┌──────────────┐   ┌──────────────────────────┐      ┌──────────────────────────┐
│              │   │ Stage 1 Screening: Title  │─────▶│ Records excluded          │
│              │   │ and abstract (n = 643)    │      │ (n=604)                   │
│ Screening    │   └──────────────────────────┘      └──────────────────────────┘
│              │
│              │   ┌──────────────────────────┐      ┌──────────────────────────────────┐
│              │   │ Stage 2 Screening: Full   │─────▶│ Records excluded (n = 29)         │
│              │   │ text review (n = 39)      │      │ Reasons for exclusion:            │
└──────────────┘   └──────────────────────────┘      │ *No investigation of therapist    │
                                              │        │ effects on outcome; internet-based │
┌──────────────┐                              │        │ interventions; meta-analysis; not  │
│              │                              │        │ primary research; summary of       │
│ Eligibility  │                              │        │ previous findings*                 │
│              │                              │        └──────────────────────────────────┘
│              │                              ▼
│              │   ┌─────────────────────────────────────────────────────────────┐
│              │   │ Studies identified through reverse-citations and reference    │
└──────────────┘   │ lists of stage 2 records                                      │
                   │ (n = 0)                                                       │
                   └─────────────────────────────────────────────────────────────┘
                                              │
                                              ▼
┌──────────────┐   ┌─────────────────────────────────────────────────────────────┐
│              │   │ Number of studies included in quality assessment and          │
│ Included     │   │ qualitative synthesis                                         │
│              │   │ (n = 10)                                                      │
└──────────────┘   └─────────────────────────────────────────────────────────────┘
```

Results

The results of the search are presented below. Initially the features of the search results are summarised and then a summary is provided of the papers, structured by the categories consistent with the main review (Chapter 2).

Study Characteristics

Ten papers were included in the final review and are described in Table 10.1. Three controlled trials were included in the review with the remainder being observational studies. Sample sizes in the studies ranged from 15–466 therapists, with a mean of 148.5 (SD = 178.1) and 51–26,888 patients, with a mean 6904.5 (SD = 10645.5). Nine of the ten studies reported an average patients per therapist, and these ranged from 3 to 67, with a mean of 32.6 (SD = 25). Treatment settings included outpatient clinics (k = 2), primary care (k = 4), university counselling clinics (k = 4), secondary care services (k = 2), other psychological therapy centres (k = 3). A range of psychological interventions were provided, including low intensity CBT (k = 3), CBT (k = 4), counselling (k = 3), Interpersonal Psychotherapy (IPT, k = 3), Eye Movement Desensitisation and Reprocessing (EMDR, k = 3), psychodynamic psychotherapy (k = 1) and mixed or non-identified psychological therapies (k = 5). Patient presentations were primarily depression and/or anxiety disorders (k = 6) or mixed presentations (k = 4).

Table 10.1

Study Characteristics (2018 – 2021)

Study	No. of patients	No. of therapists	Mean patients per therapist	Diagnosis	Outcome measure(s)	Intervention(s)	Treatment Setting(s)	Therapist effects analysis	Significant therapist effects found
RCT									
Brattland et al. 2018	170	20	8	Mixed	BASIS-32; ORS	Mixed psychotherapy	Outpatient psychiatric clinic	MLM	Yes
Brattland et al. 2019	170	20	8	Mixed	BASIS-32; ORS	Mixed psychotherapy	Outpatient psychiatric clinic	MLM	Yes
Delgadillo, de Jong et al. 2018	2233	77	31	Depression / anxiety	PHQ-9; GAD-7	Low intensity CBT, CBT, IPT, Counselling, EMDR	Primary care IAPT services	MLM	Yes
Observational									
Chui & Hill, 2020	51	15	3	Mixed	SES; WAI-SR	Psychodynamic psychotherapy	University training clinic	MLM	Yes
Delgadillo, Saxon et al. 2018	2509	49	45	Depression / anxiety	PHQ-9; GAD-7	Low intensity CBT, CBT, IPT, Counselling, EMDR	Primary care IAPT service	MLM	Yes

Delgadillo et al. 2020	4052	69	Not reported	Depression / anxiety	PHQ-9; GAD-7	Low intensity CBT, CBT, IPT, Counselling, EMDR	Primary care IAPT service	MLM	Yes
Firth et al. 2019	26888	462	58	Depression / anxiety	CORE-OM	Mixed psychotherapy	Primary care, secondary care, university, voluntary, workplace	MLM	Yes
Firth et al. 2020	26814	466	58	Depression / anxiety	CORE-OM	Mixed psychotherapy	Primary care, secondary care, university, voluntary, workplace	MLM	Yes
Goldberg et al. 2018	5078	238	67	Mixed	OQ-45	Mixed psychotherapy	University counselling centre	MLM	Yes
Ziem & Hoyer, 2020	1080	69	16	Depression / anxiety	BSI; SWLS; CGI	CBT	Psychological therapy clinic	MLM	No

Note: BASIS-32 = Behavioural and Symptom Idenitification Scale; BSI = Brief Symptom Inventory; CBT = Cognitive Behavioural Therapy; CGI = Clinical Global Impression Scale; CORE-OM = Clinical Outcomes in Routine Evaluation-Outcome Measure; EMDR = eye movement desensitisation and reprocessing; GAD-7 = Generalised Anxiety Disorder-7; IAPT = Improving Access to Psychological Therapies; IPT = Interpersonal Psychotherapy; MLM = Multi-level Modelling; ORS = Outcome Rating Scale; OQ-45 = Outcome Questionnaire-45; PHQ-9 = Patient Health Questionnaire-9; SES = Session Evaluation Scale; SWLS = Satisfaction with Life Scale; WAI-SR = Working Aliiance Inventory – Shortened Version;

Intrapersonal Characteristics

Chui and Hill (2020) investigated the impact of therapist calmness and fatigue on patients' ratings of session quality and alliance. A sample of 51 patients and 15 therapists was included and analysed using MLM. This study did not explore the therapist effect (though this was reported) and did not use a symptom-based outcome measure due to investigating a state rather than trait-based phenomenon. However, it did pose the suggestion that therapist-specific intrapersonal factors could impact on patients. The study found that therapists who were calm prior to the session and became calmer during the session, had high quality sessions and strong working alliance as rated by therapist and patient. Therapist pre-session fatigue was not associated with patient ratings of session quality or working alliance, though therapists who reported an increase in fatigue after sessions were judged as having better sessions by patients. Due to the established link between alliance and outcome, this study concluded that calmer, less fatigued therapists would have better sessions, alliance, and by association, outcomes. There was significant variability between therapists, particularly on therapist self-rated measures. Therapist level effects accounted for 26% of therapist-rated working alliance variability and 13% of therapist-rated session quality. In contrast, therapist effects accounted for just 3% of variance on patient-rated working alliance, and 1% of patient-rated session quality. Although therapist variability was not the focus of this study, the impact of therapist calmness and fatigue on outcomes suggested that an intervention that aimed to improve therapist wellbeing may be helpful and is supported by this evidence. Small sample sizes limited this study in terms of being able to identify therapist effects.

In a study of 2509 patients and 49 therapists, Delgadillo, Saxon et al. (2018) investigated associations between therapists' occupational burnout and job satisfaction with patient outcomes. They observed a 5.9% therapist effect, 31-39% of which was explained by increased disengagement (burnout) and lower job satisfaction. Increased levels of disengagement (burnout) were associated with poorer depression and anxiety outcomes for patients, and lower job satisfaction was associated with poorer depression, but not anxiety,

outcomes. The study utilised a large patient and therapist cohort overall, though the number of patients per therapists may not have been sufficient for reliable estimates of all model parameters. Two plausible conclusions from the study were proposed. First, that more stressful working conditions may increase burnout (specifically disengagement), which may in turn limit clinical outcomes for patients due to the potential impact on the therapeutic alliance of stressful disengagement. Alternatively, the context of having poorer patient outcomes (due to other factors) may lead the therapist to experience burn out or lower job satisfaction. In either case, the findings would indicate that an emphasis on how services can support therapists to feel more engaged and positive about their work, whether through their own wellbeing or sense of competence, may well pay dividends both for the wellbeing of the individual therapist, but also for the outcomes of patients.

Delgadillo et al. (2020) assessed if therapist effects might be influenced by therapists' personality (based on the 'big five' personality traits), after controlling for experience, technical competence, and reflective ability. This retrospective observational study comprising a sample size of 4052 patients and 69 trainee and recently qualified therapists found that, for Psychological Wellbeing Practitioners (PWPs), an above average level 'agreeableness' was significantly associated with poorer patient outcomes, and for CBT therapists, an above average 'openness to experience' was significantly associated with poorer outcomes. Neither experience, technical competence, nor reflective ability were associated with better outcomes, though it should be noted that technical competence for CBT therapists and reflective ability for both were based on training assessments (i.e., therapist selected sessions).

As a complement to the Nissen-Lie, Monsen et al. (2013), and Nissen-Lie et al. (2017) studies investigating professional self-doubt, Ziem and Hoyer (2020) studied the potential for illusory superiority within therapists and its relationship to patient outcomes. The sample size for the study comprised 1080 patients and 69 therapists. The study found no significant therapist effect, so there are limited conclusions in relation to the impact of therapists' self-assessments of patient progress. In addition, there was some evidence of a

positive association between therapists' lower estimation of patient outcomes and their therapeutic effectiveness. However, due to the lack of therapist effect, the authors concluded this may be more a feature of effective therapist- patient dyads rather than a stable therapist characteristic.

Context of the Therapist: Use of Feedback and Activities to Deliberately Improve Practice

Brattland et al. (2018) conducted a naturalistic randomised trial investigating the impact of providing therapists with sessional feedback, on patient ratings of their symptoms and experience of the session. The sample size for this study was 170 patients and 20 therapists which, although small for a study of therapist effects, focused on the comparison of a therapist level intervention versus treatment as usual whilst controlling the 'undefined' therapist effect. Two groups: treatment as usual (TAU) and the feedback condition were compared. The feedback condition also involved regular supervision of therapists where they were encouraged to deliberately practice skills based on the feedback from their patients. Patients in the feedback condition demonstrated superior outcomes to those in the TAU condition, even when therapist variability and initial severity of outcomes were controlled. Therapist variability explained 9-10% of the outcome, but was not significant in relation to how feedback influenced outcomes.

A follow up analysis of the Brattland et al. (2018) study, investigated whether the working alliance had a mediating effect on routine outcome monitoring (ROM), in terms of patient symptoms, thus including both the use of feedback and the impact of the alliance on therapy outcomes (Brattland et al., 2019). This study found that patients' alliance ratings increased more in the ROM condition, and a greater alliance increase predicted less post-treatment impairment. The therapist effect was even larger in this analysis than in the 2018 study, with 17% of variance on the patient wellbeing and alliance measures being due to the therapist effect. This suggests that a more individualised approach to how therapists are supported, particularly with the use of ROMs, may be advised. In addition, the authors pointed out that improving therapists' outcomes and alliance may be more about supporting

therapists' response to alliance-related feedback, rather than simply getting the feedback itself.

Delgadillo, de Jong et al. (2018) also investigated the impact of the use of feedback-informed treatment on patient symptoms in a randomised controlled multi-site trial with a sample size of 2233 patients and 77 therapists. A control condition involved therapists offering TAU and standard use of sessional outcome measures with patients. This was compared with a condition where therapists were provided with feedback on whether patients were 'on-track', based on an automated computer algorithm which alerted therapists to patients who were 'not-on-track'. This then triggered the therapist to conduct conversations with the patient and within supervision, to identify potential blocks to improvement and adaptations that could be made to support improvement. The study found that the use of the enhanced feedback system was significant in improving outcomes for patients who were 'not-on-track', but had no significant impact on other patients' outcomes. The study reported a therapist effect of 11% when case mix was controlled for, which the authors noted is larger than the effect of feedback - indicating that further investigation of factors that characterise underperforming (or highly effective) therapists would be warranted.

Firth et al. (2019) found a significant clinic effect in addition to the therapist effect in an observational study involving over 26,000 patients, 462 therapists and 30 clinics. It is of note that, although adding the clinic level reduced the therapist effect, the therapist effect was generally consistent despite the addition of other variables. By contrast, the clinic effect reduced significantly when patient severity, employment, and community variables of ethnicity mix and type of clinic were added to the model. Despite this reduction the clinic effect remained significant (i.e., there were factors contributing to patient outcome at the clinic level that remained unexplained by any of the variables measured).

A follow up analysis of the Firth et al. (2019) study (Firth et al., 2020) demonstrated that there were different levels of therapist effects in different clinic contexts, with the largest in primary care settings (after controlling for baseline patient variables and clinic variability). The finding that the differences in therapist effectiveness varies according to the setting they

are working within, may have implications for the intervention development of the current study. One of the recommendations of the Firth et al. (2020) study was, where larger therapist effects are found (i.e., primary care settings), to focus more resource on: understanding the differences in effectiveness of therapists; understanding potential differences in case-load or care provision; improving outcomes of less effective therapists through deliberate practice; and incorporating practices of highly effective therapists. This is in support of the intervention developed within the current study. Variability in therapists' effectiveness in working with patient with high financial distress was investigated by Goldberg et al. (2018). Although there was variability in therapists' ability to engage patients beyond one session, there was no significant variability in outcomes between therapists associated with patient reported financial distress.

Study and Review Limitations

The limitations of these studies and the review itself mirrors those identified in the initial 2012–2018 search so will not be repeated here.

Therapist Effects Review 2021

Subsequent to the validation review described above, the 7th edition of *Bergin & Garfield's Handbook of Psychotherapy and Behavior Change* (Barkham et al., 2021) was published at the end of the same year. As summarised in Chapter 2, the 6th edition of the *Handbook* (Lambert, 2013) included an updated review of the therapist effects literature, following chapters in previous editions reviewing therapist variables (Beutler et al. 1994, 2003). The 7th edition chapter (Wampold & Owen, 2021) provides a comprehensive review of the history and study of the therapist effect (summarised briefly in Chapter 1), as well as emerging understandings of the characteristics of effective therapists. This was, therefore, a key text in understanding the position of the therapist effects research context at the point of completion of the study Phases and data analysis, and was used as an additional validation tool for the results of this work.

The Wampold and Owen (2021) review confirms that the most significant development in the therapist effects field since the 2013 review, has been the identification

of the characteristics and actions of effective therapists. This was in its infancy in 2013, but is one of the key areas of interest of this thesis, and has been summarised in these review chapters (Chapters 2 and 3). Wampold and Owen (2021) concluded that, based on the studies reviewed, there are areas which appear to have potential in advancing our current understanding of the association between therapist attributes and activities and effective outcomes/effective therapists. These are first, interpersonal skills in challenging contexts; and second, attitudes towards professional development that lead to activities which aim to improve skills and practice. Additionally, the ability to work with ethnically/culturally diverse patients appears varied, with some therapists being more effective than others with these patients. In terms of how these areas may impact training and practice, they suggest that training should focus on the acquisition of interpersonal skills to at least an equivalent level as the training on manualised treatment protocols. They conclude that *"given the evidence reviewed that indicates that therapists, as a general rule, do not improve over the course of their careers, efforts need to be made to create environments that encourage improvement."* (p.320). In addition, they suggest from the review that professional humility and a desire to improve are important, but not sufficient, and that skills need to be actively worked on and practiced, such as through the means of deliberate practice. Similarly, in relation to the information about their patients' outcomes, the authors suggested that having this intelligence available to therapists is crucial. However, the potential for this type of information to be used by services or managers to performance manage, or even dismiss, therapists could be counterproductive, creating low morale and burnout. Instead, this needs to be approached carefully in a way that supports professional development and learning.

Summary

This chapter presented literature included in a validation literature review as a follow up to the original literature review presented in Chapter 2. The reason for separating the reviews in this way was to provide the basis of the development of the collaborative research intervention in 2018, followed by additional literature that was published subsequently. Literature published subsequently to the research Phase did not provide any evidence that

would have contradicted the decisions made based on the original review. Of note, there was additional evidence related to the impact of therapist wellbeing; use of ROMs and use of deliberate practice in the validation review, further reinforcing the chosen elements of the research intervention package.

Search strategy used to search papers, abstracts and key-terms

1 "therapist effects" OR "differential effects of therapists" OR "therapist outcome" OR "therapist variance" OR "effective therapist" OR "ineffective therapist"

2 "intraclass correlation" OR "multilevel model" OR "hierarchical linear model" OR "mixed models"

3 1 AND 2

Excluded studies with reason (Full-text eligibility review at stage 2)

First author	DOI	Reason for exclusion
Adler, G.	10.1097/NMD.0000000000000814	No investigation of therapist effects on outcomes
Anderson, T.	10.1002/capr.12302	Summary of previous findings
Bennett-Levy, J.	10.1016/j.jbtep.2018.08.004	Not primary research
Chui, H.	10.1037/cou0000393	No investigation of therapist effects on outcomes
Coyne, A. E.	10.1037/cou0000457	No investigation of therapist effects on outcomes
Coyne, A. E.	10.1037/int0000125	No investigation of therapist effects on outcomes
Easden, M. H.	10.1080/10503307.2018.1540895	No investigation of therapist effects on outcomes
Finegan, M.	10.1080/10503307.2019.1649500	No investigation of therapist effects on outcomes
Fisher, H.	10.1037/cou0000377	No investigation of therapist effects on outcomes
Gonzalez, J.	10.15241/jg.8.4.314	No investigation of therapist effects on outcomes

Hardy, G. E.	10.1080/10503307.2017.1393575	No investigation of therapist effects on outcomes
Katz, M.	10.1037/cou0000299	No investigation of therapist effects on outcomes
Kivlighan, D.	10.1037/pst0000197	No investigation of therapist effects on outcomes
Lingu, A.	10.2196/18723	Internet-based intervention
Nikendei, C.	10.1002/capr.12232	No investigation of therapist effects on outcomes
Nissen-Lie, H. A.	10.1080/10503307.2020.1823030	No investigation of therapist effects on outcomes
Ottoman, K. E.	10.1016/j.brat.2019.103531	No investigation of therapist effects on outcomes
Paulick, J.	10.1037/int0000099	No investigation of therapist effects on outcomes
Petrowski, K.	10.1002/ccp.2334	No investigation of therapist effects on outcomes
Probst, T.	10.1002/cpp.2441	No investigation of therapist effects on outcomes
Rozental, A.	10.1192/bjo.2018.42	No investigation of therapist effects on outcomes
Santoft, F.	10.1016/j.janxdis.2019.102118	No investigation of therapist effects on outcomes
Sauer, E. M.	10.1037/pst0000304	No investigation of therapist effects on outcomes
Schwartz, C.	10.1037/ccp0000302	No investigation of therapist effects on outcomes
Sembill, A.	10.1080/10503307.2017.1405170	No investigation of therapist effects on outcomes
Snowdon, D. A.	10.1186/s12913-019-4873-8	No investigation of therapist effects on outcomes
Stuart, S.	10.1007/s10597-017-0220-x	No investigation of therapist effects on outcomes
Tschuschke, V.	10.1097/NMD.0000000000001111	No investigation of therapist effects on outcomes

| Walsh, L. M. | 10.1080/10503307.2018.1469802 | Meta-analysis |

Note: Only the first author, DOI and reason for exclusion have been given for conciseness.

Appendix B: Primary Literature Review Method

Search strategy used to search papers, abstracts and key-terms

1 "therapist effects" OR "differential effects of therapists" OR "therapist outcome" OR "therapist variance" OR "effective therapist" OR "ineffective therapist"

2 "intraclass correlation" OR "multilevel model" OR "hierarchical linear model" OR "mixed models"

3 1 AND 2

Excluded studies with reason (Full-text eligibility review at stage 2)

First author	DOI	Reason for exclusion
Ehlers, A.	10.1016/j.brat.2013.08.006	No explanatory therapist factors investigated.
Gibbons, M. B.	10.1037/a0039302	No investigation of therapist effects on outcomes
Hara, K. M.	10.1080/16506073.2016.1253605	Primarily investigates patient effects
Hayes, J. A.	10.1037/2Fcou0000098	No additional explanatory therapist factors investigated in addition to Hayes (2015) paper included in review
McClintock, A. S.	10.1037/cou0000188	No investigation of therapist effects on outcomes
Minami, T.	10.1007/s11135-011-9548-4	No explanatory therapist factors investigated.
Nissen-Lie, H. A.	10.1002/cpp.1891	No investigation of therapist effects on outcomes
Tschacher, W.	10.1002/cpp.1822	No investigation of therapist effects on outcomes
Werbart,A.	10.1037/2Fa0031386	No explanatory therapist factors investigated.

Note: Only the first author, DOI and reason for exclusion have been given for conciseness.

Risk of bias and quality assessment summary table

Risk of Bias for RCTs

First Author and Year	Selection Bias	Performance Bias	Detection Bias	Attrition Bias	Reporting Bias	Other Bias	Overall Quality Rating
Errázuriz et al. (2018)	M[a]	H[b]	U[c]	L	U[d]	L	Moderate
Zilcha-Mano et al. (2015)	M[a]	H[b]	U[c]	L	U[d]	L	Moderate

[a]Therapists were given discretionary permission to disclose intervention condition to patients.

[b]Patient may have been aware of intervention if disclosed by therapist. Therapist aware of intervention condition (not possible to blind).

[c]Not reported.

[d]Findings reported but unclear if protocol available.

Quality Assessment Rating of Observational Studies

Author and Year	Overall Quality Rating	Rationale if not rated Good
Ali et al. 2014	Good	
Chow et al. 2015	Fair	Sample size justification and effect estimates not provided. Exposure of interest not measured prior to outcome measurement. Exposure assessed only once.
Cologon et al. 2017	Good	
Erekson et al. 2017	Good	
Firth et al. 2015	Good	
Goldberg, Hoyt et al. 2016	Good	
Goldberg, Rousmaniere et al. 2016	Good	
Green et al. 2014	Good	
Hayes et al. 2015	Fair	Sample size justification and effect estimates not provided.
Kraus et al. 2016	Good	

Laska et al. 2013	Good	
Nissen-Lie, Havik et al. 2013	Fair	Sample size justification and effect estimates not provided. Limited reliability of independent variable
Nissen-Lie, Monsen et al. 2013	Fair	Sample size justification and effect estimates not provided. Limited reliability of independent variable
Nissen-Lie et al. 2016	Good	
Nissen-Lie et al. 2017	Fair	Sample size justification and effect estimates not provided. Limited reliability of independent variable
Odyniec et al. 2017	Good	
Owen et al. 2016	Good	
Pereira et al. 2017	Fair	Participation rate lower than 50% Sample size justification and effect estimates not provided Over a third of treatment was 1 session (limited timeframe to measure exposure and outcome)
Rousmaniere et al. 2016	Good	
Saxon & Barkham, 2012	Good	
Saxon, Barkham et al. 2017	Good	
Saxon, Firth et al. 2017	Good	
Slone & Owen 2015	Fair	Sample size justification and effect estimates not provided.
Xiao et al. 2017	Good	
Zeeck et al. 2012	Fair	Indications that the two study settings may have led to two different patient populations being recruited. Sample size justification and effect estimates not provided.
Zimmerman et al. 2017	Good	

Appendix C: Ethics Approval Phase 1 and 2 July 2017

NHS
Health Research Authority

Mrs Katy James
65 Rosebery Rd
Norwich
NR3 3AB

Email:
hra.approval@nhs.net

05 July 2017

Dear Mrs James

Letter of HRA Approval

Study title:	**The contribution of therapist factors towards patient outcomes during routine psychological therapy practice.**
IRAS project ID:	**221057**
REC reference:	**17/EE/0251**
Sponsor	**University of Sheffield**

I am pleased to confirm that **HRA Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications noted in this letter.

Participation of NHS Organisations in England

The sponsor should now provide a copy of this letter to all participating NHS organisations in England.

Appendix B provides important information for sponsors and participating NHS organisations in England for arranging and confirming capacity and capability. **Please read *Appendix B* carefully**, in particular the following sections:

- *Participating NHS organisations in England* – this clarifies the types of participating organisations in the study and whether or not all organisations will be undertaking the same activities
- *Confirmation of capacity and capability* - this confirms whether or not each type of participating NHS organisation in England is expected to give formal confirmation of capacity and capability. Where formal confirmation is not expected, the section also provides details on the time limit given to participating organisations to opt out of the study, or request additional time, before their participation is assumed.

- *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria) -* this provides detail on the form of agreement to be used in the study to confirm capacity and capability, where applicable.

Further information on funding, HR processes, and compliance with HRA criteria and standards is also provided.

It is critical that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details and further information about working with the research management function for each organisation can be accessed from www.hra.nhs.uk/hra-approval.

Appendices

The HRA Approval letter contains the following appendices:

- A – List of documents reviewed during HRA assessment
- B – Summary of HRA assessment

After HRA Approval

The document *"After Ethical Review – guidance for sponsors and investigators"*, issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

In addition to the guidance in the above, please note the following:

- HRA Approval applies for the duration of your REC favourable opinion, unless otherwise notified in writing by the HRA.
- Substantial amendments should be submitted directly to the Research Ethics Committee, as detailed in the *After Ethical Review* document. Non-substantial amendments should be submitted for review by the HRA using the form provided on the HRA website, and emailed to hra.amendments@nhs.net.
- The HRA will categorise amendments (substantial and non-substantial) and issue confirmation of continued HRA Approval. Further details can be found on the HRA website.

Scope

HRA Approval provides an approval for research involving patients or staff in NHS organisations in England.

If your study involves NHS organisations in other countries in the UK, please contact the relevant national coordinating functions for support and advice. Further information can be found at http://www.hra.nhs.uk/resources/applying-for-reviews/nhs-hsc-rd-review/.

If there are participating non-NHS organisations, local agreement should be obtained in accordance with the procedures of the local participating non-NHS organisation.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/.

HRA Training

We are pleased to welcome researchers and research management staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

Your IRAS project ID is **221057**. Please quote this on all correspondence.

Yours sincerely

Catherine Adams

Senior Assessor

Email: hra.approval@nhs.net

Copy to: Deborah McClean, Sponsor Contact
*   Dr Bonnie Teague, Norfolk and Suffolk Foundation Trust*

Appendix A - List of Documents

The final document set assessed and approved by HRA Approval is listed below.

Document	Version	Date
Covering letter on headed paper [Covering letter to REC Committee]	1	22 June 2017
IRAS Application Form [IRAS_Form_01062017]		01 June 2017
Other [Second Supervisor Info]	1	31 May 2017
Other [SoA dated]	1	05 July 2017
Other [SoE dated]	1	05 July 2017
Referee's report or other scientific critique report [Statistician's Letter]	1	03 March 2017
Research protocol or project proposal [Phase 1 & 2 Protocol]	3	22 June 2017
Summary CV for Chief Investigator (CI) [Summary CV]	1	03 March 2017
Summary CV for supervisor (student research) [Supervisor Info M.Barkham]	1	31 May 2017

Appendix B - Summary of HRA Assessment

This appendix provides assurance to you, the sponsor and the NHS in England that the study, as reviewed for HRA Approval, is compliant with relevant standards. It also provides information and clarification, where appropriate, to participating NHS organisations in England to assist in assessing and arranging capacity and capability.

For information on how the sponsor should be working with participating NHS organisations in

England, please refer to the, *participating NHS organisations, capacity and capability* and *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* sections in this appendix.

The following person is the sponsor contact for the purpose of addressing participating organisation questions relating to the study:

Deborah McClean
E-mail d.mcclean@sheffield.ac.uk
Telephone 01142221449

HRA assessment criteria

Section	HRA Assessment Criteria	Compliant with Standards	Comments
1.1	IRAS application completed correctly	Yes	No comments
2.1	Participant information/consent documents and consent process	Yes	No comments
3.1	Protocol assessment	Yes	No comments
4.1	Allocation of responsibilities and rights are agreed and documented	Yes	A statement of activities will act as agreement of an NHS organisation to participate. The sponsor is not requesting and does not expect any other site agreement.
4.2	Insurance/indemnity arrangements assessed	Yes	Where applicable, independent contractors (e.g. General Practitioners) should ensure that the professional indemnity provided by their medical defence organisation covers the activities expected of them for this
Section	HRA Assessment Criteria	Compliant with Standards	Comments
			research study
4.3	Financial arrangements assessed	Yes	No funding is to be provided as detailed in the Statement of Activities.
5.1	Compliance with the Data Protection Act and data security issues assessed	Yes	No comments
5.2	CTIMPS – Arrangements for compliance with the Clinical Trials Regulations assessed	Not Applicable	No comments

5.3	Compliance with any applicable laws or regulations	Yes	No comments
6.1	NHS Research Ethics Committee favourable opinion received for applicable studies	Yes	No comments
6.2	CTIMPS – Clinical Trials Authorisation (CTA) letter received	Not Applicable	No comments
6.3	Devices – MHRA notice of no objection received	Not Applicable	No comments
6.4	Other regulatory approvals and authorisations received	Not Applicable	No comments

Participating NHS Organisations in England

This provides detail on the types of participating NHS organisations in the study and a statement as to whether the activities at all organisations are the same or different.

This is a single site study and therefore there is only one 'site-type' undertaking activities detailed in the protocol.

The Chief Investigator or sponsor should share relevant study documents with participating NHS organisations in England in order to put arrangements in place to deliver the study. The documents should be sent to both the local study team, where applicable, and the office providing the research management function at the participating organisation. For NIHR CRN Portfolio studies, the Local LCRN contact should also be copied into this correspondence. For further guidance on working with participating NHS organisations please see the HRA website.

If chief investigators, sponsors or principal investigators are asked to complete site level forms for participating NHS organisations in England which are not provided in IRAS or on the HRA website, the chief investigator, sponsor or principal investigator should notify the HRA immediately at hra.approval@nhs.net. The HRA will work with these organisations to achieve a consistent approach to information provision.

Confirmation of Capacity and Capability

This describes whether formal confirmation of capacity and capability is expected from participating NHS organisations in England.

Participating NHS organisations in England **will be expected to formally confirm their capacity and capability to host this research**.

- Following issue of this letter, participating NHS organisations in England may now confirm to the sponsor their capacity and capability to host this research, when ready to do so. How capacity and capacity will be confirmed is detailed in the *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* section of this appendix.
- The Assessing, Arranging, and Confirming document on the HRA website provides further information for the sponsor and NHS organisations on assessing, arranging and confirming capacity and capability.

Principal Investigator Suitability

This confirms whether the sponsor position on whether a PI, LC or neither should be in place is correct for each type of participating NHS organisation in England and the minimum expectations for education, training and experience that PIs should meet (where applicable).

The Chief Investigator will retain responsibility for research activity at the participating organisation. GCP training is not a generic training expectation, in line with the HRA statement on training expectations.

HR Good Practice Resource Pack Expectations

This confirms the HR Good Practice Resource Pack expectations for the study and the pre-engagement checks that should and should not be undertaken

No Honorary Research Contracts, Letters of Access or pre-engagement checks are expected for local staff employed by the participating NHS organisations.

Other Information to Aid Study Set-up

This details any other information that may be helpful to sponsors and participating NHS organisations in England to aid study set-up.

The applicant has indicated that they do not intend to apply for inclusion on the NIHR CRN Portfolio.

Appendix D Full List of IST Service Recommendations in Phase 2

Clinical and Pathway Recommendations

1. Revise the patient pathway to ensure it is simplified, clear and more linear for the patient

2. Ensure that all patients are assessed on entering the service

3. Ensure that all patients are clustered at assessment

4. Problem descriptors should be recorded at assessment and decision to treat should be based on the problem descriptor and NICE guidance

5. Patients on the enhanced care pathway who receive an IAPT psychological therapy should be included in IAPT data

6. Some of the 'social' care pathway could be considered for inclusion in IAPT data set e.g. groups/interventions that are behavioural activation and supervised by a qualified member of staff

7. Address long waits and consider an interim pathway for longest waits

8. Ensure all interventions/treatments are evidence-based and the full dosage available to patients

9. Counsellors should be required to train in an IAPT modality and a training plan put in place

Admin and Management Recommendations

1. Examine reasons behind high attrition rates and declining referrals, as well as under provision to older adults and BME communities

2. Revisit demand and capacity planning to address root cause of long waits

3. Allocate increased analytical resource to IAPT within NSFT, ensure that IAPT data is easily accessible and, where possible, automated through trust data warehouse

4. Use NHS Digital data in internal reports within both NSFT and CCGs, with local data only used for most recent months where NHS Digital data is not yet available

5. Carry out monthly reconciliation of NHS Digital and local data using the PAVE report

6. Align data quality measures to priorities such as problem descriptor, cluster and scores

7. Develop more sophisticated waiting list management processes and reports for all waits

8. Strengthen case management to include monitoring of delivered and forthcoming sessions, outcomes and DNA rates

9. Develop therapist dashboards to support case management

10. Regularly analyse outcomes by team/modality/therapist in order to focus service improvement

Appendix E: Ethics Approval Phase 3 September 2018

Ymchwil Iechyd
a Gofal Cymru
Health and Care
Research Wales

NHS
Health Research
Authority

Mrs Katy James
65 Rosebery Rd
Norwich
Norfolk
NR3 3AB

Email:
hra.approval@nhs.net
Research-
permissions@wales.nhs.uk

17 September 2018

Dear Mrs James

HRA and Health and Care

Research Wales (HCRW)   Approval Letter

Study title:	**Can psychological therapists outcomes be improved and variability reduced in routine psychotherapy practice?**
IRAS project ID:	**248085**
REC reference:	**18/NS/0104**
Sponsor	**University of Sheffield**

I am pleased to confirm that **HRA and Health and Care Research Wales (HCRW) Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

How should I continue to work with participating NHS organisations in England and Wales? You should now provide a copy of this letter to all participating NHS organisations in England and Wales, as well as any documentation that has been updated as a result of the assessment.

Following the arranging of capacity and capability, participating NHS organisations should **formally confirm** their capacity and capability to undertake the study. How this will be confirmed is detailed in the "*summary of assessment*" section towards the end of this letter.

You should provide, if you have not already done so, detailed instructions to each organisation as to how you will notify them that research activities may commence at site following their confirmation of capacity and capability (e.g. provision by you of a 'green light' email, formal notification following

a site initiation visit, activities may commence immediately following confirmation by participating organisation, etc.).

It is important that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details of the research management function for each organisation can be accessed here.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?

HRA and HCRW Approval does not apply to NHS/HSC organisations within the devolved administrations of Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report (including this letter) has been sent to the coordinating centre of each participating nation. You should work with the relevant national coordinating functions to ensure any nation specific checks are complete, and with each site so that they are able to give management permission for the study to begin.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?

HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your nonNHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?

The document "*After Ethical Review – guidance for sponsors and investigators*", issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including: ☐   Registration of research

- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

I am a participating NHS organisation in England or Wales. What should I do once I receive this letter?

You should work with the applicant and sponsor to complete any outstanding arrangements so you are able to confirm capacity and capability in line with the information provided in this letter.

The sponsor contact for this application is as follows:

Name: Dr Thomas Webb

Tel: 01142222000

Email: t.webb@sheffield.ac.uk

Who should I contact for further information?

Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **248085**. Please quote this on all correspondence.

Yours sincerely

Aliki Sifostratoudaki

Assessor

Email: hra.approval@nhs.net

Copy to: Dr Thomas Webb, University of Sheffield, Sponsor contact

*   Dr. Bonnie Teague, Norfolk & Suffolk NHS Foundation Trust, R&D contact*

List of Documents

The final document set assessed and approved by HRA and HCRW Approval is listed below.

Document	Version	Date
Covering letter on headed paper [Letter to REC]	1	17 September 2018
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Public Liability Insurance Certificate]		05 September 2017
HRA Schedule of Events [248085_SOE_All site activities_Assessed by HRA.docx]	1	17 September 2018
HRA Statement of Activities [248085_SOA_All site activities_Assessed by HRA]	1	17 September 2018
IRAS Application Form [IRAS_Form_07092018]	248085/1247396/37/742	07 September 2018
Letter from sponsor [Sponsor Letter]		20 August 2018
Letter from statistician [Letter from Statistician]		05 July 2018
Non-validated questionnaire [Therapist Evaluation v.1 Sept 2018]	1	07 September 2018
Other [Letter to HRA]	1	17 September 2018

Participant information sheet (PIS) [Participant Information Sheet]	2	17 September 2018
Referee's report or other scientific critique report [Letter of Scientific Approval]		06 August 2018
Research protocol or project proposal [Research Protocol]	1 * Aug 18 *	01 August 2018
Summary CV for Chief Investigator (CI) [Katy James]	1 * Aug 18 *	01 August 2018
Summary CV for supervisor (student research) [Supervisor CV M.Barkham]	2	07 September 2018
Summary CV for supervisor (student research) [Supervisor CV S.Kellett]	1	07 September 2018
Summary CV for supervisor (student research) [Supervisor CV D.Saxon]	1	07 September 2018
Validated questionnaire [Patient Health Questionnaire - 9 (PHQ-9)]	* date received*	30 August 2018
Validated questionnaire [GAD-7]	* date received*	30 August 2018

Summary of assessment

The following information provides assurance to you, the sponsor and the NHS in England and Wales that the study, as assessed for HRA and HCRW Approval, is compliant with relevant standards. It also provides information and clarification, where appropriate, to participating NHS organisations in England and Wales to assist in assessing, arranging and confirming capacity and capability.

Assessment criteria

Section	Assessment Criteria	Compliant with Standards	Comments
1.1	IRAS application completed correctly	Yes	No comments
2.1	Participant information/consent documents and consent process	Yes	No comments
3.1	Protocol assessment	Yes	No comments

Section	Assessment Criteria	Compliant with Standards	Comments
4.1	Allocation of responsibilities and rights are agreed and documented	Yes	The Sponsor contact has confirmed that the Statement of Activities and the Schedule of Events will form the agreement between the Sponsor and the research sites. No judgement on the cost attributions has been made.
4.2	Insurance/indemnity arrangements assessed	Yes	No comments
4.3	Financial arrangements assessed	Yes	This study is not receiving external funding.
5.1	Compliance with the Data Protection Act and data security issues assessed	Yes	The Applicant clarified that the local data manager at the NHS Trust will code and distribute the questionnaires to each participant. These will then be handed back to the host NHS organisation via the data manager who will pass these to the research team. The external research team will not need to identify participants. The participant's ID number will allow the research team to link questionnaire responses to outcome data.
5.2	CTIMPS – Arrangements for compliance with the Clinical Trials Regulations assessed	Not Applicable	No comments
5.3	Compliance with any applicable laws or regulations	Yes	No comments
6.1	NHS Research Ethics Committee favourable opinion received for applicable studies	Not Applicable	The REC Favourable Opinion letter has been issued.

6.2	CTIMPS – Clinical Trials Authorisation (CTA) letter received	Not Applicable	No comments
6.3	Devices – MHRA notice of no objection received	Not Applicable	No comments
6.4	Other regulatory approvals and authorisations received	Not Applicable	No comments

Participating NHS Organisations in England and Wales

This provides detail on the types of participating NHS organisations in the study and a statement as to whether the activities at all organisations are the same or different.

There is one site type in this study – research sites. Research sites will be responsible for all activity as listed in the Protocol.

The Chief Investigator or sponsor should share relevant study documents with participating NHS organisations in England and Wales in order to put arrangements in place to deliver the study. The documents should be sent to both the local study team, where applicable, and the office providing the research management function at the participating organisation. Where applicable, the local LCRN contact should also be copied into this correspondence.

If chief investigators, sponsors or principal investigators are asked to complete site level forms for participating NHS organisations in England and Wales which are not provided in IRAS, the HRA or HCRW websites, the chief investigator, sponsor or principal investigator should notify the HRA immediately at hra.approval@nhs.net or HCRW at Research-permissions@wales.nhs.uk. We will work with these organisations to achieve a consistent approach to information provision.

Principal Investigator Suitability

This confirms whether the sponsor position on whether a PI, LC or neither should be in place is correct for each type of participating NHS organisation in England and Wales, and the minimum expectations for education, training and experience that PIs should meet (where applicable).

A Principal Investigator (PI) would be expected at this site type. A PI would be expected to extract the required data from the clinical system, to code the data and disseminate the questionnaires to the therapists. A PI will also be expected to undertake the optional focus groups and supervisor training.

A Local Collaborator (LC) would not be expected at this site type as the local NHS staff will undertake the study activities at the site.

GCP training is not a generic training expectation, in line with the HRA/HCRW/MHRA statement on training expectations.

HR Good Practice Resource Pack Expectations

This confirms the HR Good Practice Resource Pack expectations for the study and the pre-engagement checks that should and should not be undertaken

All activity will be conducted by staff employed by the site, therefore no access arrangements are required.

Other Information to Aid Study Set-up

This details any other information that may be helpful to sponsors and participating NHS organisations in England and Wales to aid study set-up.

The applicant has indicated that they do not intend to apply for inclusion on the NIHR CRN Portfolio.

Appendix F: Focus Group Resources & Questions

Focus Group Information Sheet

Invitation:

You are being invited to take part in this focus group which is part of a larger research project investigating therapist variability within routine psychotherapy practice in the NHS. This focus group is being carried out to help inform the construction of a 'package' of development opportunities for therapists within Wellbeing Norfolk & Waveney, with the aim of supporting therapists to further develop their skills/expertise, and ultimately to help clients who access psychotherapy, even more.

What is involved:

The focus group will last approximately 60 minutes and is an informal opportunity to think about this potential package of support. You will be asked to reflect on, and discuss with the rest of the group, a number of questions related to your own clinical work and what a support/development package might consist of and look like. You will not be asked to disclose details about your clients, please make sure that anything you share protects clinical confidentiality. You are not required to answer these questions if you do not feel comfortable to do so.

Risks & Benefits:

There are no risks to taking part in this focus group, and benefits may include; time to reflect on positive therapy experiences, and the sharing of ideas with colleagues.

Use and storage of your data:

Your comments in this focus group are strictly for the use of the research team and are not related to your performance or work from a service or management perspective. Your comments will not be shared with the service outside those within the focus group.

Anonymised comments may be written into an academic report or publication. Your anonymity will be ensured in the way described in the consent information below. Please read this information carefully and, if you would like to take part in the focus group, please sign to show that you have understood the information sheet and that you consent to take part as described below.

Research Team:

This focus group is being conducted by Katy James, Lead Investigator for the project. She is supervised by the Project Research Team, based at the University of Sheffield: Professor Michael Barkham, Dr. Stephen Kellett and Dr. David Saxon. Chris Davis, Counsellor, is an informal member of the research team, providing additional analysis of the focus group material for development of the package. If you have question or concerns about this focus group, you can contact Katy James (kmjames1@sheffield.ac.uk), or the research team (m.barkham@sheffield.ac.uk).

Focus Group Participant Consent Form

- I understand that by signing below I am agreeing to take part in the University of Sheffield research described here, and that I have read and understood this information sheet.

- I understand that my participation is entirely voluntary, that I can choose not to participate in part or all of the focus group, and that I can withdraw at any stage without having to give a reason.

- I understand I can request without penalty that my comments be withdrawn from consideration in the project, any time up until the focus group feedback is analysed (July 2018).

- I understand that this focus group will be audio recorded and this will be stored in a de-identified way (e.g. using ID numbers not names), and kept separate from my consent form. The recording will be transferred from the audio recording device (the recording device will then be wiped) and stored on a password protected computer, accessible only to the Lead Investigator, until completion of analysis, at which point it will be deleted. Consent forms will be stored in a locked filing cabinet accessible only to the Lead Investigator.

- I understand that my identity will remain confidential in any written reports of this research.

- I understand that my name and data will not be shared with any third parties outside the research group outlined above.

_____

Name of Participant

_____

Signature

_____

Date

Focus Group Questions

1. Thinking of a recent piece of therapy work that you completed with a client that was a success (resulted in positive outcomes for the client including reliable improvement in their symptoms of depression and/or anxiety):

- What do you think you did (or the theory that you used) that contributed to the positive outcome? (prompt: skills you used, qualities you have that contributed)

- Do you think these things are the common features or factors for an effective therapy outcome across the other successful clinical work that you've undertaken in IAPT? If not, are there things that you do (skills you use, qualities you have) that you think are important in helping clients make improvements in therapy?

Please talk about the contribution made by clinical supervision (if any) in this outcome.

2. Thinking of a recent piece of therapy work that you completed with a client that was not a success (resulted in little improvement or termination of therapy):

- Was there anything that you think you could have done differently on reflection that would have changed this outcome?

- If you had had the chance to hone the skills and qualities you mentioned when thinking about a successful case, do you think this would have made a difference with this client?

Please talk about the contribution made by clinical supervision (if any) in this outcome.

3. This research is looking at ways that we can support therapists to hone their clinical skills and develop greater clinical expertise and competence and provide more consistent outcomes for clients across the service. This will result in a package of support/development opportunities that will then be offered to all the therapists in the service. There are a number of things from the literature that appear to be linked with more effective therapists, these are:

- Higher resilience levels

- Lower therapist stress

- Quality of the therapists learning environment

- Professional self doubt and self-compassion

- The nature of learning e.g. deliberate practice

Given these things, together with some of your reflections from the previous questions, if you were to have the opportunity to hone your clinical competency, how would you like to be supported with this and what methods would be most useful?

4. Are there any other things that contribute to your successful work with clients that you think could be considered within this package of support/development opportunities?

5. Is there anything that would concern you about undertaking development opportunities within a package as described (i.e. anything the team needs to consider that might make you not want to be involved?)?

6. Your managers are committed to supporting therapists in this potential 'development package', but is there anything you think the research team needs to consider in terms of making the package practical for you as a therapist?

Thank you for taking part.

Appendix G: Deliberate Practice Hints and Tips and Recording Form

Deliberate practice hints & tips

Identifying a micro-skill:

1. Pick something small (rather than a large area of practice). E.g. summarising at the end of a session, interrupting a client, how you open a session.

2. Pick one skill and focus on this for a few months – that way you will get a chance to really feel confident with the skill, and it also means you don't have to be coming up with a new micro-skill each month.

3. Your micro skill is something you want to develop yourself, it doesn't have to be something that other people in the group want to work on.

Structuring the deliberate practice group sessions:

1. Split the time between the number of participants and nominate a time keeper so that everyone has their time (on a rotating basis so the same person doesn't have to time each month). The time keeper should also have a slot each month.

2. Prior to the session, think about how you want to practice and what you want from the other members of the group. I.e. do you want one person to role play a client while you practice your skill? Or do you want to practice to an empty chair and have feedback from the group on your skills etc.

3. Make a note of what you are working on and how you are practicing so you can pick this up again easily at the next session.

4. Don't worry if you feel a bit silly, this is a time to play and experiment so have fun with it too.

Ideas of how to practice:

- Have one person role play your client in the specific scenario you want to practice (give specific instructions to your 'client') and repeat the microskill you are focussing on for a specific time period (e.g. 5 mins) then get feedback from your 'client' and group members.

- Go around the group (repeatedly, not just once!) trying out the specific microskill you are working on, e.g. practicing the first sentence opener of your session. That way you can learn from other people's ideas and have a chance to practice yourself.

- Listen to a short section of tape repeatedly and describe your internal emotions or cognitions whilst listening (this would be ideal for a client presentation that is particularly activating you emotionally, such as making you doubt yourself, feel irritated/hopeless etc). The section of tape would need to be short so you have a chance to repeat this quite a few times, to allow a level of habituation. (See 3min Tony Rousmaniere clip starting **from 47min in**)
https://www.youtube.com/watch?v=eR5OQ_mWros

Further reading:

Deliberate Practice for Psychotherapists: A Guide to Improving Effectiveness (2017) by Tony Rousmaniere

Deliberate Practice Recording Sheet	
Date:	**Colleagues Present (if applicable):**

Example of an issue I have noticed in session:

*Client DF – felt like I was saying too much, trying to rescue – felt a bit anxious in the session, like I had to **do** something to make him feel better.*

Description of micro skill to practice:

1. *E.g. when a client is talking about their sadness about something, giving time for the emotion rather than jumping in to fix it to make us both 'feel better' in the moment.*

How I practiced it:

E.g. role play within the deliberate practice group. Another member of the group role played a client that they are seeing that is depressed and sad and I practiced not jumping in but instead focusing on my own bodily responses, and my breath while listening. Allowing silences. Did this for 10mins then gave and received feedback to/from the group.

E.g. listened to the recording of my client talking about a sad experience x5. Focused on my own bodily sensations and breathing whilst client was talking.

Reflections or further practice plans:

E.g. I noticed that I become really tense and find it difficult to stay 'present' when someone is expressing intense sadness. I am going to experiment with my reactions to my own sad feelings as part of my own meditation practice/discuss with my therapist/be mindful of this with clients and keep practicing connecting with my breath when in this situation.

Deliberate Practice Recording Sheet

Date:	**Colleagues Present (if applicable):**

Example of an issue I have noticed in session:

Multiple clients: end of session feedback "yeah, it was fine", "it was really helpful to talk things through" - I feel frustrated, like I'm not missing something, the end of the session then feels a bit 'false' or 'fake'.

Description of micro skill to practice:

E.g. end of session feedback – I've noticed I tend to get fairly superficial responses, how can I get more depth?

How I practiced it:

E.g. asked for examples of how other colleagues ask for end of session feedback.

Practiced different wording with the group – we went round in turn trying out different wording of asking for feedback, did this repeatedly in turn for 10mins.

Reflections or further practice plans:

E.g. Rather than asking something that could invite a compliment or complaint (like " how did you find the session today"), asking something like "what are you going to take away from the session", "what are the main things you've learnt today" or more things about the relationship like "can you think of any times in the session where you felt we were not really in tune/on the same page today" or "if we could have done one thing differently/could have changed one thing in the session today that would have made it more useful, what would it have been?" – tends to give more indepth responses.

Experiment with using these more creative questions in next weeks client sessions.

Deliberate Practice Recording Sheet
Date:	**Colleagues Present (if applicable):**
Example of an issue I have noticed in session:
Description of micro skill to practice:
How I practiced it:
Reflections or further practice plans:

Appendix H: Core Therapists Full MLM

Core Therapists MLMs

The tables below show the MLMs calculated from the Core Therapists datasets for each Phase. Each table presents: the best fit model (referred to as the 'Full Model') with all significant predictor variables included; the best fit model with predictive variables only (that is, excluding any variables that would not be known at the beginning of therapy); the 'Final' model, that is the model that was used to compare phases in the main analysis; and the MCMC calculation.

Phase 1

MLM	Full Model			Full Predictive Model			Final Model			MCMC Core 1		
MLM Values	Value	S.E.	*p*	Value	S.E.	*p*	Value	S.E.	*p*	Value	S.E.	*p*
Fixed Part												
Average therapist PHQ9 Change (intercept)	6.101	0.292	0.00	6.443	0.326	0.00	6.134	0.302	0.00	6.137	0.31	0.00
First PHQ9	0.601	0.036	0.00	0.569	0.037	0.00	0.564	0.037	0.00	0.564	0.037	0.00
First Phobia	-0.142	0.031	0.00	-0.135	0.033	0.00	-0.139	0.033	0.00	-0.139	0.033	0.00
First WSAS	-0.076	0.026	0.00	-0.065	0.027	0.01	-0.065	0.027	0.02	-0.065	0.027	0.01
Interaction First PHQ9 / First WSAS	-0.007	0.003	0.02	-0.006	0.003	0.00	-0.006	0.003	0.04	-0.006	0.003	0.02
Deprivation Quintile Group				-0.95	0.401	0.02						
Number of missed sessions	-0.567	0.091	0.00									
Number of attended sessions	0.292	0.048	0.00									
Interaction First PHQ9 / Number attended sessions	0.028	0.008	0.00									
Random Part												
Level 2 (therapist) variance	1.423	0.615		1.453	0.644		1.502	0.658		1.682	0.812	
Level 1 (patient) variance	27.041	1.278		29.222	1.381		29.382	1.387		29.604	1.407	
N Therapists	35			35			35			35		
N Patients	930			930			930			930		
Estimation:	IGLS			IGLS			IGLS			MCMC		
-2*loglikelihood:	5734.487			5805.544			5811.147					
Chain Length:										14000		

Phase 2

MLM	Full Model			Full Predictive Model			Final Model			MCMC Core 2		
MLM Values	Value	S.E.	*p*	Value	S.E.	*p*	Value	S.E.	*p*	Value	S.E.	*p*
Fixed Part												
Average therapist PHQ9 Change (intercept)	6.923	0.278	0.00	6.959	0.283	0.00	6.665	0.263	0.00	6.667	0.27	0.00
First PHQ9	0.593	0.032	0.00	0.588	0.032	0.00	0.593	0.033	0.00	0.593	0.033	0.00
First Phobia							-0.05	0.028	0.08	-0.05	0.028	0.04
First WSAS	-0.119	0.02	0.00	-0.111	0.021	0.00	-0.098	0.022	0.00	-0.098	0.022	0.00
Interaction First PHQ9 / First WSAS	-0.006	0.003	0.02	-0.007	0.003	0.01	-0.006	0.003	0.01	-0.006	0.003	0.01
Referral Source Group	-0.959	0.325	0.00	-0.929	0.332	0.00						
Number attended sessions	0.214	0.039	0.00									
Number missed sessions	-0.459	0.077	0.00									
Random Part												
Level 2 (therapist) variance	1.287	0.498		1.326	0.517		1.32	0.516		1.452	0.616	
Level 1 (patient) variance	25.529	1.046		26.654	1.092		26.775	1.097		26.935	1.109	
N Therapists	35			35			35			35		
N Patients	1225			1225			1226			1226		
Estimation:	IGLS			IGLS			IGLS			MCMC		
-2*loglikelihood:	7479.364			7531.908			7543.425					
Chain Length:										21000		

Phase 3

MLM	Full Model			Full Predictive Model			Final Model			MCMC Core 3		
MLM Values	Value	S.E.	*p*	Value	S.E.	*p*	Value	S.E.	*p*	Value	S.E.	*p*
Fixed Part												
Average therapist PHQ9 Change (intercept)	6.829	0.214	0.00	6.865	0.206	0.00	6.865	0.206	0.00	6.864	0.202	0.00
First PHQ9	0.601	0.032	0.00	0.587	0.033	0.00	0.587	0.033	0.00	0.589	0.033	0.00
First Phobia	-0.084	0.027	0.00	-0.082	0.028	0.00	-0.082	0.028	0.00	-0.082	0.028	0.00
First WSAS	-0.104	0.022	0.00	-0.099	0.022	0.00	-0.099	0.022	0.00	-0.1	0.022	0.00
Interaction First PHQ9 / First WSAS	-0.011	0.003	0.00	-0.011	0.003	0.00	-0.011	0.003	0.00	-0.011	0.003	0.00
Number attended sessions	0.158	0.036	0.00									
Number missed sessions	-0.408	0.071	0.00									
Interaction First PHQ / Number attended sessions	0.017	0.006	0.00									
Random Part												
Level 2 (therapist) variance	0.611	0.319		0.475	0.291		0.475	0.291		0.425	0.359	
Level 1 (patient) variance	24.061	0.989		25.22	1.036		25.22	1.036		25.452	1.057	
N Therapists	35			35			35			35		
N Patients	1217			1217			1217			1217		
Estimation:	IGLS			IGLS			IGLS			MCMC		
-2*loglikelihood:	7345.805			7398.754			7398.754					
Chain Length:										100100		

Appendix I: Core Therapists Tests of Model Assumptions

The figures below show the tests of model assumptions for the Core Therapists models. Model assumptions were tested for both levels by visual inspection of quantile-quantile (q-q) plots to test for normality and plotting residuals across variable values to test for homoscedasticity. The tests of scedasticity are presented first, following by the q-q plots.

It should be noted that the tests of scedasticity show a fanning out pattern at the patient level. This indicates that there is more variability between patients in the PHQ Change scores when intake scores are higher. It could also be influenced a small number of patients with very small amounts of pre-post change. This was considered within the analysis, however the benefit of transforming the data or modelling non-linear relationships at the patient level, given the focus of the thesis on therapist effects, was considered to be minimal, and was not therefore undertaken.

Tests of Scedasticity
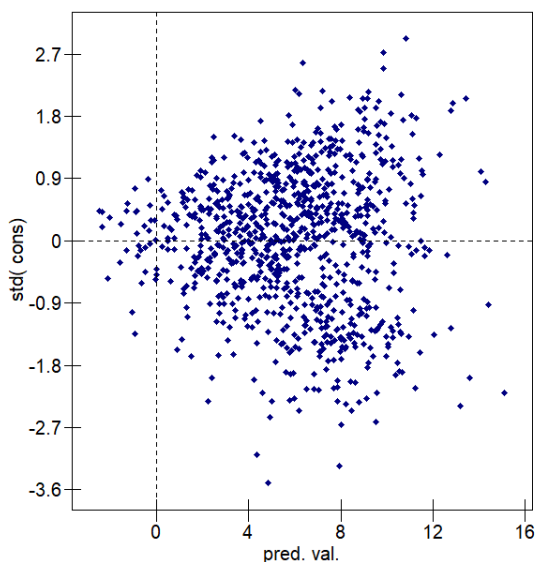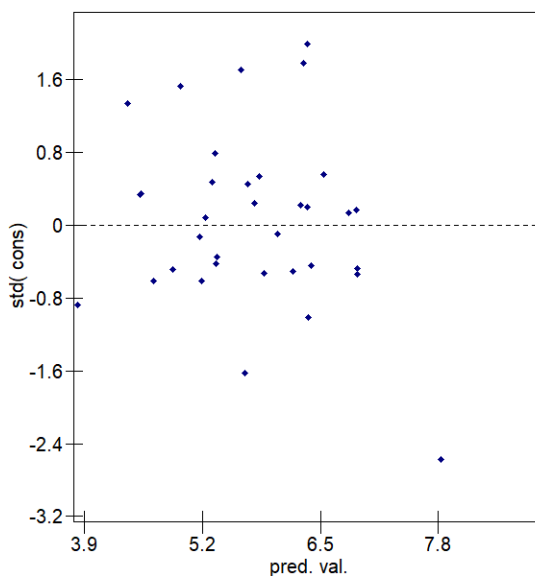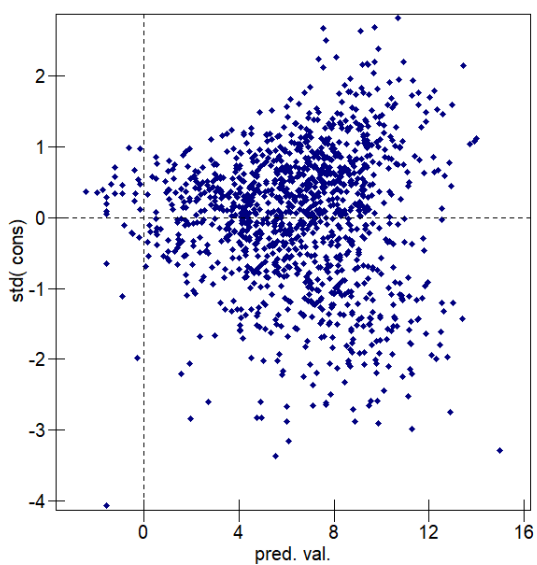
Phase 1 patients

Phase 1 therapists

Phase 2 patients
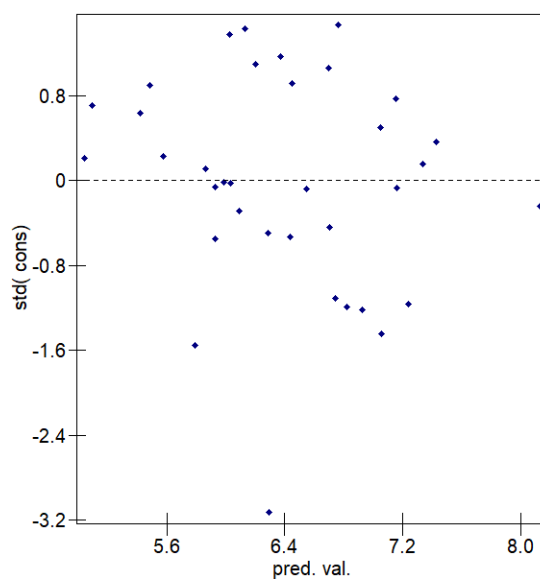
Phase 2 therapists
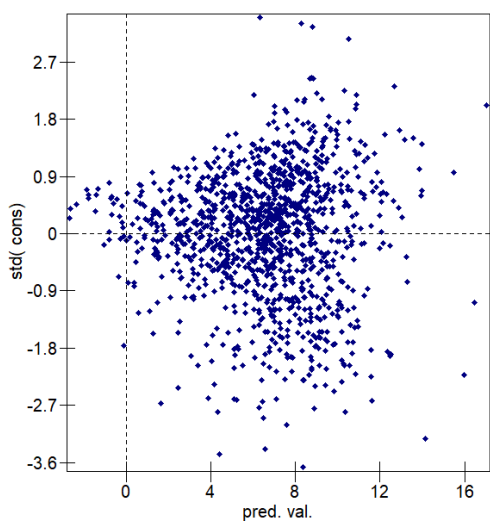
Phase 3 patients
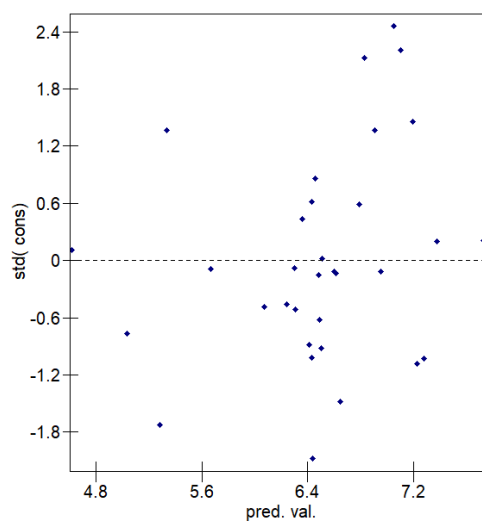
Phase 3 therapists

QQ Plots
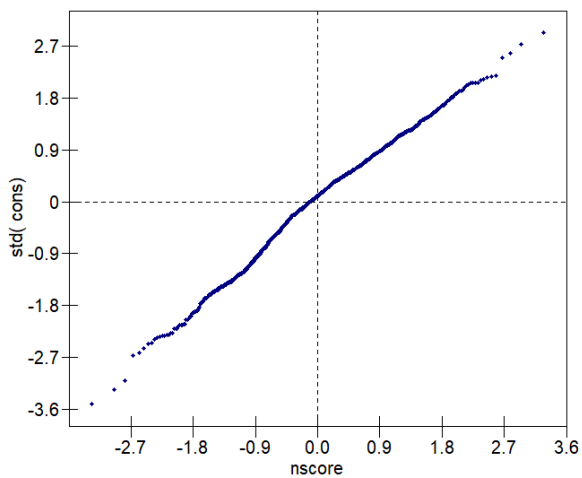
Phase 1 patients
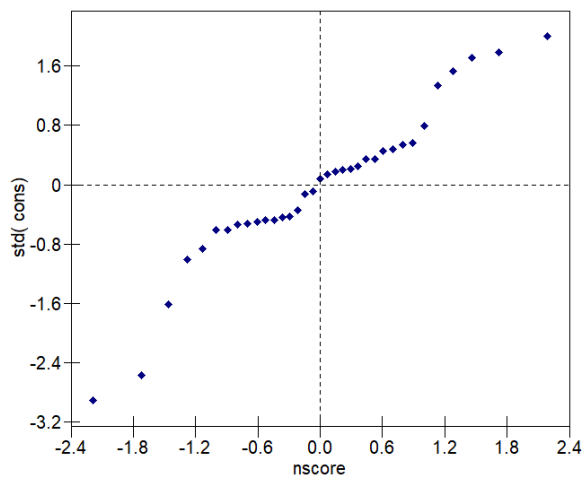
Phase 1 therapists
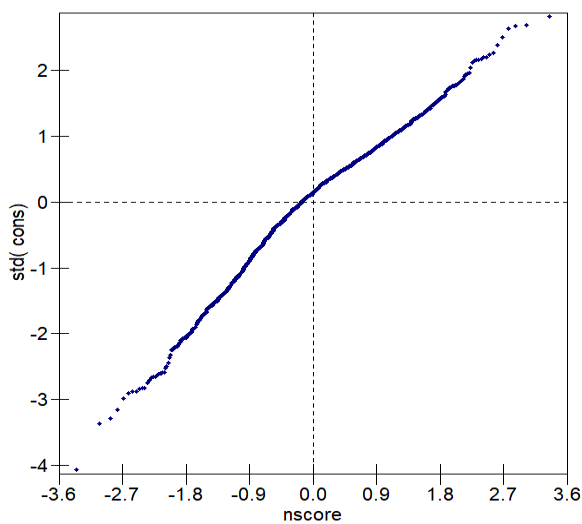
Phase 2 patients
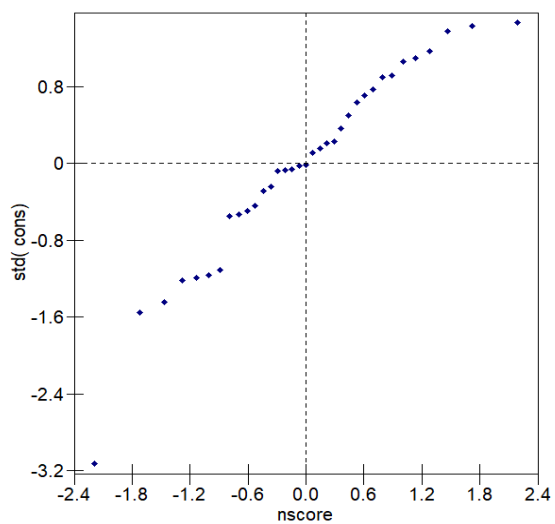
Phase 2 therapists

Phase 3 patients
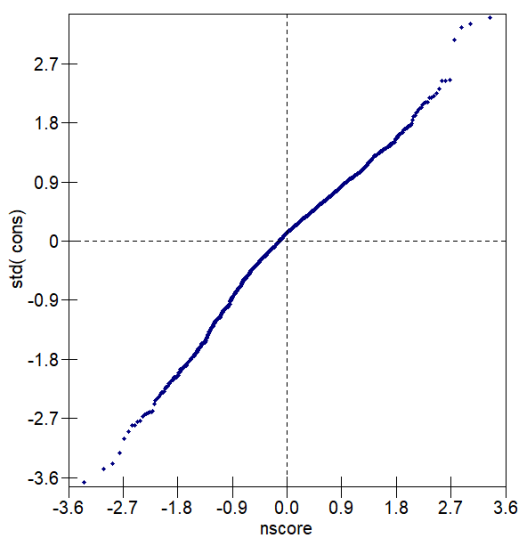
Phase 3 therapists
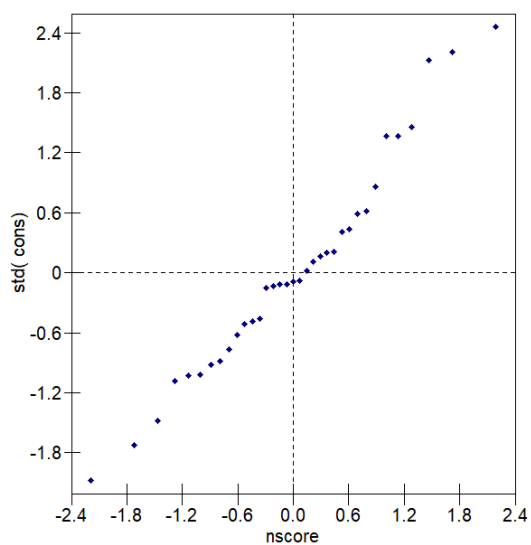
Appendix J: Core Therapists Sensitivity Tests

Sensitivity test 1: Core Therapists MLMs with trainee therapists removed

The following figures present the MLWin output for the sensitivity test conducted on the All Therapists dataset, removing the trainee therapists from the sample and repeating the primary MLMs. The therapist effect for each model has been calculated and reported for each phase.

Phase 1: Therapist effect = 4.3%

$$\text{PHQ Change}_{ij} = \beta_{0j} + 0.551(0.035)(\text{First PHQ9-gm})_{ij} +$$
$$-0.077(0.025)(\text{First WSAS-gm})_{ij} +$$
$$-0.123(0.030)(\text{First Phobia-gm})_{ij} +$$
$$-0.007(0.003)(\text{First PHQ9-gm}).(\text{First WSAS-gm})_{ij} + e_{ij}$$
$$\beta_{0j} = 6.153(0.265) + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \quad \sigma_{u0}^2 = 1.379(0.586)$$
$$e_{ij} \sim N(0, \sigma_{e}^2) \quad \sigma_{e}^2 = 30.708(1.335)$$
-2*loglikelihood = 6945.728(1104 of 1104 cases in use)
UNITS:
  Therapist Code: 53 (of 53) in use

Phase 2: Therapist effect = 5.9%

$$\text{PHQ Change}_{ij} = \beta_{0j} + 0.569(0.031)(\text{First PhQ9-gm})_{ij} +$$
$$-0.045(0.026)(\text{First Phobia-gm})_{ij} +$$
$$-0.089(0.021)(\text{First Wsas-gm})_{ij} +$$
$$-0.005(0.002)(\text{First PhQ9-gm}).(\text{First Wsas-gm})_{ij} + e_{ij}$$
$$\beta_{0j} = 6.395(0.253) + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \quad \sigma_{u0}^2 = 1.753(0.571)$$
$$e_{ij} \sim N(0, \sigma_{e}^2) \quad \sigma_{e}^2 = 28.128(1.054)$$
-2*loglikelihood = 9138.236(1472 of 1472 cases in use)
UNITS:
  Therapist Code: 53 (of 53) in use

Phase 3: Therapist effect 3.2%

$$\text{PHQ Change}_{ij} = \beta_{0j} + 0.594(0.031)(\text{First PHQ9-gm})_{ij} +$$
$$-0.099(0.021)(\text{First WSAS-gm})_{ij} +$$
$$-0.089(0.026)(\text{First Phobia-gm})_{ij} +$$
$$-0.010(0.003)(\text{First PHQ9-gm}).(\text{First WSAS-gm})_{ij} + e_{ij}$$
$$\beta_{0j} = 6.439(0.216) + u_{0j}$$

$$u_{0j} \sim N(0, \sigma^2_{u0}) \quad \sigma^2_{u0} = 0.883(0.378)$$
$$e_{ij} \sim N(0, \sigma^2_e) \quad \sigma^2_e = 26.454(1.011)$$
$$-2*loglikelihood = 8662.805(1412 \text{ of } 1412 \text{ cases in use})$$

UNITS:

        Therapist ID new: 53 (of 53) in use

Sensitivity test 2: Expanded dataset

As with the primary analysis, MCMC estimations were used on the expanded datasets. The following figure shows the MCMC for the expanded dataset. As with the MCMC for the main analysis, the PrI's for the sensitivity test of expanded datasets overlapped so the differences based on these estimations cannot be considered significant.

MCMC Therapist Effect (with 95% Probability Intervals) for Sensitivity Datasets Showing Each Study Phase
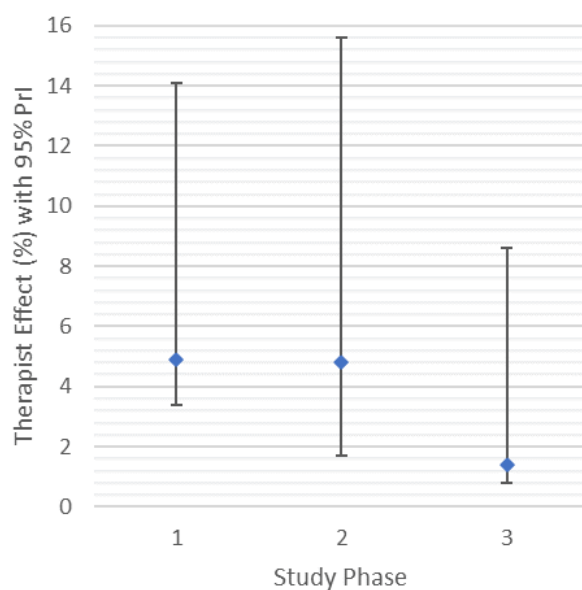
Appendix K: All Therapists Full MLM

The tables below show the MLMs calculated from the All Therapists datasets for each Phase. Each table presents: the best fit model (referred to as the 'Full Model') with all significant predictor variables included; the best fit model with predictive variables only (that is, excluding any variables that would not be known at the beginning of therapy); the 'Final' model, that is the model that was used to compare phases in the main analysis; and the MCMC calculation.

Phase 1

MLM	Full Model			Full Model Predictive			Final Model			MCMC model		
MLM Values	Value	S.E.	*p*	Value	S.E.	*p*	Value	S.E.	*p*	Value	S.E.	*p*
					Fixed Part							
Average therapist PHQ9 Change (intercept)	6.253	0.214	*0.00*	6.3	0.224	*0.00*	5.975	0.203	*0.00*	5.979	0.21	*0.00*
First PHQ9	0.556	0.03	*0.00*	0.538	0.03	*0.00*	0.532	0.026	*0.00*	0.532	0.026	*0.00*
First Phobia	-0.073	0.022	*0.00*	-0.078	0.022	*0.00*	-0.09	0.022	*0.00*	-0.09	0.022	*0.00*
First WSAS	-0.091	0.018	*0.00*	-0.08	0.018	*0.00*	-0.077	0.018	*0.00*	-0.077	0.018	*0.00*
Interaction First PHQ9 / First WSAS	-0.005	0.002	*0.01*	-0.004	0.002	*0.03*	-0.004	0.002	*0.04*	-0.004	0.002	*0.02*
Deprivation Quintile Group	-0.808	0.272	*0.00*	-1.003	0.279	*0.00*						
Number of attended sessions	0.269	0.032	*0.00*									
Number of missed sessions	-0.428	0.063	*0.00*									
Interaction First PHQ9 / Number attended sessions	0.017	0.005	*0.00*									
					Random Part							
Level 2 (therapist) variance	1.422	0.42		1.629	0.464		1.438	0.45		1.5	0.493	
Random Slope on First PHQ9	0.019	0.008		0.018	0.008							

Level 1 (patient) variance	28.8	0.943	30.499	0.998	31.49	1.02	31.61	1.03
N Therapists	81		81		81		81	
N Patients	1982		1982		1982		1982	
Estimation:	IGLS		IGLS		IGLS		MCMC	
-2*loglikelihood:	12351.164		12462.31		12519.29			
Chain Length:							15500	

Phase 2

MLM	Full Model			Full Predictive Model			Final Model			MCMC		
MLM Values	Value	S.E.	*p*	Value	S.E.	*p*	Value	S.E.	*p*	Value	S.E.	*p*
						Fixed Part						
Average therapist PHQ9 Change (intercept)	6.748	0.243	*0.00*	6.378	0.234	*0.00*	6.035	0.213	*0.00*	6.036	0.216	*0.00*
First PHQ9	0.576	0.025	*0.00*	0.563	0.026	*0.00*	0.559	0.026	*0.00*	0.558	0.026	*0.00*
Random Slope on First PHQ9	-0.007	0.003	*0.03*	-0.007	0.003	*0.02*						
First WSAS	-0.117	0.017	*0.00*	-0.109	0.017	*0.00*	-0.104	0.017	*0.00*	-0.104	0.017	*0.00*
Random Slope on First WSAS	-0.004	0.001	*0.00*	-0.004	0.001	*0.00*						
First Phobia	-0.062	0.021	*0.00*	-0.059	0.022	*0.01*	-0.06	0.022	*0.01*	-0.06	0.022	*0.00*
Number attended sessions	0.403	0.039	*0.00*									
Random Slope on Number attended sessions	-0.026	0.005	*0.00*									
Number missed sessions	-0.37	0.057	*0.00*									
Interaction First PHQ9 / Number attended sessions	0.021	0.005	*0.00*									
Interaction First PHQ9 / First WSAS							-0.007	0.002	0.00	-0.007	0.002	0.00

	Random Part							
Level 2 (therapist) variance	0.813	0.281	0.946	0.31	0.946	0.31	0.964	0.358
Level 1 (patient) variance	25.668	0.779	27.059	0.816	27.059	0.816	27.162	0.822
N Therapists	74		74		74		74	
N Patients	2267		2267		2267		2267	
Estimation:	IGLS		IGLS		IGLS		MCMC	
-2*loglikelihood:	13838.617		13961.72		13961.72			
Chain Length:							31000	

Appendix L: All Therapists Tests of Model Assumptions

The figures below show the tests of model assumptions for the All Therapists models. Model assumptions were tested for both levels by visual inspection of quantile-quantile (q-q) plots to test for normality and plotting residuals across variable values to test for homoscedasticity. The tests of scedasticity are presented first, following by the q-q plots. See previous comments re implications of patient level tests of scedasticity.

Tests of Scedasticity

Phase 1 patients

Phase 1 therapists

Phase 2 patients

Phase 2 therapists

Phase 3 patients

Phase 3 therapists

QQ Plots

Phase 1 patients

Phase 1 therapists

Phase 2 patients

Phase 2 therapists

Phase 3 patients

Phase 3 therapists