Reviewing the evidence for NICE recommended psychotherapies for depression

Kofi Kramo

Student number: 09274982

A thesis submitted in partial fulfilment of the requirements of the University of Hertfordshire for the degree of Doctor of Clinical Psychology

December 2016

Contents

1. Abstract

Introduction: Depression is a common mental health problem affecting 1 in 6 people in the UK, which represents a considerable burden (Lépine and Briley, 2011). Cognitive behavioural therapy (CBT) is recommended by the National Institute for Heath and Care Excellence (NICE) as first line treatment for depression on the basis of the evidence available in the depression guidelines (NCCMH, 2010). However, there remains controversy and debate about the strength of the evidence supporting CBT, the relative efficacy of psychotherapies and concerns about the current evidence-based practice paradigm used by NICE to recommend psychotherapies. The current project aimed to examine the secondary evidence base for CBT and other psychological therapies recommended by NICE for the treatment and management of adult depression.

Methods: An exploratory data analysis was conducted to assess the strength of the evidence, using meta-analytic outcomes as units of analysis (effect sizes and risk associations) for CBT and other psychotherapies. Further analysis examined the relationship between different evidence characteristics (including treatment comparators, study quality etc.) and meta-analytic outcomes for the two psychotherapy groups.

Results: The analyses revealed significant differences in the overall mean effect sizes for CBT and other psychotherapies, which consisted of larger mean effects within the other psychotherapy group. However, the evidence base within the CBT group was stronger than the other psychotherapies group when the two groups were compared to medication comparators. Furthermore, significant relationships were found between the psychotherapy effects and evidence quality, suggesting that greater amounts of low quality evidence associated with favourable effects for CBT.

Implications: The findings in this review question the strength of the evidence base for CBT as a front line psychological intervention for the treatment of depression, particularly when considered against the collective evidence for other psychotherapies. The findings highlight how guideline evidence used to recommend psychological treatments of depression are constructed to fit within a medical context and the impact that this has on the choice of psychotherapies available to clients and practitioners are considered.

2. Introduction

2.1 Chapter outline

This chapter will introduce the topics central to this project. First, it is necessary to place this project firmly within the social constructionist stance within which this research has been conducted. This will involve a brief introduction to social constructionism and the type of social constructionist ideas that have informed my thinking and provides a critical thread on the topics that follow. The chapter then provides a brief introduction to psychotherapy in the context of the National Health Service (NHS). An overview of evidence-based practice will be covered, considering its implications for psychotherapy. I will then outline clinical practice guidelines as they apply to psychotherapy and consider the main areas of controversies and debates. Finally, I will state the aims of this project and the rationale for the current project.

2.2 A social constructionist stance taken within this research project

Social constructionism refers to a theory of knowledge that examines how individuals and groups mutually construct understandings of the world that form the basis for shared assumptions about reality. Although the theory is fundamental to sociology, its application to the field of clinical psychology has been relatively more recent. Social constructionism is a broad church (Locke and Strong, 2010) that draws on a range of disciplines. On this basis Burr (2003) asserts that it is helpful to consider social constructionist approaches as sharing one or more of its four broad tenets: 1) a critical stance towards taken-for-granted knowledge; 2) the historical and cultural specificity of knowledge; 3) meaning is sustained through social interactions of discourse and other symbolic forms; and 4) knowledge and social action go together rather than the former informing the latter.

In addition to these characteristics, Locke and Strong emphasise the uneasy relationship between social constructionist approaches in psychology and essentialism, i.e. the notion that one of the major goals of psychology is to uncover the essential characteristics of people. This latter tenet can be extended further to traditional scientific approaches to psychotherapy research, particularly those exercised within dominant evidence paradigms. A review of this type of evidence is the focus of this study. Locke and Strong emphasise a political component in psychology in relation to the 'facts' not being neutral and awaiting discovery but rather constructed in fields of activities, and assembled ideologies that benefit some at the expense of others. A major focus of social constructionism is the manner in which individuals and groups interact to construct a perceived social reality. It involves exploring how social phenomena are created, institutionalized, understood, and made into traditions. The ideas that social constructionism embodies have an intuitive appeal to me. It is through a social constructionist lens that this research project has been undertaken.

The social constructionist stance adopted within this study is similar to what Danziger (1997) described as 'light' social constructionism: a body of thought that emphasises the ongoing construction of meaning in current discursive practices. This strand of social constructionism places emphasis on the dependence of current patterns of interaction on rigid power structures, which have historical foundations afforded protection through institutionalised practices and textual conventions. According to Locke and Strong its roots lie within traditional and pragmatic concerns that stem from empirical traditions, whilst avoiding much of the philosophical 'quicksand' that accompanies other more epistemologically challenging, albeit important, strands of social constructionism.

2.3 Psychotherapy

Defining psychotherapy is a difficult and daunting task (Andersson and Cuijpers, 2009; Wampold, 2001). A well-known, often cited definition of psychotherapy is provided by Strupp (1978), who describes it as "an interpersonal process designed to bring about modifications of feelings, cognitions, attitudes and behaviours that have proved troublesome to the person seeking help from a trained professional" (p.3). Roth and Fonagy (2005) elaborated on this definition, suggesting that psychotherapy contains three key components: the therapist-patient relationship, an interpersonal context, and a theoretical model that guides the therapist's action, which in turn generates procedures for relieving distress.

Another helpful and uncontroversial definition is provided by Wampold (2001), who defines psychotherapy as primarily "an interpersonal treatment based on psychological principles and involves a trained therapist and a client who has a mental disorder, problem or complaint" (p. 3). Further, Wampold emphasises that psychotherapy is intended to be remedial and is adapted for a client's individual's problem or circumstances.

These broad basic definitions accommodate the inherent complexity that psychotherapy is as an enterprise and acknowledge the interpersonal processes, diverse models and theories that underpin them.

2.3.1 Psychotherapy in the context of the National Health Service

There has been a proliferation of psychotherapy over the past 60 or so years. Some estimates have placed this at over 400 different brands (Karasu, 1986). However, these are mostly subclasses of seven major therapeutic orientations (DoH, 2001; Roth and Fonagy, 2013), which consist of psychodynamic psychotherapy; behavioural and cognitive-behavioural therapy; interpersonal psychotherapy;

strategic or systemic psychotherapies; supportive and experiential psychotherapies (humanistic); and counselling.

In addition to traditional orientations, there are integrative psychotherapies that have achieved successful integration of two or more major theoretical orientations in order to address complex problems and patients encountered in routine clinical practice. Cognitive behaviour therapy (CBT) represents the most established of the integrative therapies and cognitive analytic therapy (Ryle, 1982) is also a commonly established therapy. Moreover, other psychotherapies have developed that differ from traditional theoretical orientations through being rooted within postmodern epistemologies that emerged within the second half of the twentieth century and thus sit outside of these traditional theoretical orientations such as narrative therapy (White & Epston, 1990). Parry, Roth and Fonagy (2005) argue that continual theoretical development makes traditional demarcations of psychotherapy less clear as theoretical orientations tend to incorporate the strengths of other theoretical orientations in order to strengthen and refine their approaches.

Cognitive behavioural therapies (CBT), psychodynamic psychotherapies and systemic therapies are most widely practiced within the NHS (Department of Health, 2001). Integrative and eclectic approaches are common in everyday clinical practice due to complexity of presentations, comorbidity and the chronic nature of conditions seen in routine clinical practice (Roth and Fonagy, 2013). These therapies can be described as individually tailored and formulation driven with greater emphasis placed on the 'non-specific' aspects of therapy rather than specific techniques.

2.3.2 Psychotherapy outcome research: Principal methods and findings

Assessing outcome has been a primary concern throughout the development of traditional psychotherapy orientations (see Constonguay & Association, 2010). The

basic methodological approach to this involves comparing patient scores on an indicator of wellbeing, which could be a measure of behavioural or subjective experiences, before and after therapy. Such research has repeatedly found, across different mental health difficulties, that people are better off after receiving psychotherapy (Cooper, 2008). However, more sophisticated procedures involving randomisation of patients to experimental or control groups are required in order to determine treatment efficacy and effectiveness[1, 2]. The use of these procedures to assess the benefits of intervention are known as randomised controlled trials (RCT) and are considered the 'gold standard' of research (these are considered in more detail later in this chapter; see section 2.4.2).

The past forty years has seen the development of meta-analysis (Smith and Glass, 1977; Smith, Glass and Miller, 1980) as a technique enabling psychotherapy researchers to combine the findings from numerous psychotherapy studies using standardised measures of the size of the relationship between psychotherapy and outcome, which is known as an effect size (e.g. Cohen's *d*). Meta-analysis is considered the most reliable evidence due its ability to draw from an extensive body of data. The procedure has been applied to assess the general efficacy of psychotherapy, the efficacy of specific interventions (classes of interventions or techniques), and to assess the differential effects of psychotherapies.

There is an overwhelming body of evidence that, in general, psychotherapy is efficacious and has a positive impact on people's mental health and it had been estimated that 79 percent of people who receive psychotherapy improve to a

[1] The control group could consist of patients who have not undergone the therapeutic procedures being tested. This either consists of 'no treatment' group, usually on a waiting-list and in receipt of treatment as usual or an active control group with characteristics similar to the 'experimental group' but who do not receive the procedures being tested.

[2] There is an important distinction between the terms treatment 'efficacy' and 'effectiveness'. The former refers to whether psychotherapy is responsible for the desired outcome. The latter refers to a treatment's ability to bring about a desired outcome when used under circumstances more reflective of clinical practice.

greater extent than the average person who does not (Cooper, 2008; Lambert, Bergin and Garfield, 2013a; Wampold, 2001). Overall, the average effect of psychotherapy is estimated to be large and more effective than a number of 'evidence-based' medical procedures routinely used in healthcare (Wampold, 2007). In respect to clinical change, Hansen and colleagues' (2002) review of twenty-eight clinical trials reported that 60 percent of participants in psychotherapy made clinically significant improvements[3]. Moreover, Shadish and colleagues' (2000) extensive meta-analyses of 'real-world' data found psychotherapy in clinically representative patient populations to be no less effective than under more controlled conditions.

Research into rate of therapeutic change (also referred to as 'dose-effect' relationships) has found that sizeable proportions of participants in therapy improve after 10 sessions and 75 percent by 50 sessions (Hansen et al., 2002; Lambert and Ogles, 2004). However, this varies based on the problems and symptoms, with changes on acute and symptomatic problems occuring more quickly than changes on characterological and personality-based problems (see Kopta, Howard, Lowry and Beutler, 1994).

Psychotherapies are generally as effective as medication for psychological distress and seem to have lower rates of relapse and drop-out rates (Cooper, 2008). Findings from other research methods provide further support for the positive outcomes of psychotherapy. For example, health economic modelling has demonstrated psychotherapy to be cost effective in respect to making substantial reductions to the utilisation of medical care (e.g. Chiles et al., 1999). Furthermore, survey research of patient experiences (e.g. Consumer Reports, 1995) indicates that most clients found psychotherapy beneficial. Taken together these findings present a generally positive picture of the current status of psychotherapies' absolute efficacy and effectiveness.

[3] Clinically significant change is conceptualised as patients moving from high levels of psychological distress to what would be considered within normal range (Cooper, 2008).

2.3.3 Differential outcomes: the dodo bird verdict and specific vs common factors

Outcome research into whether or not some forms of psychotherapy are more efficacious than others has considerable practical importance for deciding the most appropriate form of psychological therapy to offer individuals with a particular problem (Cooper, 2008). RCTs and meta-analyses are conducted to answer questions of relative (or differential) efficacy and effectiveness of psychotherapies and this is a rigorously contested issue within psychotherapy research. On one side of the debate are researchers, mainly of cognitive-behavioural orientation, who argue that cognitive-behavioural interventions are more efficacious than other psychotherapies (e.g. Chambless, 2002). On the other side are psychotherapy researchers, usually from non-CBT orientations, who argue that different classes of psychotherapy are broadly equivalent in their effects (e.g. Luborksy, Rosenthal, Digeur et al, 2002; Wampold, 2001).

The body of evidence from both comparative outcome studies at primary and meta-analytic level consistently finds there to be little differences in the efficacy and effectiveness between different psychological approaches (e.g. Luborsky et al., 2002; Wampold, 2001). This is more so when all of the therapies compared are bona fide psychotherapies (Wampold, 1997). Meta-analyses have shown the average difference in effect between diverse psychotherapy orientations to be small or less (i.e Cohen's d = 0.2 or below). Moreover, further reductions in effect size have been estimated when research allegiance effects (e.g. Luborsky et al., 1999; Robinson et al., 1990) and other biasing variables are controlled for (Wampold, 2001).

The finding of general equivalence between different psychotherapy approaches is well known within the psychotherapy research literature as the 'dodo bird' verdict (Luborsky et al., 1975; Rosenzweig, 1934), after the character in *Alice in Wonderland* who declared 'everybody has won and all must have prizes'. The dodo bird verdict

has been strongly adopted by advocates of the common factors perspective (Lambert, 2013). Despite criticisms of the dodo bird verdict (e.g. Beutler et al. 2002; Chambless, 2002), other lines of research, namely dismantling studies, converge with the dodo bird verdict to add further support for the common factors perspective (e.g. Duncan, 2010) and studies of therapist effects (e.g. Crits-Christoph & Mintz, 1991; Crits-Christoph et al., 1991). Moreover, Lambert (2013b) argues that when findings from meta-analytic reviews consider outcomes of patients with specific disorders, these also point towards general equivalence of psychotherapies.

Specific and common factors perspectives also represents two major, and often competing, explanatory paradigms of how psychotherapy works. Advocates of the specific factors view psychotherapy's mechanism of change and the specific techniques associated with this, as responsible for therapeutic outcomes. For example, cognitive-behavioural interventions are explicitly derived from causal models, which is often not the focus of other theoretical orientations. Failure to demonstrate superiority risks undermining the causal models as rationales for therapeutic intervention (Budd and Hughes, 2009). The common factors perspective argues that the elements that diverse psychotherapies share, in addition to a treatment procedure (such as therapeutic alliance, exposure to anxiety-provoking situations, an explanation for an individual's distress), are responsible for therapeutic change (e.g. Frank and Frank, 1993; Wampold, 2001). Both perspectives view the other as necessary but not sufficient for therapeutic change and within the major methods of RCTs and meta-analysis various empirical strategies are associated with each (see Stiles, Shapiro and Elliot, 1986).

Gordon Paul's question 'what treatment, by whom, is most effective for this individual with that specific problem, and under which set of circumstances?' (1967, p. 111), illustrates the scale and complexity of psychotherapy research and the number of factors that are relevant to therapeutic outcomes. These undoubtedly

include participant factors (i.e. client and therapist), relationship factors and technique factors, which represent other important areas of psychotherapy research (e.g. Castonguay and Beutler, 2006) that will further improve our understanding of both specific and common factors relevant to psychotherapy outcome. However, Westen, Novotny and Thompson (2004) noted that the specific factors perspective has shifted from a purely theoretical perspective to one closely aligned to the treatment guidelines and manuals that emphasize a model of evidence-based psychotherapy based upon 'empirically supported' therapies (e.g. Barlow, 2004; Chambless and Hollon, 1998). The movement of this perspective towards practice by statute (Miller et al., 2013), has enabled it to take centre stage in healthcare policy at the exclusion of the alternative perspectives. These issues are discussed further below.

2.4 Evidence-Based Practice (EBP) and Psychotherapy

Within the National Health Service (NHS), evidence-based practice (EBP) is the standard for clinical work. EBP is defined as "an approach to decision making in which the clinician uses the best evidence available, in consultation with the patient, to decide upon the option which suits that patient best" (Gray, 2001, p.17). EBP is a derivative of evidence-based medicine (EBM; Sackett, 1996) and the term is more suited to wider disciplines including clinical psychology and mental health (Reynolds, 2000). EBP has spread into various branches of healthcare to such an extent that it has become a research paradigm (Stiwne and Abrandt, 2004).

2.4.1 The emergence of EBP in healthcare

Trinder (2000) argued that evidence-based practice is a product of its time that connects to the social, political and economic factors that coincided with advances in

information technology and the development of the internet that have enabled dissemination of research findings in a global community. Trinder attributes EBP's rapid growth in healthcare to its simple message that "practice should be based on the most up to date, reliable and valid research evidence" (p. 3).

Evidence-based practice within psychotherapy was further driven by economic pressures and the need to contain costs and to ensure clinical effectiveness within the NHS system (Parry, 1996). This was consistent with the political climate within wider healthcare of the 1990s, namely the introduction of clinical and health outcome measures to achieve improvements in the quality and delivery of care and to set professionally agreed standards for clinical care (e.g. DoH, 1997).

Within the 1990s organisational structures were established with the specific purpose of reviewing evidence. An important development was the Cochrane Collaboration (www.cochrane.org) for systematic reviews. Perhaps the most notable development was the creation of the National Institute of Health and Care Excellence (NICE), a nationally co-ordinated governmental programme, professionally underpinned, with the remit of producing authoritative guidance to the NHS.

The impact of the EBP movement on psychotherapy is significant when one considers its broader application to the contexts of healthcare policy, service commissioning, and research funding and academic training programmes. However, it remains unclear as to the extent that EBP has been able to penetrate clinical practice at 'street level' (Miller et al., 2013). This connects to the challenge that it faces in ensuring that its evidence is relevant to clinical practice (Barkham and Parry, 2008).

2.4.2 Evidence within the EBP paradigm

Within the EBP paradigm the randomised controlled trial (RCT) is considered the 'gold-standard' method to address the efficacy of therapeutic interventions (Pearce, Raman and Turner, 2015). A hierarchy of evidence is generally adopted based around RCTs, where systematic reviews and meta-analysis are considered the strongest evidence, followed by individual RCTs with definitive results and RCTs with non-definitive results. These are followed by cohort studies, case-control studies, cross sectional surveys and case reports. RCTs provide clarity around the impact of a variable of interest (e.g. class of psychotherapy as a treatment condition or as specific techniques within psychotherapy) by controlling other simultaneous variables that are operating. Thus, the strength of RCTs arguably lies within the simplicity they create within the psychotherapy research context (Lambert, Bergin and Garfield, 2004).

RCTs have broader uses within psychotherapy research (Behar and Borkovec, 2003) such as testing basic theories about change processes, mediators and moderators of outcome or for testing specific therapy techniques (as opposed to testing entire psychotherapy packages). Within each of these broader areas, RCTs have their relative merits and areas of considerations distinct to each empirical aim. However, when used in the evaluation of an entire psychotherapy package, RCTs encounter numerous challenges and internal sources of bias. A major challenge involves balancing internal consistency (e.g. selection and measurement of outcome, construct and statistical validity, etc.) and external validity (e.g. inclusion and exclusion of patient populations and clinical problems, treatment delivery, etc.). Attempts to control bias in any one particular area require compromises in others.

The use of RCTs within the EBP paradigm is open to criticism on the very same grounds as those that provide its purported strengths. Westen, Novotny and Thompson-Brenner (2004) highlight a range of conceptual weaknesses inherent

within the use of RCTs to validate psychotherapies. Westen et al. argue that in this context RCTs require a set of additional assumptions that generally are neither well validated nor broadly applicable to most disorders and treatments. This includes the assumption that RCTs provide the gold-standard for assessing treatment efficacy, which creates a broader bias based on the selective study of certain kinds of treatment at the expense of others that cannot be tested using this empirical method. They argue that this creates a situation of only being able to draw conclusions about the treatment selected within the confines of the "(small) universe of treatments that have received empirical attention" (p. 640).

Nilsson (2010) argues the basic assumptions of EBP paradigm are foreign to some forms of traditional psychotherapies such as longer-term psychodynamic psychotherapy, where it is more difficult to isolate the specific psychotherapeutic techniques from individual, therapist and therapeutic alliance factors that are likely to be treated as confounds within RCTs. As a result certain psychotherapies are less compatible with these types of evaluation. Such psychotherapies and their appropriate method of evaluation are precluded by the EBP paradigm which is predominant in healthcare.

2.4.3 Empirically supported treatments

Certain parallels can be drawn between the EBP in the UK and the empirically supported therapies (EST) approach in the United States. The EST movement developed in response to the pressures for psychotherapy to demonstrate its efficacy within a backdrop of both economic pressures and powerful influence and resources of the pharmaceutical industry (Chambless and Hollin, 1998). ESTs had an explicit focus on the promotion of treatments for specific mental disorders that were based on high quality evidence. This went a considerable way towards increasing

the status of psychological interventions in healthcare and in ensuring that resources continued to be directed towards psychotherapy services.

However, the EST movement has been controversial within the wider psychotherapy community. The EST approach implies that psychotherapies listed for a specific disorder contains within their specific treatment content specific mechanisms of change (Laska, Gurman and Wampold, 2014) that can be implemented as if they are drugs (Parry, Cape and Pilling, 2003). Laska et al. argue that widespread adoption of the EST approach has become a paradigm that dictates how psychotherapy research applied to healthcare is conducted. Thus, RCTs of whole treatment packages thought to contain specific therapeutic ingredients for specific disorders are given priority of evidence.

Miller et al. (2013) argue that with few exceptions the EBP paradigm within the UK context has equated to lists of specific treatments for specific disorders similar to the EST approach. This is contrary to the aims of EBP within the NHS context. Parry et al. argued that in the UK the EST paradigm was deliberately eschewed, in part due to the potential for this type of research evidence to mislead. Moreover, they cite its failure to take into account common factors or the strong evidence that specific factors account for a relatively small amount of variance in outcome (Norcross, 2002; Wampold, 2001). Furthermore, Parry et al. cautioned that the temptation to move in this direction is strong due to the parallels between the use of RCTs in psychotherapy and medicine. It would appear that such temptations have been realised through the evaluation of psychotherapy within a multi-modal EBP and guideline development context, which has resulted in sharper contrast of psychotherapy with pharmacological as well as social and organisational approaches to mental healthcare.

2.5 National Institute for Health and Clinical Excellence (NICE) and clinical practice guidelines

Clinical practice guidelines are defined as 'systematically developed statements to guide decisions about appropriate health care for specific and clinical purposes' (Field and Lohr, 1990, p. 38). The role of clinical guidelines is to promote clinically and cost effective care based on the best available evidence. Within the United Kingdom the government sanctioned body responsible for development of clinical guidelines is the National Institute for Health and Care Excellence (NICE). Since its inception, NICE has gained a strong reputation for the publication of evidence-based guidelines to inform clinicians on the most cost-effective and clinically efficacious interventions (Williams, 2015).

NICE guidelines are distinct in their authority and considerable weight of influence. According to Pilling (2008) clinical guidelines are the most complete manifestation of the evidence-based medicine movement. In this respect guideline development programmes provide a vehicle for evidence-base practice within healthcare. Since their establishment NICE has, in a sense, taken on the baton from the EBP paradigm in ensuring clinical practice is ‘evidence-based’.

Guideline recommendations set standards for interventions for health care professionals, which should guide their clinical behaviour. Although NICE (2007) acknowledges that clinical guidelines are not a substitute for clinical judgement in determining the most effective care for an individual, guidelines undoubtedly shape treatment options available to patients and practitioners.

2.5.1 The status of psychological interventions within NICE guidelines

Psychological interventions have been increasingly included in medically oriented guidelines (Parry, Roth and Fonagy, 2013) and are central to the treatment

recommendations for the majority of mental health conditions (Pilling, 2008). This has wide ranging implications for psychological interventions in terms of wider health policy that determines resource allocation and referral pathways. The influence of guideline evidence is exemplified by NICE guidelines' significant role in forming the evidence base for improving access to psychological therapies (IAPT; DoH, 2007). Moreover, it extends to education and the training of healthcare professionals, particularly clinical and applied psychologists who are predominantly at the front line of psychotherapy provision. A further implication is how guideline recommendations shape perceptions and expectations of other professionals and the public about which psychotherapies are likely to be helpful (Pilling, 2008). Thus, guidelines place psychotherapies under increasing scrutiny of various stakeholders including commissioners, services providers and the public and at the same time have the potential to play a significant role in shaping practice via these channels.

2.5.2 NICE psychotherapy recommendations: controversies and debates

Some take a positive view of guidelines and their role in identifying psychological interventions as mainstream treatment options within a healthcare system that is broadly diagnostic or condition based. For instance, Barlow (2004) has further suggested that 'evidence-based psychotherapies' with a clear medical objective be distinguished from others by adopting the label 'psychological treatments'. Others take a critical view, arguing that medically oriented guidelines have the potential to restrict psychotherapies that do not fit as neatly into the EBP paradigm (e.g. Nilsson 2010). Despite being central treatment options for a range of mental health conditions, many of the concerns relevant to the field of psychotherapy research including common factors and process-outcome variables, which certain theoretical orientations place a strong emphasis on, are not considered within NICE guideline evidence reviews.

NICE psychotherapy recommendations have been criticised further for representing a competitive branding and marketing of psychological therapies, which is detrimental to an integration of the multiple factors involved in psychological distress and recovery (Mollon, 2009). Guy et al. (2012) argue that NICE guidelines privilege results from RCTs at the expense of pluralistic approaches to scientific enquiry. They argue that the 'drug metaphor' for the purposes of psychotherapy research is insufficient to address the complex processes involved in psychotherapy. Furthermore, Guy and colleagues highlight the continued inflexibility of evidence hierarchies despite formal removal of this system within more recent revisions of the guideline manual due to the potential risks of discounting robust evidence from non-RCTs (see Pilling, 2008; Pearce et al. 2015 for a more detailed discussion of this).

In contrast, Pilling (2009) argues that NICE mental health guidelines have recommended a range of psychological interventions and are not limited only to recommending CBT. However, Holmes (2002), drawing on adult depression guidelines, argues that when detailed recommendations are scrutinised more closely it is apparent CBT-based interventions are promoted as the therapy of choice, in terms of breadth and scope of the recommendations relative to the other psychotherapies that are recommended within the guideline.

2.5.3 The status of cognitive behavioural therapy (CBT) in NICE guidelines

Much of the controversy of NICE recommendations of psychotherapy relates to CBT's unmatched support in the treatment of a broad spectrum of mental disorders relative to other therapeutic orientations. NICE recommends CBT as the primary treatment choice for depression and a range of mental disorders including anxiety disorders (panic and post-traumatic stress disorder), bipolar disorder and psychosis. Furthermore, NICE recommendations have been instrumental in the government investment for CBT provisions of mental health problems in England. Mollon (2009)

suggests that CBT recommendations within guidelines are harnessed to the government initiatives of the Improving Access to Psychological Therapies Programme (see Layard, Bell, Clark, Knapp, Meacher, Priebe & Wright, 2006;). The high compatibility of CBT with the evidence-base practice movement has only served to further cement its status (Stiwne and Abrandt Dahlgren, 2004).

Proponents of CBT point out that it is arguably the most widely studied form of psychotherapy, with an extensive evidence base available (Hofmann et al. 2012; Hunot et al., 2013). This evidence consists of a large amount of outcome research that has demonstrated symptom reduction, improvement in functioning and remission of the disorder. Such outcomes are consistent with a medical model of mental disorder and within this context it is easy to appreciate CBT's appeal.

2.6 Guideline development process

NICE outlines a clear protocol for the development of guidelines. This involves a number of distinct and overlapping phases from decisions on the initial scope of the guideline, selecting members of the guideline development group, developing the clinical questions to review and making recommendations. Details of the development process and the values that underpin them are described within 'The guidelines manual' (NICE, 2009; NICE, 2014) and within full versions of each guideline (i.e. NCCMH, 2010).

Alderson and Tan (2011) argue that NICE guideline reviews of evidence are distinct from Cochrane collaboration (and other systematic) reviews, due to their combined use of research evidence (i.e. systematic review and health economic analyses) and clinical expertise from healthcare professionals and patients. Similarly, Parry, Cape and Pilling (2003) assert that within the context of guidelines, the evidence base refers to both research and structured clinical opinion. Alderson and Tan describe a

further core distinctive characteristic of guideline development programmes is the scope of about 20 to 25 clinical questions that are addressed by systematic reviews. This is intended to ensure the guideline's broad relevance to clinical practice.

The manner that secondary research evidence is generated, organised and interpreted is specific to a guideline context; this makes evaluation of this type of evidence of interest within the current project. Parry et al. (2003) argue that it is difficult to predict the impact of multi-modal guidelines on the types of recommendations that emerge for psychological interventions[4]. They suggest that recommendations in part reflect the methodological approach taken, which varies across different guideline topics. It is also difficult to predict the impact of psychotherapy evidence on psychotherapy recommendations within a particular guideline, particularly when one considers the broader medical context in which psychotherapy evidence is reviewed. Therefore, examination of how psychotherapy evidence interacts with the methodological approach adopted by NICE guidelines is of particular interest.

2.6.1 National Collaborative Centre for Mental Health (NCCMH)

The National Collaborative Centre for Mental Health (NCCMH) is responsible for the development of NICE guidelines for mental health conditions. The organisation consists of technical researchers without practitioner affiliation who are responsible for systematically reviewing the relevant literature and conducting statistical analyses that are presented to multidisciplinary guideline panels. The guideline development process takes place under the auspices of two professional bodies, the Royal College of Psychiatrists (MRCPsych) and the British Psychological Society (BPS).

[4] The multi-modal nature of NICE mental health guidelines refers to broad-based approach to assessment consisting of social, organisational, psychological and pharmacological approaches to treatment and management of a problem or disorder.

For each condition or specific disorder there is a representative guideline development group (GDG) of professionals, patients and carers, who combine their expertise with evidence.

2.6.2 NICE systematic review process

Systematic reviews identify, appraise, select and synthesize research evidence relevant to a particular research question. As noted above RCTs form the evidence base for the evaluation of psychotherapy's effectiveness. The NICE systematic review process is consistent with those within the wider EBP paradigm in prioritising evidence from RCTs to evaluate the effectiveness of a particular psychotherapy. The methods used to achieve these aims are an integral part of process and outcome. Systematic reviews require complex and sophisticated search strategies to identify all of the relevant literature. National guideline development programmes have the resources necessary to perform well-designed electronically based search strategies supplemented with hand searching of the literature and regularly updated searches. However, further challenges are present in respect to the quality of the primary studies available for review and bias within the body of available evidence (Pilling, 2008).

2.6.3 Meta-analysis and NICE guidelines

Meta-analysis is an important tool within EBP and within guideline development programmes. Within guideline development its statistical aims include: obtaining more precise estimates of overall treatment effects; the evaluation of interventions on specific subgroups of patients; overcoming the difficulties of limited statistical power in small trials; and assessing safety and rare adverse events (Egger, Davey and Altman, 2001).

Although meta-analytic techniques potentially provide greater statistical power and precision, they remain vulnerable to numerous sources of bias inherent in the conduct of the primary clinical trials and wider bias within the research literature (Pearce et al., 2013). Further challenges are encountered in efforts to combine evidence that varies in selection of outcome measures, inclusion criteria (e.g. client population, diagnosis and severity), delivery of trial interventions and treatment settings (see Roth and Fonagy, 2005). Thus, variations in primary study design, study characteristics and quality require consideration within both the conduct and interpretation of meta-analytic findings.

Within the field of psychotherapy research, study characteristics of primary studies used in meta-analyses have been coded and analysed by psychotherapy researchers using various regression techniques to determine their influence on outcome (e.g. Smith and Glass, 1980; Wampold, 2001; Cuijipers, Van Straten, Warmedam, 2008). However, these techniques are not employed within guideline meta-analyses so the impact of study characteristics on guideline evidence remains unclear' at the end of the sentence.

2.7 Reviews of NICE guidelines

The widespread support for CBT-based interventions within NICE guidelines relative to other psychotherapies has prompted recent reviews of the evidence base for CBT, which have questioned the methods employed within NICE evidence reviews and the strength of the evidence (e.g. Jauhar, McKenna, Radua, Fung, Salvador and Laws, 2014; Jauhar, McKenna and Laws et al., 2016 Nel, 2014).

Jauhar et al. (2016) cautioned that judgements based on meta-analyses require that primary findings are both reliable and valid, arguing that this is not always the case. They draw on the broad area of primary study quality, arguing that its influence is

not routinely accounted for within NICE's meta-analytic reviews for bipolar disorder guideline. Jauhar et al's (2014) updated meta-analysis of CBT for schizophrenia revealed that the small treatment effect sizes reported within the NICE guideline for schizophrenia (see NCCMH, 2009) reduced further when controlling for the influence of masking at assessment, which is a known sources of bias.

Nel (2014) conducted a meta-review of the CBT evidence base for treatment of depression in children and young people. This review highlights a number of methodological issues and inconsistencies within the primary studies that are used to support recommendations. These issues include the diverse use of CBT as a treatment approach and trial conditions, small sample sizes, varying inclusion and exclusion criteria, generalisability of findings and diverse use of outcome measures. It is argued that these issues are not adequately taken into account within the quantitative evaluation of evidence within the guideline. Moreover, Nel's review suggests that evidence is misrepresented within the guideline's classification of non-active controls as 'psychological interventions' which are subsequently presented as evidence of relative efficacy of CBT within the guideline. This has further implications for the overall quality and validity of evidence presented in guidelines and the subsequent recommendations.

The author questions whether CBT recommendations can be considered 'evidence-based' on the basis of the evidence presented for individual CBT. Further, Nel argues these inadequacies raises important questions about NICE's process for ensuring that evidence is reviewed adequately and how these processes can be improved. This begs the question as to whether there is wider systematic bias inherent within the processes by which primary research evidence is evaluated by guideline developers.

Further concerns have been raised in relation to the strength of the evidence supporting CBT. Nel's review identified limited or weak evidence across the 4 RCTs.

This included non-significant treatment effects of CBT compared to control conditions, no difference in treatment effect at follow-up and inconsistent depression scores on self and parent versions of outcome measures. Furthermore, the review identified that non-statistical differences were arbitrarily judged within the guideline to indicate clinically important improvements despite these small differences lack of actual clinical significance.

These reviews illustrate the challenges that the consumer of evidence-base reviews faces in relation to the interpretation of guideline reviews and recommendations of psychotherapy. There is an inherent contradiction in a meta-analytic process where primary evidence of psychotherapy trials is extracted from diverse contexts and purposes at the level of primary studies and treated in a uniform way in order to achieve objective, unbiased and precise appraisals of evidence.

Perhaps the phases of the current EBP paradigm, from assumptions and conduct of the primary RCTs through to the meta-analytic extraction and synthesis of data, represent incremental distance between guidelines and the primary evidence within its original context. This ultimately creates distance from its intended clinical practice context, which requires further meta-analytic procedures to bridge the gulf. Within the broader field of meta-analysis, the use of meta-regression techniques enables more explicit quantitative appraisals of study characteristics, which can arguably maintain a connection to the study's original context. However, such techniques are not employed within the sets of meta-analyses conducted within NICE mental health guideline programmes.

Guideline programmes do incorporate quality rating outcomes into guideline development (e.g. Oxman and GRADE Working Group, 2004). Kendall (2016) argues that these provide an adequate level of transparency between the recommendations and the evidence presented within guidelines that enables its users to draw their own conclusions based upon the specific purpose that guidelines

are to be referred to. However, questions remain about how quality of evidence interacts with the strength of the evidence for the recommended psychological interventions.

2.8 NICE guidelines and Depression

Depression results from a complex interaction of social, psychological and biological factors (WHO, 2016). Those who have experienced adverse life events are more likely to experience depression, which in turn can lead to increased stress and dysfunction and further depression. A large body of research evidence is available for the treatment of depression from different theoretical orientations. According to Roth and Fonagy (2005) this reflects the common and chronic nature of the disorder. Major guideline development programmes have a tendency to start here (Parry et al.,2008).

The effectiveness of psychological interventions for depression has been well established, with results from meta-analyses indicating that most psychological treatments that are studied produce a considerable positive effect (Cuijpers, van Straten, Warmedam, 2008). However, research has generally provided little evidence showing any specific intervention to be both specific and efficacious (Barkham and Parry, 2008). Major reviews of the meta-analytic evidence of comparative trials of psychotherapy for depression (e.g. Hoffman, Asnaani, Vonk, Sawyer and Fang, 2012; Lambert, 2013a; Roth and Fonagy, 2005) have generally concluded that there is no clear advantage to any particular approach.

Lambert (2013b) notes that in the case of depression many elements are likely shared by diverse treatment orientations and modalities and these common factors loom large in improvement of function. For instance Cuijpers' et al. (2012b) examination of the effects of non-directive supportive therapy for adult depression,

which included comparisons of this approach with other psychotherapies, concluded that most of the effects of therapy for adult depression are realized by non-specific factors and the contribution of specific effects is limited at best. Further support for common factors in psychotherapy is found within the contextual factors model that conceptualizes psychotherapy as a socially constructed and mediated healing practice (Frank and Frank, 1993; Wampold, 2001) and the body of supporting evidence that contextual factors (i.e. therapeutic relationship, a healing setting, treatment rationale, and a procedure to resolve them, etc) contribute considerably more to psychotherapy outcomes than specific components of therapy alone (Wampold, 1997; 2002). From the common factors perspective, specific models of psychotherapy (i.e. the underlying theoretical explanation and treatment procedures) only partly represent these necessary components for therapeutic change.

NICE depression guidelines have received criticism for continuing to adopt an EBP paradigm, which prioritise RCTs of psychotherapy brands proposed to contain specific therapeutic ingredients (Laska, Gurman and Wampold, 2014), despite the overwhelming evidence that contradicts adopting such an exclusive approach. Furthermore despite the equivalence in psychotherapy outcome of a range of psychotherapies for the treatment of depression a narrative has emerged both within depression guidelines and more broadly across the spectrum of disorders of CBT's prominence as a treatment based upon the secondary evidence presented within the guidelines.

In addition to concerns about the type of evidence, questions have also been raised about the consistency of the interpretation of NICE, or more specifically NCCMH, evidence. Winter (2010) highlights a number of inconsistencies in the updated NICE guidelines on depression in adults (NCCMH, 2010), highlighting that the evidence reported regarding comparative outcome research on CBT indicates no significant

difference from a range of other psychotherapies reviewed including behavioural activation, interpersonal therapy or 'non-directive psychotherapies'. Furthermore, there were no significant differences from usual GP care or placebo although there was some indication of relative effectiveness of CBT when compared to antidepressants. Winter (2010) argues that the NICE recommendations for cognitive behavioural models of therapy appear to pay scant relation to the evidence.

2.9 Returning to social constructionism and clinical guidelines

Lock and Strong (2010) suggest that social constructionism enables a focus on the recognition of multiple possibilities for meaning and transformative action where some convention or taken-for-granted understanding holds influence. NICE's evidence base represents one of many possibilities in respect to broader based evidence for psychotherapies. It is necessary to examine this constructed meaning and the transformative actions taken within its context (i.e. within the context of the guideline development process).

The social constructionist stance adopted within this project takes a view that the current narratives and discourse of evidence-base practice, in mental healthcare and psychotherapy more specifically, are dependent upon NICE as an institutional structure. Danziger (1997) describes this strand of social constructionism as concerned with the ongoing construction of meaning in current discursive practices. Slade and Priebe (2000) provide an indirect example of this through describing the function achieved by importing an evidence-based approach into mental health, namely strengthening the position of pharmacological interventions, which by their very nature are standardised and well-defined relative to psychological and social interventions. They argue that this serves a further function of underlining the link between psychiatry and other specialities.

The establishment of government funded bodies such as NICE that are rooted within the same evidence-based paradigm provides the framework for the ongoing constructions of evidence. This is further supported by its own internal infrastructures that enable self-regulation of its own guideline development processes and textual conventions initially through the guideline publications in the various formats and extending to broader mechanisms for the implication of recommendations (i.e. care pathways, standards and indicators and NHS evidence). This is further maintained and enhanced by the ability to consolidate and disseminate information rapidly through a range of web-based resources (Leng, 2009).

In relation to the construction of evidence, Moreira's (2007) ethnographic study of knowledge construction in systematic and meta-analytic reviews sheds light on how evidence contributes to NICE as a broader institutional structure. Moreira argued that in the development of clinical guidelines, knowledge is constructed via two parallel processes: firstly, the disentanglement of data from its original context and secondly, through a process of qualification of evidence that involves endowing data with new qualities such as precision, where it becomes 'unbiased' and fair. Moreira argues that these processes become vehicles of 'persuasive power' that meta-analytic and systematic reviews have in contemporary healthcare systems. This situates NICE guideline evidence within the politics of healthcare, where it becomes a tool for shaping the structure of the debates about knowledge and evidence.

This reflects how within guidelines evidence is not only constructed but re-constructed for the purposes of guideline recommendations. Moreira's description of these processes is of particular relevance to the current study: the re-qualifying of data. In this respect systematic reviews and meta-analysis mediate between particular types of evidence and the politics of healthcare (Moreira, 2007). Systematic reviews are a central part of a set of processes where data from primary

studies are disentangled from their original context and meta-analyses provide the means of statistical manipulation and re-qualification of the data. Moreira's findings support Lock and Strong's (2010) assertion of a political component to social science in general and psychology in particular, whereby 'facts' are not neutrally awaiting discovery. Instead, such 'facts' are constructed in fields of activities, and assembled into ideologies that benefit some positions whilst disempowering others.

It is possible that similar processes occur within the construction of guideline evidence and therefore these constructed meanings remain unclear without direct exploration. Direct analysis of the secondary evidence generated within guideline development programmes as a basis for psychotherapy recommendations has seldom been the direct focus of review. This is surprising given their ability to impact policy and practice.

2.10 Study Aims and Research Questions

2.10.1 Study aims

This project aims to examine the secondary evidence base generated by NCCMH, which forms the basis for NICE guideline recommendations for psychological interventions in the management of adult depression. This review of the evidence will focus on exploring the similarities and differences between that for cognitive-behavioural based interventions and for other psychological interventions recommended within the guideline.

2.10.2 Research questions

The primary research questions are:

1) How strong is the evidence within the guideline in support of CBT-based interventions and how does this compare with the strength of evidence in support of other psychological interventions recommended?
2) Are there identifiable patterns of difference in the evidence for CBT and other psychotherapies?
3) Are specific characteristics of NICE reviews associated with the differences observed between the strength of evidence for CBT and other psychotherapies?
4) Does the guideline's approach to constructing evidence create bias against other psychotherapies in comparison to CBT?

2.10.3 Why this review is necessary

Although some direct reviews of the evidence base have been conducted for other guidelines, as discussed earlier in this chapter, no reviews have explored the evidence generated by NICE in the process of developing psychotherapy recommendations for adult depression. Therefore, currently there is a dearth of research on this particular aspect of the guideline development process. This is of particular importance when one considers the impact of recommendations that are presented as 'evidence-based' on policy and practice regarding depression, which represents one of the most common mental health problems encountered within the NHS.

This project aims to explore the relationship between the evidence presented for recommended psychotherapies and features/characteristics of the meta-analytic review. Considering that there is evidence that study characteristics can influence psychotherapy outcome and interpretation very little direct research examining these factors associated with NICE's secondary evidence has been conducted. Thus this review aims to understand and demystify the evidence base for NICE

psychotherapy recommendations within its context. As this review has highlighted, the secondary evidence generated by NICE is unique to the guideline development context. This is in part by virtue of the broad remit of guidelines within healthcare. This distinguishes it from the majority of evidence reviews in depression that occur in parallel to this context of the wider field of psychotherapy research.

A key factor of clinical guidelines is transparency and that the evidence base (research or clinical opinion) is clearly indicated in each recommendation, so that users of the guideline can evaluate it (Parry et al., 2003). Therefore, guideline evidence should be able to demonstrate robust evidence to support recommended psychological interventions. Moreover, any issues in the strength (and quality) of evidence for a recommended psychotherapy should be consistent within all the broad classes of psychotherapy that are evaluated within the guideline development process.

3. Methodology

3.1 Chapter outline

This chapter begins by considering the research design, methodological and data issues relevant to conducting a study of this kind. I will then summarise the stages involved in conducting this review of the secondary evidence for psychological interventions using exploratory data analysis; this involves a number of distinct stages utilising quantitative and qualitative methods. This will be followed by an outline of the statistical strategy employed. Finally, I will discuss my perspective within the researcher process and reflect on how this could have impacted on the direction of exploration during the review process.

3.2 Design and methodological considerations

This section will discuss the research design and approaches that were employed at different stages of the review, focusing on the appropriateness of the methodological stance taken throughout the project.

3.2.1 Exploratory Data Analysis (EDA)

This project used exploratory data analysis to examine the evidence presented in NICE documentation, which forms the basis of psychotherapy recommendations for depression. More specifically, this project explores the quantitative, meta-analytic secondary evidence relating to recommended psychotherapies for the NICE guideline adult depression update (NCCMH, 2010).

Exploratory Data Analysis (EDA) refers to the tradition dating back to the early 1960s developed by John Tukey. According to Behrens (1997) 'this tradition can be loosely characterized by an emphasis on (a) substantive understanding of the data to answer

the broader question of 'what is going on here?'; (b) an emphasis on graphic representations of data; (c) a focus on tentative model building and hypothesis generation in an iterative process of model specification, residual analysis, and model re-specification; (d) use of robust measures, re-expression, and subset analysis; and (e) positions of skepticism, flexibility, and ecumenism regarding which methods to apply.' (p. 132). However, Tukey (1980) asserts that EDA is characterized more by the attitude taken to research rather than a particular set of techniques adopted. EDA enables phenomena observed within the data to generate questions or hypotheses about possible causes rather than imposing these a priori (Vigni, Durante and Cocchi, 2013). According to Behrens (1997) EDA emphasizes that at different stages of research there are different types of questions, different levels of hypothesis specificity used, and different levels of conclusions warranted. This makes EDA an appropriate fit for the current project's focus.

3.2.2 EDA's appropriateness for this research

The above-mentioned characteristics of the EDA tradition hold several advantages within the context of the current project. EDA is a data-driven approach, which enables closer examination of relationships between recommended psychotherapies, effect sizes and different meta-analytic characteristics. A large volume of data is generated through the guideline development process and it is argued here that this data has not been held to as much scrutiny as the primary evidence that it is based upon. Thus, an approach that enables the data to 'talk' has the potential to offer new insights into guideline evidence that psychotherapy recommendations are based upon. The scope for re-expression of data and graphical representations to enable the data to talk differently also makes EDA an appropriate approach. Furthermore, an EDA tradition offers flexibility in respect to

methodological and philosophical stances taken by the researcher (discussed later within this chapter).

The EDA tradition, however, is not a primary approach used within psychological research. Confirmatory approaches are predominant within psychotherapy research and systematic reviews including the use of meta-analysis. King and Resick (2014) assert that most psychology researchers associate the notion of EDA with preliminary descriptive statistics prior to the use of more substantive tests. However, they argue that EDA provides the fuel for confirmatory research by generating stronger questions, which can lead to more refined study designs. When applied to psychotherapy, EDA approaches have been utilized to find the unexpected (Pokorny, 2015). This may involve a considerable amount of 'data mining', revealing unforeseen but critical insights as to with *whom* and under *which* circumstances treatments are beneficial. They also have been used to highlight sources of bias within treatment studies. Other examples of where EDA and data mining have been utilized to good effect in psychotherapy outcome research are in the exploration of randomized controlled trial (RCT) data and for preliminary evidence of treatment moderators and mediators (e.g. Kraemer, Wilson, Fairburn, and Agras, 2002). This is vital in understanding how and when these treatments have beneficial effect. Such data mining of treatment study results has been employed, albeit in a confirmatory fashion, in change process research (e.g. Elliott, 2009). These approaches have also been integral to meta-analyses used to demonstrate common factors between different classes of psychotherapies (e.g. Wampold, 2001). Data mining techniques, such as working with means and collapsing data into smaller samples to find new meanings essentially require the researcher to stay with the messiness of the data.

The current research project differs in its direct focus on secondary evidence generated through the guideline development process. This introduces a new level of abstraction in order to explore relationships between recommended

psychotherapies and the strength of the evidence by exploring how they interact with specific meta-analytic factors. This not only enables exploration of whether or not patterns of difference exist between recommended psychotherapies, but also provides an understanding within a guideline specific context, which can be used further to consider how this relates to wider contexts of psychotherapy research.

EDA offers the means to look at data from a new perspective. This makes it an appropriate fit with this project's broader aims of understanding evidence within its context. Moreover, EDA enables an inductive approach to the analysis of data in order to develop clearer hypotheses about which aspects of the evidence base are influencing psychotherapy recommendations within the guideline development process. It enables the researcher the flexibility to 'let the data talk' and to explore new and unexpected avenues that emerge through observation of the data within the context of the guideline documents.

Within the current psychotherapy literature the relationship between guideline evidence and psychotherapy recommendations is rarely examined. Although some commentators have suggested that there is a general bias towards CBT in guideline recommendations and, as reviewed in the earlier chapter, such assertions find some support in recent reviews that contested the strength of the evidence-base for CBT as a basis for wider recommendations over other psychotherapies, direct examinations of this within specific guideline evidence contexts are lacking. An exploratory approach will enable the development of a set of procedures concerning data collection methods, and the selection of variables and statistical strategies specific to this context that could further our understanding of this area. Moreover, this could enable an inductive approach to any further question or hypothesis formation in this area.

3.3 Procedure

3.3.1 Data Sources

The NICE Guideline document 'Depression: the Treatment and Management of Depression in Adults, Update' (NCCMH, 2010) was the primary data source for the quantitative review of evidence including the guideline's supporting appendices of evidence profiles and forest plots (Appendix 16b and Appendix 19b respectively). These contained a large set of meta-analyses that were conducted by NCCMH in development of the depression guideline's psychotherapy recommendations. The data was extracted from the evidence profiles as described below for each of the psychological interventions that were recommended within the full guideline.

3.3.2 Raw data extraction process

Two Excel/ SPSS databases were created for data extraction purposes. Raw data were organized into separate databases depending on broad type of outcome, i.e. continuous or dichotomous outcomes[1]. This enabled subsequent data analysis to be based upon similar metrics, i.e. risk ratios associations or effect sizes for dichotomous and continuous outcomes respectively. All subsequent analysis of the extracted data was carried out within each database separately.

3.3.3 Data extraction of meta-analysis characteristics

Characteristics of meta-analyses for recommended psychotherapies were extracted from the NICE evidence profiles into the relevant database. The following characteristics were extracted, which are described in turn below:

[1] Continuous outcomes are based upon numerical scores on a scale measuring a domain of change such as symptoms of depression. A dichotomous outcome refers to binary measures used to determine if a patient has improved or not using a variety of criteria.

- Treatment intensity (high or low intensity depression)
- Specific type of psychotherapy recommended
- Raw effect size and/ or associations (for continuous or dichotomous outcomes)
- Treatment comparator
- Outcome/ risk type (dichotomous outcomes only)
- GRADE quality rating.

3.3.3.1 Treatment intensity

Treatment intensity was either 'low' or 'high' intensity depending on whether the intervention targeted mild- moderate or moderate- severe depression in the population sampled within the primary studies used in NICE's meta-analysis. Treatment intensity relates to the severity of depression in the population sampled within each primary study used in the set of meta-analyses that comprise secondary evidence.

3.3.3.2 Specific type of psychotherapy recommended

Details were extracted from the evidence profiles of the specific type of psychotherapy that outcomes were presented for. For low intensity depression this included:

- Computerised CBT (CCBT)
- Individual guided self-help based upon CBT principles

- Group structured physical activity programme[2].

For high intensity depression recommendations included:

- CBT
- Interpersonal therapy (IPT)
- Behavioural activation
- Counselling[3]
- Couples Therapy (based on behavioural principles)
- Short-term psychodynamic psychotherapy (STTP)

3.3.3.3 Raw effect sizes and risk associations

The raw effect size and risk association was extracted from each meta-analysis of recommended psychotherapy and inserted in the relevant database for continuous and dichotomous outcomes. At this stage it was necessary to capture further details relevant to these meta-analytic outcomes.

An effect size measures the size of the difference between two treatment conditions. The measure of effect size within the meta-analyses performed by NICE was Cohen's d, which is calculated by taking the difference between the two treatments conditions' means divided by the average of their standard deviations. Therefore, a Cohen's d of 1 indicates that the two group means differ by one standard deviation; Cohen's d of 0.5 means that the two groups' means differ by half a standard deviation; and so on.

[2] Although physical activity is not a form of psychotherapy, as it was evaluated and recommended as a psychological intervention within NICE (2009) it was included within the analysis.
[3] Although counselling was recommended as a treatment option for mild to moderate depression, this was based on evidence for high intensity depression interventions.

A relative risk (RR) is the ratio of the treatment event rate to the control event rate of a negative outcome (e.g. leaving the study early). An RR of less than one indicates a reduced risk of negative outcome. For example, an intervention with an RR of 0.75 indicates that the event rate 'leaving the study early' in the intervention group is about three quarters of that in the control group or, in other words, the RR reduction is 27%. An RR value of 1 indicates no difference between treatment and control.

3.3.3.4 Treatment comparator

Data on treatment comparators were extracted for each meta-analysis into the relevant database (i.e. continuous and dichotomous outcomes) with the corresponding effect size or risk association. There were a range of different treatment comparators that the recommended psychotherapies were compared against; these included:

- *vs* placebo control
- *vs* non-active psychotherapy control
- *vs* active control
- *vs* psychotherapy (including other recommended psychotherapy)
- *vs* medication
- *vs* medication combined with non-active comparators
- Recommended psychotherapy combined with medication *vs* psychotherapy
- Recommended psychotherapy combined with medication *vs* medication
- Recommended psychotherapy combined with medication *vs* psychotherapy combined with another psychotherapy

3.3.3.5 Risk outcome type (dichotomous outcomes only)

As the dichotomous data consisted of different risks of a negative treatment outcome, it was necessary to extract additional details from the evidence profiles about the type of risk outcome that a given risk association value related to. These included:

- Depression score (above/ below cut-off score)
- Leaving the study early
- Leaving treatment early
- Relapse
- Recurrence of depression
- not achieving remission

3.3.3.6 Grading of recommendations assessment, development and evaluation (GRADE) quality rating

Data on the quality rating of each individual meta-analysis was extracted from the evidence profiles. These ratings were based upon the Grading of Recommendations Assessment, Development and Evaluation (GRADE, 2004) system that is incorporated into NICE's evidence review process. GRADE classifies the overall 'quality' of underlying evidence into 'high', 'moderate', 'low', and 'very low'. This is based upon an aggregation of factors relating to the quality of the primary studies (randomization, trial treatment protocols, etc.) and the statistical strength of subsequent meta-analysis that they performed (power and significance, etc.). Quality ratings were extracted for each corresponding meta-analytic outcome (i.e. raw effect size or risk association) for recommended psychotherapies presented within the evidence profiles.

3.3.4 Coding of meta-analysis characteristics

Following the initial extraction of data further classification and re-coding of variables was required for the purposes of exploratory data analysis. These are described for each of the aforementioned characteristics.

3.3.4.1 Recommended psychotherapies

Recommended psychotherapies were re-coded into two broader groups in order to provide meaningful units of examination of guideline evidence for the purposes of this project. The two broad psychotherapy groupings consisted of specific types of psychotherapies (as listed in 3.3.3.2) being re-coded as those that were broadly cognitive *and* behavioural (i.e. CBT group) and those that were from other theoretical traditions (i.e. other psychotherapies group).

The CBT group included:

- computerized-CBT,
- group-based CBT interventions
- third-wave CBT approaches
- CBT interventions that targeted specific aspects of depression (e.g. mindfulness-based CBT for relapse prevention)

The other psychotherapies group consisted of the following interventions:

- Interpersonal therapy (IPT)
- Behavioural activation[4]
- Couples Therapy (based on behavioural principles)
- Short-term psychodynamic psychotherapy (STTP)

[4] Psychotherapies that used a solely behavioural approach were considered distinct from cognitive behavioural therapies, thus their inclusion in the other psychotherapies group.

- Physical activity
- Counselling

This decision to place specific therapies into distinct groups was taken in line with the project's aims and research questions (see sections 2.10.1 and 2.10.2). Moreover, the decision to place specific treatments into 'cognitive and behavioural' based interventions or 'other psychotherapies' was informed by how the specific treatments were organised and assessed within the NICE guideline's own evidence review (see NCCMH, 2010). Those that were reviewed by NICE as cognitive behavioural therapies were placed within this group and those that were not were placed in the other psychotherapies. Thus, even though some treatment shared similar components to CBT but was not assessed under this class of therapy by NICE, the specific therapy was coded under the other psychotherapies group: this was the case with behavioural activation and couples therapy (based on behavioural principles).

3.3.4.2 Raw effect size

Within the continuous outcome database, the raw effect size data extracted for each meta-analysis performed were further described using Cohen's (1988) designations 'small', 'medium', and 'large' to reflect the effect size magnitudes. 'Small' denotes an effect-size of above -0.2 and below -.0.5, 'medium' denotes an effect size above -0.5 and below -0.8, and 'large' denotes an effect-size of -0.8 or above[5]. Effect sizes below -0.2 were coded as 'none' effect, as were effect size values that indicated an equivalence with the treatment comparator (i.e. an effect size of 0) and effect sizes

[5] Within NICE guideline evidence negative effect sizes indicate a recommended psychotherapy's superiority over its comparator condition.

that indicated an inferiority of a recommended psychotherapy to a treatment comparator (i.e. effect sizes of 0.2 and above).

3.3.4.3 Raw relative risk ratios

Within the dichotomous outcome database, the raw relative risk data for each meta-analysis was further described as 'more' or 'less' for likelihood of a negative depression-related outcome. This was based upon their values being above 1 (i.e. indicating an increased risk) or below 1 (i.e. indicating a decreased risk).

These re-coded outcomes (effect sizes and relative risks) represent a more qualitative, albeit rather crude, description of the evidence supporting recommended psychotherapies. These descriptions arguably mirror the level of information accessible to guideline development panels once translated from graphically represented data in evidence profiles. Furthermore, the transformation of continuous data to discrete, categorical (more specifically, ordinal) data enabled each meta-analysis to contribute to an overall headcount of outcomes. This was necessary to understand the data in its context and draw meaningful comparisons between the two recommended psychotherapy groups for the purposes of this project.

3.3.4.4 Treatment comparators re-coded

For the purposes of further analysis, the data extracted on treatment comparators (see section 3.3.3.4) were re-coded into broader groups. The decision to pool the data in some groups was based upon small amounts of meta-analytic data available in some of the original treatment comparator groups particularly within the other psychotherapy group. Therefore, this resulted in the following treatment comparators:

- non-active comparators (pooling non-active psychotherapy controls and placebo)
- active comparators (pooling active controls/ psychotherapy)
- medication
- medication combined with non-active comparators
- psychotherapy combined with medication

3.3.4.5 Risk outcome type re-coded (dichotomous outcomes only)

Similarly, data extracted on risk outcome type (see subsection 3.3.3.5) was re-coded for the purposes of data analysis, which included the pooling of some smaller categories to create larger groups. Therefore, the original six sub-types of risk outcome were regrouped as follows:

- Depression score
- Leaving study/ treatment early
- Relapse/ recurrence
- Not achieving remission

3.3.4.6 Other characteristics

The remaining meta-analytic characteristics were retained as originally coded on data extraction. These included the following:

- Treatment intensity
- Low intensity depression
- High intensity depression

- GRADE quality rating
- low
- medium
- high

3.3.5 Statistical analysis

This section outlines the statistical strategy that was employed following the extraction and coding of data. As discussed earlier an exploratory approach was taken that enabled the flexibility to observe and follow patterns of interest within the data. Employing an EDA approach meant that data was examined at different levels, namely interval and ordinal levels. Therefore, I shall clarify the level of data that was analyzed throughout this section.

I shall summarize the descriptive analyses performed on the raw data to examine for patterns of differences in effect sizes/ relative risk ratios for the two psychotherapy groups and the tests of difference that were performed. This was in reference to the initial research question as to whether or not differences exist in the strength of evidence between CBT and other recommended psychotherapies. I shall then summarize the test procedures used to explore the strength of evidence for each group in relation to various meta-analytic characteristics. Finally, I shall summarize the descriptive and correlation analysis performed on the data at a categorical level to examine the effect-size/ risk associations.

The statistical strategy summarized represents a process of exploration, which was recursive, rather than a step-by-step account of analysis undertaken in the strategy. The statistical tests summarized were often carried out as 'post-hoc' examinations in response to patterns observed from initial tests of difference between two

recommended psychotherapy groups and were used to expand the range of data exploration.

The majority of the statistical analyses were performed in SPSS®. EPI Info 7 StatCalc was also used to compute 2x2 tables and to perform Chi-square tests of independence/ Fisher's exact tests in order to explore differences in effect size magnitude and risk associations for the two groups of recommended psychotherapies and whether there was an interaction with other meta-analytic characteristics.

3.3.5.1 Descriptive statistics

Initially, univariate analyses were used to compute summaries about the raw effect sizes and relative risk ratios (continuous level data) separately for each recommended psychotherapy group (CBT and other psychotherapies). This enabled examination of the overall amount of meta-analytic evidence within each group and the amount within treatment intensity sub-sets. Mean effect sizes and risk associations were also computed for the two psychotherapy groups and for various sub-group analyses. Further descriptive analyses were performed on categorical level data, namely the proportions of effect size magnitudes and the categories of risk association of a negative outcome, for the two psychotherapy groups. These descriptive data formed the basis to explore broad similarities and differences within secondary evidence for the two recommended psychotherapy groups.

3.3.5.2 Independent samples t-test

I was interested in the difference in the mean effect sizes and risk associations between the two recommended psychotherapy groups (i.e. CBT and other psychotherapies). Therefore an initial step in the analysis was to perform an

independent sample t-test on mean effect sizes for the two psychotherapy groups to determine whether or not there was a significant difference. Although the data were not drawn from standardized populations, the raw effect sizes were continuous (interval) data and the sample sizes in each group were large enough to perform a parametric test on non-standardized data. Whereas non-parametric tests assume equal variance within two conditions, this parametric test calculates Levene's Test for Equality of Variances to determine if the two conditions have about the same or different amounts of variability between scores. This enabled selection of the appropriate significance value.

Data for high and low intensity depression for the two psychotherapy groups were analyzed in two different ways. Initially effect sizes of both intensities were pooled together for analysis. This enabled exploration of similarities and differences between the broader classes of recommended psychotherapy across the guideline evidence in its entirety. The data was then split by intensity to control for this variable within the analysis. This was done to reflect that low and high intensity recommended psychotherapy groups were comprised of different sets of psychotherapies. This enabled me to explore whether effect sizes differed specifically within treatment intensity.

3.3.5.3 Kruskal-Wallis H Test

This test was performed to determine whether the raw effect sizes (interval data) differed significantly when controlling for the following meta-analytic characteristics: treatment comparator (see section 3.3.4.5), risk types (see section 3.3.4.6) and quality outcome rating (see section 3.3.4.7). Initially all data were pooled for both psychotherapy groups (CBT and other psychotherapies) in order to determine whether meta-analytic characteristics had a general influence on the ranked-means of effect sizes/risk associations. Kruskal-Wallis tests were repeated for two

recommended psychotherapy groups separately to examine within group differences of effect sizes on a particular meta-analytic characteristic. As before, data were initially pooled and analyzed for both low and high treatment intensity depression; the test was then repeated controlling for depression treatment intensity.

3.3.5.4 Mann-Whitney tests for pairwise comparisons

Mann-Whitney tests were performed in response to statistically significant differences revealed on Kruskal-Wallis test. As these consisted of multiple sub-categories for each of the meta-analytic characteristics tested, subsequent pairwise comparisons were required to determine specific differences in effect sizes and risk associations between sub-groups within a meta-analytic category (e.g. specific treatment comparator types).

3.3.5.5 Cross tabulation analyses (descriptive analyses)

Correlation analyses were performed at a categorical level to investigate the relationship between recommended psychotherapy group (i.e. CBT and other psychotherapies) and outcome. Two-by-two (2x2) contingency tables were constructed in order to examine the relationship between recommended psychotherapy group and outcomes (i.e. effect-sizes or relative risks). On a descriptive level this provided frequency counts of effect size magnitudes (in addition to overall effectiveness) and relative risks within each psychotherapy group and between the two psychotherapy groups. Moreover, it enabled observation of both within and between group differences.

Contingency tables were constructed in different ways for continuous outcomes (i.e. effect sizes) and dichotomous outcomes (i.e. risk associations). For continuous data,

recoded into magnitudes of effect (i.e. none, small, medium and large; see section 3.3.4.2), these were organized into contingency tables to examine overall effectiveness (i.e. none x all magnitudes) and relative magnitude (i.e. small vs medium/large). This was done in response to initial examination of the data, namely disparities between magnitudes of effects (see results chapter). For dichotomous data, recoded into risk likelihood (less or more; see section 3.3.4.3), the contingency tables were constructed in a way that was consistent with original categories.

In order to further explore the relationship between the two groups of psychotherapy and outcomes (effect size magnitudes/ risk associations), meta-analytic characteristics described earlier (see section 3.3.3.4 – 3.3.3.6) were controlled within cross tabulation analyses. Through exploring the data a hypothesis was emerging that these characteristics could provide an alternative explanation to any observed differences in outcomes: rather than straightforward differences between recommended psychotherapies, these characteristics could potentially act as sources of bias in the interpretation of evidence and subsequent recommendations.

Of course, study characteristics have previously been shown to interact with outcome in primary studies (i.e. treatment comparators used in primary RCTs), and thus are integral to reviewing evidence as part of the guideline development process. However, other characteristics are more specific to the meta-analyses performed as part of the guideline development process (i.e. quality outcome ratings of evidence). In these instances effect size magnitudes and risk associations could be a direct consequence of the conditions of the meta-analysis performed within the guideline development process! Therefore, characteristics of the guideline meta-analyses, which are usually considered noise or mess within guideline reviews, are of particular interest and of relevance to reviewing the strength of evidence presented for recommended psychotherapies.

A number of contingency tables were constructed to explore the interaction between meta-analytic characteristics, effect size and psychotherapy group. The contingency tables constructed were in relation to meta-analytic characteristics that data were extracted for (see sections 3.3.3.4 – 3.3.3.6), namely treatment comparators, depression intensity and GRADE quality ratings. For relative risk outcomes (dichotomous measures), the type of risk was an additional characteristic of interest, which a series of contingency tables were constructed for.

3.3.5.6 Fisher's exact test

Fisher's exact tests (FET), two-tailed, were performed to measure independence between psychotherapy group and effect size magnitude/ relative risks. This was repeated for each contingency table that controlled for the aforementioned meta-analytic characteristics and to examine their interaction with psychotherapy group and effect. FET was selected as the appropriate test from the family of chi-square tests due to the sample that the data was drawn from not meeting the assumption of normal distribution. Furthermore, many of the contingency tables controlling for specific meta-analytic characteristics contained small sample sizes.

3.4 My position as researcher

Similar to characteristics of a qualitative paradigm, within this project I view the data encountered as more of a dynamic reality constructed through the process of primary research conducted and secondary review through guideline development process. The process of generating evidence that was interpreted by researchers and reviewers prior to my own explorations and subsequent findings adds further to this dynamic reality. My perspective and actions throughout the research process are influenced by my professional and world view, and my theoretical allegiances. I

am aware that these guided my decisions at key steps of the research procedure. These decisions were made as an actor within the research process (Burr, 2003) and inevitably influence the dynamic reality of the data further.

My perspective and actions throughout the research process are influenced by my professional and world view and my own theoretical allegiances. Self-reflexivity is integral to qualitative research approaches (Alvesson and Skoldberg, 2000). Although this is not typically applied to quantitative methodology it is of particular relevance to this exploratory research project. As Elliot et al. (1999) suggest it was necessary for me to 'own' my perspective within the research by recognizing my values, interests and assumptions and the role these preconceptions may have had in influencing how the data was gathered, interpreted, and presented (Tufford and Newman, 2012).

I am a thirty-five year-old black male; a second-generation Ghanaian. In early childhood my family home was in Walthamstow, a working class north east London suburb with a relatively diverse ethnic composition. When I was eight the family moved to Chingford, which was predominantly a white working class town. I grew up between multiple identities and contexts in respect to my ethnicity, culture, education and religion (Roman Catholic), family, social and personal situation. Reflecting on these experiences I recognise a complex interplay between what ecological systems theorists term 'proximal processes' (Bronfenbrenner & Morris, 2006), various ecological systems (or contexts) and how I navigated various challenges I encountered in both adaptive and maladaptive ways.

Due to my personal experiences I embraced the social constructionist course philosophy on embarking on clinical training. Social constructionism as described by Burr (2005) encourages a critical stance towards taken-for-granted knowledge; a historical and cultural specificity to understanding; knowledge sustained by social processes; and knowledge and social action in synchrony. When applied to

psychotherapy research, social constructionism questions the traditional scientific paradigm firmly committed to the formulation of research questions that can be empirically pursued by quantitative measurements in a hypothetico-deductive framework of experimentation with individuals and groups. The knowledge that is gained through this process is viewed as representative of facts waiting to be found or discovered that are separate from the actions of the researchers (Lock and Strong, 2010).

Prior to clinical training my work experiences were a contrast to the course philosophy. I worked for five years within a research network supporting large-scale NHS research trials, which promoted evidence-base medicine (EBM) models as the gold standard for evaluating the effectiveness of psychological interventions. The complexities (and limitations) inherent within adopting an empirical approach to psychotherapy outcome research became apparent to me in respect to various stages of the research process. I observed considerable disconnect between psychotherapy researchers (including research support staff), service managers and clinicians providing routine clinical practice and service users who were targeted for recruitment into trials. I was concerned that there was too much emphasis on psychotherapy outcome research within the portfolio, which taken on its own, paints a very incomplete picture of what actually occurs within primary research studies. I was aware of the wider context of government policy and strategy to create an infrastructure for 'high quality' mental health research that has the ability to penetrate healthcare policy.

I gradually became jaded by the daily rigours of recruitment, assessment and administration on clinical trials and jumped at the opportunity to train as a research therapist for a trial of CBT adapted for health anxiety. However, even within this position of treatment delivery I began to question whether an RCT approach warranted the 'gold-standard' tag within the complicated terrain of psychotherapy.

The University of Hertfordshire's (UH) doctorate in clinical psychology embodies a course philosophy that takes a critical stance towards dominant evidence-based paradigms such as those adopted by guidelines, namely on the grounds that they lack a pluralistic approach to evidence and reality that subsequently prioritises some forms of psychotherapy over others. This stimulated my interest in examining NICE's evidence base from a social constructionist perspective. Furthermore, my interest in this connected to my previous work experiences of contributing to an evidence-based research infrastructure. My curiosity developed further out of discussions with my supervisor, who had a similar interest in this area.

The process of developing and carrying out this research generated mixed feelings throughout. I came to this project sceptical about evidence supporting NICE guidelines and keen to uncover underlying sources of bias that disadvantage the type of therapeutic pluralism that attracted me to the profession and more specifically to the University of Hertfordshire doctoral course. I was also daunted by the prospect of challenging the evidence presented and interpreted within an organisation so well structured and resourced and comprising of experts in the area.

My attempts to understand the guideline development process involved extensively reading NICE documents, and articles written by proponents of guidelines that place guidelines into context. I began to feel somewhat in awe by what had been achieved through the NICE guidelines that had been developed. Furthermore, the literature from dissenting camps, which had initially fuelled my motivation to question the evidence base, faded somewhat in weight and significance. However, there remained a dissonance between what guidelines purportedly achieved through the evidence reviewed and my experiences delivering problem-based psychotherapies prior to and over the course of my clinical training. Furthermore, I was drawn to alternative psychotherapy research paradigms and evidence for a spectrum of

common factors inherent in psychotherapy that are less privileged within evidence-based narratives.

This provided a rationale for exploring whether or not differences exist between the evidence base for CBT (granted frontline treatment status within the NICE depression guideline) and the aggregated evidence base for other psychotherapies not afforded the same status. Although an adversarial element remains within my study design (i.e. CBT 'versus' other psychotherapies), my focus shifted away from purely exploring the data for differences towards looking for patterns within CBT and other psychotherapy data sets that would enable a clearer understanding of their interaction with meta-analytic characteristics.

I was mindful that my position throughout this research process was similar to the challenge that Boltanski and Thevenot (cited in Moreira, 2007) describe in respect to sociological analyses that 'focus on how actors draw upon common modes of judgement to orientate their involvement in disputes' (pg. 3). Therefore, I was mindful of the need to reflect on my role as a social 'actor' in constructing and judging the strength of evidence that I was presented with. This required me to remain open-minded and to attempt to balance my own perspective and assumptions throughout review of evidence.

4. Results

4.1 Chapter Outline

As outlined in the previous chapter, this project was interested in the strength of secondary evidence presented for CBT-based interventions and other psychotherapies recommended by NICE for the treatment of depression in adults. The primary variable of interest is differences in effect sizes and risk associations between the groups of recommended psychotherapies (CBT and other psychotherapies). As the process of exploratory data analysis developed, so did my interest in interaction between recommended psychotherapies, the evidence and various meta-analytic characteristics. Thus, it is important to also understand the data within the contexts of the review characteristics.

This chapter is split into two parts. The first part summarises the findings of continuous outcomes (i.e. effect sizes) for the two groups of recommended psychotherapies. The second part summarises the findings of dichotomous outcomes (i.e. risk associations). Both parts of the chapter begin with a summary of the descriptive statistics and statistical tests of difference for the overall raw outcomes for two groups of recommended psychotherapies. These are followed by summaries of the findings from further post hoc analyses, where carried out in response to the prior findings. Findings are then presented for the categorical data. This consists of cross tabulation analyses of the overall evidence for each recommended psychotherapy group and tests of independence. This is followed by a summary of findings from further cross tabulation analyses used to explore the interaction of psychotherapy group, evidence strength and the different review characteristics of interest.

PART ONE: Evidence on continuous outcomes

4.2 Raw effect size analyses

The CBT group (*N* = 130) was associated with a mean effect size *M* = -0.16 (*SD* =0.31). By comparison the other psychotherapies group (*N* = 93) was associated with a numerically larger mean effect size *M* = -0.29 (*SD* = 0.50). An independent samples t-test was performed to test for a difference between the mean effect sizes for CBT and other psychotherapies (see Appendix 1), and showed that the other psychotherapies group was associated with larger effect sizes (for the reduction of depression) than the CBT group[1], $t(143.37) = 2.224$, $p = .026$. A significant difference between the two groups was maintained when favourable raw effect size values (i.e. 0.2 or above) were analysed separately, $t(66.419) = 4.457$, $p < 0.01$ (see Appendix 1)

When the t-tests were repeated separately for each quality rating (see Appendix 1), a significant difference between the mean effects for the two psychotherapy groups was maintained for moderate and high quality outcome ratings for both combined, high and low treatment intensity.

4.2.1 Post hoc analysis

Post hoc analyses were performed in order to determine whether the significant difference between recommended psychotherapies was influenced by specific meta-analytic features and to explore the influence of 'actual' psychotherapies on the mean effect size within the other psychotherapies group.

[1] A negative value indicates a superior effect size and the larger the negative value the greater the effect size.

4.2.1.1 Treatment intensity

Although a similar trend of larger effect sizes was observed within other psychotherapies than the CBT group when controlling for treatment intensity, the independent t-tests revealed no significant difference between the two recommended psychotherapies for either high or low intensity depression (see Appendix 2). However, there was a significant difference between the two psychotherapy groups when favourable effects (i.e. 0.2 and above) were analysed separately for both high and low intensity depression. This indicated larger mean effects within the other psychotherapies group.

4.2.1.2 T-tests (between psychotherapy group differences based on treatment comparators)

A series of post hoc t-tests were performed in order to examine differences between the two psychotherapy groups controlling for treatment comparators; these were done in order to examine whether differences between the mean effects occurred as a function of the class of interventions that the recommended psychotherapies were compared against (see Appendix 3). For active interventions, the difference in mean effects between the two recommended psychotherapies overall (i.e. for low and high intensity depression combined) was approaching significance $t(50) = 2.005$, $p = .050$. This indicated that the other psychotherapies group was associated with larger mean effect sizes ($M = -0.1478$; $SD = 0.074$) than the CBT group ($M = 0.0138$; $SD = 0.041$). There were no significant differences for any of the remaining overall mean effects of the two psychotherapy groups based on treatment comparators.

Separate analysis of high and low intensity depression based on each treatment comparator revealed a significant difference in the mean effect sizes within the

medication sub-group for high intensity depression[2], where $t(31.857) = -3.275$, $p = -0.003$. This indicated that the CBT group was associated with larger mean effects ($M = 0.30$; $SD = 0.06$) than the other psychotherapies group ($M = 0.13$; $SD = 0.04$). A further significant difference was found within high intensity mean effect sizes based on comparisons between recommended psychotherapy combined with medication and medication alone $t(21) = -2.178$, $p = 0.041$. This indicated that the CBT group was associated with larger mean effects (M = -.2719; SD = .07178) than other psychotherapies group ($M = .0500$; $SD = .40776$). There were no other significant differences between the two recommended psychotherapies based on treatment comparators when depression intensity was analysed separately.

4.2.1.3 Kruskal-Wallis and Mann-Whitney tests for within group differences based on treatment comparators

In order to examine the overall differences in effect sizes between treatment comparators within recommended psychotherapies, a Kruskal-Wallis test (see Appendix 4) was performed on effect sizes based on the four treatment comparator conditions for the two psychotherapy groups combined[3]. The test, which was corrected for tied ranks, was statistically significant $X^2(2, N = 217) = 31.392$, $p < .001$.

A series of Mann-Whitney tests were performed to examine pairwise differences between treatment comparators (see Appendix 4). A Bonferroni correction of $p = .0083$ as the threshold value was applied to the alpha level. This indicated that recommended psychotherapy effect sizes were significantly greater when the psychotherapy was compared to non-active interventions than when compared to the other three treatment comparator groups. None of the other pairwise

[2] As there was no meta-analytic evidence for low intensity CBT compared to medication tests of differences were not computed.

[3] A seventh treatment comparator group, vs medication/ non-active intervention was excluded due to the considerably small sample (N = 6) relative to the other five groups.

comparisons of treatment comparators indicated a significant difference (see Appendix 4).

4.2.1.4 Kruskal-Wallis and Mann-Whitney tests based on treatment comparators controlling for treatment intensity

The Kruskal-Wallis test was repeated controlling for treatment intensity (high and low intensity depression), which revealed a statistically significant difference for high intensity depression only, $X^2(2, N = 148) = 11.971$, $p = .007$. A series of Mann-Whitney tests were performed to examine pairwise differences in treatment comparators further to the statistical differences observed for high-intensity depression on the Kruskal-Wallis test (see Appendix 5). A Bonferroni correction of $p = .0083$ as the threshold value was applied to the alpha level. Pairwise comparisons indicated that recommended psychotherapy effect sizes were significantly greater when the psychotherapy was compared to non-active comparators (*Mdn* = 26.91) than when compared to active treatment comparators (*Mdn* = 41.23), $U = 357.50$, $p = .004$.

A second pairwise comparison indicated that the recommended psychotherapies' effect sizes were significantly greater when the therapies were compared to non-active comparators (*Mdn* = 24.32) than when compared to medication (*Mdn* = 37.41), $U = 275.00$, $p = .004$. This reveals a pattern of lower mean ranks for the non-active comparators, indicating larger effect sizes for recommended psychotherapies against this treatment-comparator condition than against other treatment-comparator groups[4]. None of the other pairwise comparisons indicated a significant difference between any of the remaining treatment comparators (see Appendix 5).

[4] Lower ranks are indicative of larger effect sizes due to the negative values indicating a reduction in depression scores.

4.2.1.5 Kruskal-Wallis and Mann-Whitney tests for within group differences based on treatment comparators (controlling for recommended psychotherapy group)

The Kruskal-Wallis test was repeated controlling for recommended psychotherapy group (CBT and other psychotherapies; see Appendix 6). This indicated a statistically significant difference between the effect sizes within the 'other psychotherapies' group for high intensity depression as a function of treatment comparators, $X2(2, N = 55) = 17.086$, $p = .001$[5]. Therefore, Mann-Whitney tests were performed to evaluate pairwise differences between treatment comparators within the other psychotherapies group for high intensity depression (see Appendix 6). Bonferroni correction of $p = .0083$ as the threshold value was applied to the alpha level. Pairwise comparisons indicated that within the other psychotherapies group effect sizes were significantly greater when the psychotherapies were compared against non-active comparators (*Mdn* = 10.72) than when compared against medication comparator (*Mdn* = 22.00), $U = 22.000$, $p = .001$. A further pairwise comparison indicated that the effect sizes within the other psychotherapies group were significantly greater when based on active comparators (*Mdn* = 10.62) than effect sizes based on medication comparators (*Mdn* = 20.50), $U = 27.500$, $p = .002$.

4.2.1.6 Actual psychotherapies

Further post hoc analyses were performed within the 'other psychotherapies' group for high-intensity depression[6]; this was in order to explore whether there were

[5] A statistically significant difference was found between non-active and active interventions within the CBT group for low intensity depression. However, as there was no other meta-analytic data for the other treatment comparators, this equated to a pairwise comparison and therefore was not analysed any further. There was no statistical difference between treatment comparators for low intensity depression within the other psychotherapies group.

[6] This analysis was not performed for low-intensity psychotherapy as the only treatment within this group was physical activity.

differences between the effect sizes for specific (or actual) types of psychotherapies that this group comprised of, namely behavioural activation, couples therapy, interpersonal therapy, counselling and short-term psychodynamic psychotherapy. A Kruskall-Wallis test was conducted to explore these differences (see Appendix 7). The test, which was corrected for tied ranks, was statistically significant, $X^2(2, N= 55) = 11.88$, $p = 0.018$.

Mann-Whitney tests were performed to evaluate pairwise differences between actual psychotherapies within the 'other psychotherapies' group (see Appendix 7). However, when a Bonferroni-correction was applied to the alpha (using .005 as the threshold value), the p-values indicated that there were no significant differences between any of the 'actual' psychotherapies within the 'other psychotherapies' groups.

Summary of raw effect size analyses

The other psychotherapies group was associated with larger mean effect sizes than the CBT group. Further post hoc analysis revealed a statistically significant difference for treatment comparators within the high intensity depression treatment effect sizes. The CBT group was associated with larger effect sizes than other psychotherapies when compared to medication comparators. Effect sizes for recommended psychotherapies were significantly greater when the psychotherapies were compared to non-active interventions than when they were compared to active treatments. Effect sizes for recommended psychotherapies were also significantly greater when the psychotherapies were compared to non-active interventions than when compared to medication. When controlling for recommended psychotherapy group (i.e. CBT and other psychotherapies), this indicated a significant difference between treatment comparators in the other psychotherapies group only. Pairwise differences within this group revealed that

other psychotherapies' effect sizes were greater when the psychotherapies were compared to non-active comparator than when compared to a medication comparator. Furthermore, effect sizes were significantly greater when the psychotherapies were compared to active treatment comparators than when compared to medication comparators.

4.3 Overall effectiveness of recommended psychotherapies (categorical data)

Tables 4.3 show the proportion of the evidence by effect size for the two groups of recommended psychotherapies (CBT and other psychotherapies) that were presented within the evidence profile for depression guidelines. Section A (table 4.3) shows the proportion of effect sizes for the high and low intensity depression treatments combined. This reveals that a larger number of meta-analyses were performed for the CBT group (*N* = 130) than for the other psychotherapies group (*N* = 93).

Section A (Table 4.3) shows the proportion of meta-analyses within the two groups (CBT and other psychotherapies) that were classified as 'none' effects and the number of meta-analyses that indicated a favourable effect of any magnitude. A comparable proportion of none effects were observed for CBT and other psychotherapies (51.5% and 50.5% respectively). Therefore, the effect sizes that were either equivalent to a treatment comparator, minimally different or inferior accounted for approximately half of the meta-analytic data in both the CBT and the other psychotherapies group.

Table 4.3 Proportion of none-effects + favourable effects (small, medium, large) for recommended psychotherapies

Section A. High and Low Intensity Combined

			Cohen's Effect size Categories		
			none	small, medium, large	Total
Recommended Psychotherapy	CBT	Count	67	63	130
		% within Recommended Psychotherapy	51.5%	48.5%	100.0%
	Other psychotherapies	Count	47	46	93
		% within Recommended Psychotherapy	50.5%	49.5%	100.0%
Total		Count	114	109	223
		% within Recommended Psychotherapy	51.1%	48.9%	100.0%

Fisher's Exact, p = 0.89 (Non-significant)

Section B. High Intensity

			Effectiveness		
			none	small, medium, large	Total
Recommended Psychotherapy	CBT	Count	58	41	99
		% within Recommended Psychotherapy	58.6%	41.4%	100.0%
	Other psychotherapies	Count	34	21	55
		% within Recommended Psychotherapy	61.8%	38.2%	100.0%
Total		Count	92	62	154
		% within Recommended Psychotherapy	59.7%	40.3%	100.0%

Fisher's Exact, p = 0.73 (Non-significant)

Section C. Low intensity

			Effectiveness		
			None	small, medium, large	Total
Recommended Psychotherapy	CBT	Count	9	22	31
		% within Recommended Psychotherapy	29.0%	71.0%	100.0%
	Other psychotherapies	Count	13	25	38
		% within Recommended Psychotherapy	34.2%	65.8%	100.0%
Total		Count	22	47	69
		% within Recommended Psychotherapy	31.9%	68.1%	100.0%

Fisher's Exact, p = 0.80 (Non-significant)

Section B (Table 4.3) shows the proportion of effect sizes for high intensity depression. Within the two groups of recommended psychotherapies there is an increased proportion of 'none' effect sizes for CBT and other psychotherapies (58.6% and 61.8% respectively) relative to the proportion of favourable effects of any magnitude (41.4% and 38.2% for CBT and other psychotherapies respectively).

Section C (Table 4.3) shows the proportion of effect sizes for low intensity depression. Within the two groups of recommended psychotherapies there is an increased proportion of evidence that show a favourable effect of any magnitude for CBT and other psychotherapies (71% and 65.8% retrospectively).

Fisher's exact test of independence revealed that there was not a significant relationship between psychotherapy group (CBT and other psychotherapies) and effectiveness (i.e. 'none' effects and favourable effect by any magnitude) for high intensity depression treatments, low intensity depression treatments or for the two treatment intensities combined (see Table 4.3). Thus, favourable effects were no more likely in either the CBT or other psychotherapies groups.

Summary of overall effectiveness of recommended psychotherapies

A considerable volume of evidence was presented in the guideline for the CBT group and a comparable volume of evidence was available by pooling all of the other recommended psychotherapies together. The cross-tabulation analysis of recommended psychotherapies revealed comparable proportions of favourable and none effects within both the CBT and other psychotherapies group for the overall evidence. For high intensity depression there was an increased proportion of none effects for each psychotherapy group. For low intensity depression increased proportions of favourable effects were present in both psychotherapy groups. There were no significant relationship between psychotherapy and effect for the overall evidence or within either treatment intensity.

4.3.1 Overall effect size magnitude

Table 4.3.1 shows the proportion of evidence for the two psychotherapy groups by effect size magnitude. Section A (Table 4.3.1) shows the proportion of effect size magnitudes for treatment intensity combined. Within the CBT group the proportion of small effects accounted for more than 81% of the evidence in favour of the intervention. This was a considerably higher proportion than the 41% categorised as small effect size within the other psychotherapies group. In contrast, within the other psychotherapies group a higher proportion of medium and large[7] effects wasobserved, accounting for 58.7% of the evidence in favour of the intervention. This indicates a reverse pattern in proportion of effect size magnitudes where a greater proportion of small effects exists within the CBT group and there is a greater proportion of evidence of a medium or large effect within the other psychotherapies group.

[7] As medium and large effects comprised a relatively small proportion within both groups they were pooled together for analysis.

Fisher's exact test of independence revealed a significant relationship between effect size magnitude and psychotherapy for combined high and low intensity depression treatment.

Table 4.3.1 Proportion of effects by magnitude (small vs medium/ large) for recommended psychotherapies

Section A. Combined Intensity (high and low)

			Effect size magnitude		
			small	medium or large	Total
Recommended Psychotherapy	CBT	Count	51	12	63
		% within Psychotherapy	81.0%	19.0%	100.0%
			72.8%	30.7%	57.8%
	Other psychotherapies	Count	19	27	46
		% within Psychotherapy	41.3%	58.7%	100.0%
		% between Psychotherapy	27.14%	69.23%	42.2%
Total		Count	70	39	109
		% within Psychotherapy	64.2%	35.8%	100.0%
			100.0%	100.0%	100.0&

Fisher's Exact, p = 0.00002 (Significant)

Section B. High Intensity

			Cohen's with two groups		
			small	medium or large	Total
Recommended Psychotherapy	CBT	Count	37	4	41
		% within Psychotherapy	90.2%	9.8%	100.0%
		% between Psychotherapy	78.7%	26.6%	66.1%
	Other psychotherapies	Count	10	11	21
		% within Psychotherapy	47.6%	52.4%	100.0%
		% between Psychotherapy	21.2%	73.3%	33.8%
Total		Count	47	15	62
		% within Psychotherapy	75.8%	24.2%	100.0%
		% between Psychotherapy	100.0%	100.0%	100.0%

Fisher's Exact, p = 0.00042 (Significant)

Section C. Low Intensity

			Cohen's with two groups		
			small	medium or large	Total
Recommended Psychotherapy	CBT	Count	14	8	22
		% within Recommended Psychotherapy	63.6%	36.4%	100.0%
	Other psychotherapies	Count	9	16	25
		% within Recommended Psychotherapy	36.0%	64.0%	100.0%
Total		Count	23	24	47
		% within Recommended Psychotherapy	48.9%	51.1%	100.0%

Fisher's Exact, p = 0.31 (Non-significant)

A similar pattern was observed in the proportion of effect sizes within the two psychotherapy groups when controlling for high intensity depression (Section B of table 4.3.1). Within the CBT group there is a noticeable difference in the proportion of small effect sizes and the proportion of medium/ large effect sizes (90.2% and 9.8% respectively). Within the other psychotherapies group a similar proportion of small and medium/ large effects is observed as those for combined depression treatment intensity (47.6% and 52.4% respectively). Fisher's exact test of independence revealed a significant relationship between effect size magnitude and psychotherapy for high intensity depression.

For low intensity depression (Section C of table 4.3.1), the proportions within the two groups were almost reversed, where there was a greater proportion of small effect sizes within the CBT group (63.6%) and a greater proportion of medium/ large effect sizes within the other psychotherapies group (64%). However, Fisher's exact test revealed no significant relationship between psychotherapy and effect magnitude for low intensity depression.

Summary of effect size magnitudes

The cross-tabulation analysis of effect size magnitudes revealed increased proportions of small effects compared to medium/ large effects within the CBT group. Within the other psychotherapies group a greater proportion of medium/ large effects was observed than the proportion of small effects. This pattern was maintained when treatment intensity (high and low intensity) was controlled for. Furthermore, a statistically significant difference between the psychotherapy groups for effect size magnitudes was revealed for combined low intensity and high intensity depression.

4.3.2 Treatment comparators (categorical data)

The overall proportions of effect sizes and the proportion of the effect size magnitude only tell part of the story. In order to understand the overall differences observed in the evidence for CBT and other psychotherapies it was necessary to examine the evidence in relation to the treatment comparators. The following sub-sections in 4.3.2 examine this.

4.3.2.1 Psychotherapies compared to non-active interventions (effectiveness)

Table 4.3.2.1 shows the proportion of effects ('none' and favourable effects) for CBT and other psychotherapies compared to non-active treatments. Section A (Table 4.3.2.1) reveals a greater proportion of favourable effects to 'none' effect sizes within the CBT group (70.3% to 29.7% respectively). Within the other psychotherapies group there is a similar proportion of favourable effects to 'none' effects (73.7% and 27.3% respectively).

Section B (Table 4.3.2.1) reveals a contrasting pattern of effects within the two psychotherapy groups for high intensity depression. Within the CBT group a greater proportion of none to favourable effects was observed (60% and 40% respectively) whereas within the other psychotherapies group a greater proportion of favourable to none effects is observed (72.7% and 27.3% respectively). However, Fisher's exact test revealed there to be no significant relationship between psychotherapy and effect when compared to non-active interventions for high or low intensity depression or for the two combined.

Table 4.3.2.1 Proportion of effect sizes for recommended psychotherapies vs non-active comparator (effectiveness)

Section A. Combined Intensity (proportion of none + favourable effects)

			Effectiveness		
			none	small, medium, large	Total
Recommended Psychotherapy	CBT	Count	11	26	37
		% within Recommended Psychotherapy	29.7%	70.3%	100.0%
	Other psychotherapies	Count	12	32	44
		% within Recommended Psychotherapy	27.3%	72.7%	100.0%
Total		Count	23	58	81
		% within Recommended Psychotherapy	28.4%	71.6%	100.0%

Fisher's Exact, p = 0.81044 (Non-significant)

Section B. High Intensity (proportion of none + favourable effects)

			Effectiveness		
			none	small, medium, large	Total
Recommended Psychotherapy	CBT	Count	6	4	10
		% within Recommended Psychotherapy	60.0%	40.0%	100.0%
	Other psychotherapies	Count	5	13	18
		% within Recommended Psychotherapy	27.8%	72.2%	100.0%
Total		Count	5	13	18
		% within Recommended Psychotherapy	27.8%	72.2%	100.0%

Fisher's Exact, p = 0.12453 (Non-significant)

Section C. Low Intensity (proportion of none + favourable effects)

			Effectiveness		
			none	small, medium, large	Total
Recommended Psychotherapy	CBT	Count	5	22	27
		% within Recommended Psychotherapy	18.5%	81.5%	100.0%
	Other psychotherapies	Count	7	19	26
		% within Recommended Psychotherapy	26.9%	73.1%	100.0%
Total		Count	7	19	26
		% within Recommended Psychotherapy	26.9%	73.1%	100.0%

Fisher's Exact, p = 0.52558 (Non-significant))

4.3.2.2 Psychotherapies compared to non-active interventions (effect size magnitude)

Section A (Table 4.3.2.2) shows the proportion of small and medium/large effects for CBT and other psychotherapies compared to non-active treatments for combined intensity depression. Within the other psychotherapies group there was an increased proportion of medium/large effects to small effect sizes (59.4% to 40.6% respectively). The opposite pattern is observed in the CBT group, where there was a greater proportion of small to medium/ large effect sizes (65.4% to 34.8% respectively).

Section B (Table 4.3.2.2) shows a similar pattern within the CBT group for high intensity depression, where the proportion of small to medium/ large effects observed is 75% and 25% respectively. However, this is based on a small number of meta-analyses. Within the other psychotherapies group there is a more comparable proportion of effects with a slightly greater proportion of medium/ large to small effects (53.8% and 46.2% respectively). The same pattern is observed in the low intensity data set (Section C of table 4.3.2.2), where the proportion of large/ medium

effects is 63.2% within the other psychotherapies group whereas within the CBT group the proportion of small effects observed was 63.6%.

Fisher's exact test of independence revealed that there was no relationship between psychotherapy and the magnitude of effects when compared to non-active interventions. The relation between these variables was not significant for high or low intensity depression or for the two combined.

Table 4.3.2.2 Proportion of effect sizes for recommended psychotherapies vs non-active comparator (effect size magnitude)

Section A. Combined Intensity (small vs medium/ large)

			Effect size magnitudes		
			small	medium or large	Total
Recommended Psychotherapy	CBT	Count	17	9	26
		% within Recommended Psychotherapy	65.4%	34.8%	100.0%
	Other psychotherapies	Count	13	19	32
		% within Recommended Psychotherapy	40.6%	59.4%	100.0%
Total		Count	30	28	58
		% within Recommended Psychotherapy	51.7%	48.3%	100.0%

Fisher's Exact, p = 0.071 (Non-significant)

Section B. High Intensity

			Effect size magnitudes		
			small	medium or large	Total
Recommended Psychotherapy	CBT	Count	3	1	4
		% within Recommended Psychotherapy	75.0%	25.0%	100.0%
	Other psychotherapies	Count	6	7	13
		% within Recommended Psychotherapy	46.2%	53.8%	100.0%
Total		Count	9	8	17
		% within Recommended Psychotherapy	52.9%	47.1%	100.0%

Fisher's Exact, p = 0.58 (Non-significant)

Section C. Low Intensity

			Effect size magnitudes		
			small	medium or large	Total
Recommended Psychotherapy	CBT	Count	14	8	22
		% within Recommended Psychotherapy	63.6%	36.4%	100.0%
	Other psychotherapies	Count	7	12	19
		% within Recommended Psychotherapy	36.8%	63.2%	100.0%
Total		Count	21	20	41
		% within Recommended Psychotherapy	51.2%	48.8%	100.0%

Fisher's Exact, p = 0.12(Non-significant)

4.3.2.3 Psychotherapies compared to active interventions (effectiveness)

Table 4.3.2.3 shows the proportion of 'none' effects and small effect sizes for the meta-analyses performed when the two groups of recommended psychotherapies were compared with active interventions. (This included some direct comparisons between CBT and other psychotherapies recommended within the guideline).

Section A (Table 4.3.2.3) shows that within both recommended psychotherapy groups there was an increased proportion of 'none' effects when compared with another active intervention. Within the other psychotherapies groups the proportion of none to small effects was 70% to 30% respectively. Within the CBT group the proportion of 'none' to small effects was 82.8% and 17.2% respectively.

A similar trend was observed controlling for high intensity depression treatments where within the CBT group an increased proportion of 'none' to small effects was observed (80% and 20% respectively) and similarly increased proportions were observed within the other psychotherapies group.

Table 4.3.2.3 Proportion effect sizes for recommended psychotherapies vs active/ psychotherapy comparator (effectiveness)

Section A. Combined Intensity (High and Low)

			Effectiveness		
			none	small	Total
Recommended Psychotherapy	CBT	Count	24	5	29
		% within Recommended Psychotherapy	82.8%	17.2%	100.0%
	Other psychotherapies	Count	14	6	20
		% within Recommended Psychotherapy	70.0%	30.0%	100.0%
Total		Count	38	11	49
		% within Recommended Psychotherapy	77.6%	22.4%	100.0%

Fisher's Exact, p = 0.32(Non-significant)

Section B. High Intensity

			Effectiveness		
			none	small	Total
Recommended Psychotherapy	CBT	Count	20	5	25
		% within Recommended Psychotherapy	80.0%	20.0%	100.0%
	Other psychotherapies	Count	11	4	15
		% within Recommended Psychotherapy	73.3%	26.7%	100.0%
Total		Count	31	9	40
		% within Recommended Psychotherapy	77.5%	22.5%	100.0%

Fisher's Exact, p = 0.71(Non-significant)

Section C. Low Intensity

			Effectiveness		
			none	small	Total
Recommended Psychotherapy	CBT	Count	4	0	4
		% within Recommended Psychotherapy	100.0%	0.0%	100.0%
	Other psychotherapies	Count	3	2	5
		% within Recommended Psychotherapy	60.0%	40.0%	100.0%
Total		Count	7	2	9
		% within Recommended Psychotherapy	77.8%	22.2%	100.0%

Fisher's Exact, p = 0.44(Non-significant)

Section C (Table 4.3.2.3) shows the proportion of none to favourable effects for low intensity depression treatments. Within the CBT group it is observed that all of the evidence was categorised as none effects. Within the other psychotherapies group there is a marginal difference in the proportion of none to small effects (60% and 40% respectively). However, due to the small amount of evidence available within the low intensity depression group these proportions are difficult to interpret.

Fisher's exact tests performed to examine the relationship between psychotherapy and effects were non-significant when the psychotherapy was compared to active interventions for combined, high or low intensity depression.

4.3.2.4 Psychotherapies compared to active interventions (effect size magnitudes)

Table 4.3.2.4 shows the proportion of small and medium/large effect sizes when the CBT and other psychotherapies groups were compared with active interventions. Section A shows a small amount of meta-analytic data available, which makes the proportion of effect sizes difficult to interpret.

Section B (Table 4.3.2.4) shows the proportion of effect magnitudes controlling for high intensity depression. Within the CBT group the entire evidence is categorised as small effects. Within the other psychotherapies group the proportion of small effects to medium/ large is 66.7% to 33.3% respectively. Fisher's exact test revealed no significant relationship between the variables psychotherapy and effect size magnitude.

There was no meta-analytic data within the CBT group for low intensity depression and only three meta-analyses within the other psychotherapies group for small and medium/ large effects (66.7% and 33.3% respectively). Therefore, it was not possible to perform a cross-tabulation analysis.

Table 4.3.2.4 Proportion of relative effect sizes for recommended psychotherapies vs active/ psychotherapy comparator

Section A. Combined Intensity (High and Low)

			Effect size magnitudes		
			small	medium or large	Total
Recommended Psychotherapy	CBT	Count	5	0	5
		% within Recommended Psychotherapy	100.0%	0.0%	100.0%
	Other psychotherapies	Count	6	3	9
		% within Recommended Psychotherapy	66.7%	33.3%	100.0%
Total		Count	11	3	14
		% within Recommended Psychotherapy	78.6%	21.4%	100.0%

Fisher's Exact, p = 0.26 (Non-significant)

Section B. High Intensity

			Effect size magnitudes		
			small	medium or large	Total
Recommended Psychotherapy	CBT	Count	5	0	5
		% within Recommended Psychotherapy	100.0%	0.0%	100.0%
	Other psychotherapies	Count	4	2	6
		% within Recommended Psychotherapy	66.7%	33.3%	100.0%
Total		Count	9	2	11
		% within Recommended Psychotherapy	81.8%	18.2%	100.0%

Fisher's Exact, p = 0.45(Non-significant)

4.3.2.5 Psychotherapies compared to medication (effectiveness and effect size magnitudes)

Table 4.3.2.5 shows the proportion of effect sizes for recommended psychotherapies compared to medication for high intensity depression treatments[8]. Section A (Table 4.3.2.5) shows that there is a noticeable difference between the effect sizes of the two groups. Although the majority of the evidence was categorised as none (61.9%),

[8] There was no meta-analytic data within the CBT group for low intensity depression to perform a cross-tabulation analysis.

the proportion of small effects within the CBT group (38.1%) was considerably higher than the proportion of small effects within the other psychotherapies group (0%).

A Fisher's exact test was performed to examine the relationship between psychotherapy and effect (none vs small) when compared to medication. This revealed a significant relationship between these two variables when recommended psychotherapies are compared to medication.

Section B (Table 4.3.2.5) shows the proportion of none to favourable effects (of any magnitude) for the two recommended psychotherapies. Within the CBT group the proportions of none and favourable effects (56.5% and 43.3% respectively) are relatively more comparable than those observed within the other psychotherapies group (85.7% and 14.3% respectively). However, Fisher's exact test revealed no significant relationship between psychotherapy and effect (none vs favourable effect) when compared to medication.

Table 4.3.2.5 Proportion of effect sizes for recommended psychotherapies vs medication

Section A. Proportion of none and small effects for recommended psychotherapies

			Effectiveness		
			none	small	Total
Recommended Psychotherapy	CBT	Count	13	8	21
		% within Recommended Psychotherapy	61.9%	38.1%	100.0%
			52.0%	100.0%	63.4%
	Other psychotherapies	Count	12	0	12
		% within Recommended Psychotherapy	100.0%	0.0%	100.0%
			48.0%	0.0%	36.3%
Total		Count	25	8	33
		% within Recommended Psychotherapy	75.8%	24.2%	100.0%
			100.0%	100.0%	100.0%

Fisher's Exact, p = 0.030 (Significant)

Section B. Proportion of none effects and favourable effects for recommended psychotherapies

			Cohen's Effect sizes		
			none	small medium large	Total
Recommended Psychotherapy	CBT	Count	13	10	23
		% within Recommended Psychotherapy	56.52%	43.48%	62.16%
	Other psychotherapies	Count	12	2	14
		% within Recommended Psychotherapy	85.71%	14.29%	37.84%
Total		Count	25	12	37
		% within Recommended Psychotherapy	67.57%	32.43%	100.0%

Fisher's Exact, p = 0.083 (Non-significant)

4.3.2.6 Psychotherapies compared to medication (effect size magnitudes)

Table 4.3.2.6 shows the proportion of effect size magnitude (small and medium/ large) controlling for medication comparator. A greater proportion of effect sizes categorised as 'medium/ large' were observed within the other psychotherapies group than those observed the within the CBT group (100% and 20% respectively). However, considering the small numbers within the other psychotherapies group this result should be interpreted with caution. Fisher's exact test revealed no significant relationship between recommended psychotherapy and magnitudes of effects when compared to medication.

Table 4.3.2.6 Proportion of effect size magnitude for psychotherapies vs medication

			Effect size magnitudes		
			small	medium or large	Total
Recommended Psychotherapy	CBT	Count	8	2	10
		% within Recommended Psychotherapy	80.0%	20.0%	100.0%
	Other psychotherapies	Count	0	2	2
		% within Recommended Psychotherapy	0.0%	100.0%	100.0%
Total		Count	8	4	12
		% within Recommended Psychotherapy	66.7%	33.3%	100.0%

Fisher's Exact, p = 0.09 (Non-significant)

4.3.2.7 Psychotherapies combined with medication compared to psychotherapy comparator/ other conditions (effectiveness)

Table 4.3.2.7 shows the proportion of effect (none and small) for recommended psychotherapies and medication combined compared to psychotherapies for high intensity depression treatments[9]. There is a considerable difference in amount of meta-analytic data between the two groups with the vast majority of the meta-analyses being conducted within the CBT group. A greater proportion of the evidence within the CBT group was categorised as 'none' compared to evidence categorised as small effects (83.3% and 16.7% respectively). The single meta-analysis performed within the other psychotherapies group was categorised as a 'none' effect. There were no meta-analyses categorised as having medium/ large effects within either group of recommended psychotherapy.

Fisher's exact test revealed no significant relationships between the variables psychotherapy and effect when combined with medication and compared against psychotherapy.

Table 4.3.2.7 Proportion of relative effect sizes for psychotherapies and medication combined vs psychotherapies/ other conditions

			Effectiveness		
			none	small	Total
Recommended Psychotherapy	CBT	Count	15	3	18
		% within Recommended Psychotherapy	83.3%	16.7%	100.0%
	Other psychotherapies	Count	1	0	1
		% within Recommended Psychotherapy	100.0%	0.0%	100.0%
Total		Count	16	3	19
		% within Recommended Psychotherapy	84.2%	15.8%	100.0%

X^2 = 0.929, p = 0.34 (Non-significant)

Fisher's Exact, p = 1.00(Non-significant)

[9] There was no meta-analytic data for low intensity depression for either psychotherapy group.

4.3.2.8 Psychotherapy and medication compared to medication (effectiveness and effect size magnitude)

Table 4.3.2.8 shows the proportion of effects ('none' and favourable) when the two recommended psychotherapies were combined with medication and compared to medication alone for high intensity depression treatments[10]. Section A (Table 4.3.2.8) reveals an increased proportion of favourable effects to 'none' within the CBT group (81.25% and 18.75% respectively). Within the other psychotherapies group there was a decreased proportion of favourable effects to none (14.29% to 85.71% respectively).

A Fisher's exact test was performed to examine the relationship between psychotherapy (combined with medication) and effect (none vs favourable effect of any magnitude) when compared to medication alone. This revealed a significant relationship between psychotherapy and effect.

The proportions of effect size magnitudes (small vs medium/large) were considered separately (Table 4.3.2.8, Section B). Within the CBT group an increased proportion of small effects was observed compared to medium/ large (92.3% and 7.7% respectively). Within the other psychotherapies group the two medium/large effect sizes accounted for the entire meta-analytic evidence. Although Fisher's exact tests revealed a significant relationship between psychotherapy and effect size magnitude (small vs medium/ large), the small amount of data within the other psychotherapies group makes this finding difficult to interpret.

[10] There was no low intensity meta-analytic data for the CBT group and only three meta-anal in the other psychotherapies group therefore no cross-tabulation analysis was performed.

Table 4.3.2.8 Proportion of relative effect sizes for recommended psychotherapies combined + medication vs medication

Section A. Proportion of relative magnitudes of effects for recommended psychotherapies (High intensity)

			Effectiveness		
			none	small medium large	Total
Recommended Psychotherapy	CBT	Count	3	13	13
		% within Recommended Psychotherapy	18.75%	81.25%	100.0%
			33.33%	92.86%	69.57%
	Other psychotherapies	Count	6	1	2
		% within Recommended Psychotherapy	85.71%	14.29%	100.0%
			66.67%	7.14%	30.43%
Total		Count	11	15	15
		% within Recommended Psychotherapy	39.13%	60.87%	100.0%
			100.0%	100.0%	100.0%

Fisher's Exact, p = .0049 (Significant)

Section B. Proportion of relative magnitudes of effects for recommended psychotherapies (High intensity)

			Cohen's with two groups		
			small	medium or large	Total5
Recommended Psychotherapy	CBT	Count	12	1	13
		% within Recommended Psychotherapy	92.3%	7.7%	100.0%
			100.0%	33.3%	86.6%
	Other psychotherapies	Count	0	2	2
		% within Recommended Psychotherapy	0.0%	100.0%	100.0%
			0.0%	66.6%	13.33%
Total		Count	12	3	15
		% within Recommended Psychotherapy	80.0%	20.0%	100.0%
			100.0%	100.0%	100.0

Fisher's Exact, p = .02857 (Significant)

Summary of treatment comparators (categorical data)

When the recommended psychotherapy groups were compared to non-active treatments, there was a general pattern of increased proportions of small effects compared to medium/ large effect sizes in the CBT group. This was similar to the pattern observed for overall effect sizes (see section 4.3). In contrast, a greater proportion of medium/ large effects sizes was observed within the other psychotherapies group. However, Fisher's exact test did not indicate a relationship between effect size magnitude and recommended psychotherapies when compared to non-active treatments.

A similar pattern was observed when the recommended psychotherapy groups were compared to active interventions. The cross-tabulation analysis generally indicates greater proportions of favourable effects within the other psychotherapies group than those observed within the CBT group. On examination of effect size magnitudes, the proportion of medium/ large effect sizes within the other psychotherapies group was consistently greater than that observed within the CBT group, where greater proportions of small effect sizes were consistently observed. This pattern was generally maintained when controlling for treatment intensity and when the two intensities were combined.

In contrast, when the recommended psychotherapies were compared to medication there were considerably fewer favourable effect sizes of any magnitude observed within the other psychotherapies group compared to those observed within the CBT group. Fisher's exact test indicates a relationship between recommended psychotherapy group and relative effectiveness (i.e. none and small effects) when compared to medication. Therefore, this suggests that guideline evidence for CBT is stronger compared to the evidence for other psychotherapies when the two groups were compared to medication.

Further differences were observed between effect sizes within the two recommended psychotherapy groups when combined with medication and compared to medication alone. Under these conditions, there were statistically significant relationships between psychotherapy and effect, both in terms of effectiveness (none and favourable effects of any magnitude) and effect size magnitude (small and medium/ large). Therefore, this suggests some stronger evidence within the guideline evidence for CBT than other psychotherapies when therapies in the two groups were combined with medication and compared to medication alone.

4.4 Effects by quality rating (categorical data)

Table 4.4 shows the overall effectiveness (none vs favourable effect) within the two psychotherapy groups for each quality rating of evidence. The proportion of effects is presented in separate sections of the table (sections A – C) in respect to the three quality ratings: low, medium and high (as per GRADE ratings within NICE evidence profile).

Section A (Table 4.4) shows the proportion of effects for evidence rated as 'low' quality. Within both psychotherapy groups there was a greater proportion of 'none' effects. However, within the other psychotherapies a greater difference was observed in the proportion of none to favourable effects (70.2% to 29.8% respectively) compared to the proportion observed in the CBT group (58.3% to 41.7% respectively). This indicates that a higher proportion of evidence of a low quality contributes to the overall evidence within the CBT group.

Fisher's exact test was performed to examine the relationship between psychotherapy group and effectiveness (none and favourable effects) when evidence

was rated as low quality. This revealed no significant relation between these variables.

Table 4.4 Proportion of none effects and favourable effects for recommended psychotherapies

Section A. Low Quality Evidence

			Effectiveness		
			none	small, medium, large	Total
Recommended Psychotherapy	CBT	Count	42	30	72
		% within Recommended Psychotherapy	58.3%	41.7%	100.0%
	Other psychotherapies	Count	40	17	57
		% within Recommended Psychotherapy	70.2%	29.8%	100.0%
Total		Count	82	47	129
		% within Recommended Psychotherapy	63.6%	36.4%	100.0%

Fisher's Exact, p = 0.20 (Non-significant)

Section B. Moderate Quality Evidence

			Effectiveness		
			none	small, medium, large	Total
Recommended Psychotherapy	CBT	Count	21	18	39
		% within Psychotherapy	53.8%	46.2%	100.0%
		% between Psychotherapy	75.0%	43.9%	56.5%
	Other psychotherapies	Count	7	23	30
		% within Recommended Psychotherapy	23.3%	76.7%	100.0%
		% between Psychotherapy	25.0%	56.1%	43.4%
Total		Count	28	41	69
		% within Recommended Psychotherapy	40.6%	59.4%	100.0%
		% between Psychotherapy	100.0%	100.0%	100.0%

Fisher's Exact, p = 0.01(Significant)

Section C. High Quality Evidence

			Effectiveness		
			none	small, medium, large	Total
Recommended Psychotherapy	CBT	Count	4	15	19
		% within Recommended Psychotherapy	21.1%	78.9%	100.0%
	Other psychotherapies	Count	0	6	6
		% within Recommended Psychotherapy	0.0%	100.0%	100.0%
Total		Count	4	21	25
		% within Recommended Psychotherapy	16.0%	84.0%	100.0%

Fisher's Exact, p = 0.54 (Non-significant)

Section B (Table 4.4) shows the proportions of effects for recommended psychotherapy groups when the evidence was rated as 'moderate' quality. Within the CBT group a slightly greater proportion of none to favourable effects was observed (53.8% to 46.2% respectively). Interestingly within the other psychotherapies group there was a noticeable difference between the proportion of none and favourable effects (23.3% to 76.7% respectively). This indicates that when evidence is comprised of moderate quality data, the proportion of favourable evidence increases in the other psychotherapies group. In contrast, the proportion of favourable effects observed within the CBT group for low quality evidence (section A, Table 4.4) was maintained for evidence rated as moderate in quality. Fisher's exact test was performed to examine the relationship between psychotherapy group and effectiveness for moderate quality evidence, and revealed a significant relationship between these variables.

Section C (Table 4.4) shows the proportion of effects for evidence rated to be high in quality within the two recommended psychotherapy groups. There is a considerable difference in the amount of meta-analytic evidence within the CBT group compared to the other psychotherapies group (19 and 6 respectively). Within the CBT group a greater proportion of favourable effects to none were observed (78.9% to 21.1% respectively). Within the other psychotherapies groups, favourable effects were

observed for all of the high quality evidence. However, comparisons between the within-group proportions are difficult to interpret further due to the smaller amount of meta-analyses performed within the other psychotherapies group. Fisher's exact test revealed no significant relationship between the variables psychotherapy and effect when the evidence was rated as high in quality.

Summary of effects by quality

A greater proportion of favourable effects was observed within the CBT group for low quality evidence. Within the other psychotherapies group a greater proportion of favourable effects was observed for evidence rated as moderate in quality. Furthermore, there was a significant relationship between the variables psychotherapy and effect when data was moderate in quality. Both psychotherapy groups performed well when the evidence was rated as high in quality.

4.4.1 Effect size magnitude by quality rating

Table 4.4.1 shows the proportions of effect size magnitudes for the CBT and other psychotherapies groups. Section A (Table 4.4.1) shows the proportions for low quality evidence. Small effects comprise the greater proportion compared to medium/ large effects within both the CBT and other psychotherapies group (86.67% and 64.91% respectively)[11]. Fisher's exact test revealed no significant relationship between the variables recommended psychotherapies and effect size magnitude (small and medium/ large) for evidence rated as low quality.

A similar pattern is observed within the CBT group for evidence rated moderate in quality (Section B, Table 4.4.1), where an increased proportion of small effects to

[11] As medium and large effects comprised a relatively small proportion within both groups they were pooled together for analysis.

medium/ large effects (77.78% to 22.22% respectively) was observed. The reverse pattern is observed within the other psychotherapies group, where a greater proportion of medium/ large effects to small effects was observed (73.91% to 26.09% respectively). Furthermore, fourteen out of these seventeen meta-analyses were categorised as large effects. Fisher's exact test revealed a significant relationship between the variables psychotherapy group and effect size magnitudes for evidence rated as moderate in quality.

Section C (Table 4.4.1) shows the proportion of effect size magnitudes for the two recommended psychotherapy groups when the evidence was rated as high in quality. Within the CBT group small effects comprised 73.33% of the high quality evidence in favour of the intervention. Within the other psychotherapies group medium/ large effects comprised 66.67% of high quality evidence in favour of the interventions. However, as stated previously this requires a cautious interpretation due to the small number of high quality meta-analyses (*N* = 6) available within the other psychotherapies group. Fisher's exact test revealed that there was not a significant relationship between the recommended psychotherapies and effect size magnitudes when the evidence was rated high in quality.

Table 4.4.1 Proportion of effect size magnitude by quality rating

Section A. Low Quality Evidence

			Effect size magnitudes		
			small	medium/ large	Total
Recommended Psychotherapy	CBT	Count	26	4	30
		% within Psychotherapy	86.67%	13.33%	100.0%
		% between Psychotherapy	70.27%	40.0%	63.83%
	Other psychotherapies	Count	11	6	17
		% within Recommended Psychotherapy	64.91%	35.29%	100.0%
		% between Psychotherapy	29.73%	60.0%	36.17%
Total		Count	37	10	47
		% within Recommended Psychotherapy	78.72%	21.28%	100.0%
		% between Psychotherapy	100.0%	100.0%	100.0%

Fisher's Exact, p = .14 (Non-significant)

Section B. Moderate Quality Evidence

			Effect size magnitudes		
			small	medium/ large	Total
Recommended Psychotherapy	CBT	Count	14	4	18
		% within Psychotherapy	77.78%	22.22%	100.0%
		% between Psychotherapy	70.00%	19.05%	43.90%
	Other psychotherapies	Count	6	17	23
		% within Recommended Psychotherapy	26.09%	73.91%	100.0%
		% between Psychotherapy	30.00%	80.95%	56.10%
Total		Count	37	10	47
		% within Recommended Psychotherapy	48.78%	51.22%	100.0%
		% between Psychotherapy	100.0%	100.0%	100.0%

Fisher's Exact, p = 0.002 (significant)

Section C. High Quality Evidence

			Effect size magnitudes		
			small	medium, large	Total
Recommended Psychotherapy	CBT	Count	11	4	15
		% within Psychotherapy	73.33%	26.67%	100.0%
		% between Psychotherapy	84.82%	50.0%	71.43%
	Other psychotherapies	Count	2	4	6
		% within Recommended Psychotherapy	33.33%	66.67%	100.0%
		% between Psychotherapy	15.38%	50.0%	28.57%
Total		Count	13	8	21
		% within Recommended Psychotherapy	61.90%	38.10%	100.0%
		% between Psychotherapy	100.0%	100.0%	100.0%

Fisher's Exact, p = 0.15 (Non-significant)

Summary of effect size magnitudes by quality rating

The cross-tabulation analyses indicated that there is a relationship between the quality of evidence and effect size magnitudes and recommended psychotherapy group (CBT and other psychotherapies) when controlling for quality rating outcome. Within the CBT group the effect size magnitude was observed to decrease as a function of increased evidence quality. However, the reverse pattern was observed within the other psychotherapies group, in which effect size magnitudes increased as a function of increased evidence quality. Furthermore, a statistically significant relationship was revealed between psychotherapy group and magnitude of effects when the evidence was moderate in quality.

PART TWO: Evidence on dichotomous outcomes

4.5 Raw risk association analysis

The CBT group (N = 108) was associated with a mean relative risk of M = 1.34 (SD = 2.19) and the other psychotherapies group (N = 91) was associated with the slightly lower mean relative risk of depression-related outcomes of M = 1.21 (SD = 1.52). Thus, the overall mean risk associations for CBT-based interventions and other psychotherapies were 34% and 21% increased respectively. An independent sample t-test was performed to test for differences between mean risk associations for CBT and other psychotherapies (see Appendix 8). Additionally, assumptions of homogeneity of variance were tested and satisfied via Levene's F-test, F (197) = .049, p = .825. The independent samples t-test revealed no significant difference between the mean risk associations of the recommended psychotherapies, *t*(197) = .47, p = .639. Furthermore, no significant difference was found between the two groups when the t-test was repeated controlling for depression intensity, favourable outcome values (i.e. relative risks values below 1), or quality assessment outcome (See Appendices 8, 9 and 10). Therefore no further post hoc analysis was necessary.

4.6 Overall risk associations for recommended psychotherapies (categorical data)

Table 4.6 shows the risk associations for the two recommended psychotherapy groups (CBT and other psychotherapies). Section A shows the evidence for depression treatment intensities combined. As with the meta-analytic data for continuous methods described earlier, there is an increased amount of meta-analytic evidence available for CBT (*N* = 108) than for the other psychotherapies group (*N* = 91).

Section A (Table 4.6) shows a comparable proportion of 'more' risk of negative outcome for CBT and other psychotherapies following treatment (40.7% and 35.2%

respectively)[12]. A similar pattern is observed for the proportion of risk within the two groups when high intensity depression treatments were considered separately (Section B, Table 4.6). Within both the CBT and other psychotherapies groups there was a greater proportion of risk associations indicating there to be less risk of negative depression-related outcome (62.9% and 69.6% respectively).

A different pattern was observed for the proportion of risks within the two groups when controlling for low intensity treatments (section C of table 4.6). A considerably greater proportion of increased risks of negative depression-related outcomes is observed for both the CBT and other psychotherapies groups (72.7% and 66.7% respectively).

Fisher's exact tests performed to examine the relation between psychotherapy groups and risk revealed no significant relationship between the two variables (recommended psychotherapy and relative risk) when controlling for either high or low intensity depression or for the two treatment intensities combined.

Table 4.6 Overall risk ratios for recommended psychotherapy

Section A. High and low intensity combined

			risk likelihood		
			less	More	Total
Recommended Psychotherapy	CBT	Count	64	44	108
		% within Recommended Psychotherapy	59.3%	40.7%	100.0%
	Other psychotherapies	Count	59	32	91
		% within Recommended Psychotherapy	64.8%	35.2%	100.0%
Total		Count	123	76	199
		% within Recommended Psychotherapy	61.8%	38.2%	100.0%

Fisher's Exact, p = 0.55 (Non-significant)

[12] A small number of meta-analytic evidence in both groups were categorised as neither (i.e. risk association = 1). These small numbers of outcomes were excluded from analysis.

Section B. High intensity

			risk likelihood		
			less	More	Total
Recommended Psychotherapy	CBT	Count	61	36	97
		% within Recommended Psychotherapy	62.9%	37.1%	100.0%
	Other psychotherapies	Count	55	24	79
		% within Recommended Psychotherapy	69.6%	30.4%	100.0%
Total		Count	116	60	176
		% within Recommended Psychotherapy	65.9%	34.1%	100.0%

Fisher's Exact, p = 0.42 (Non-significant)

Section C. Low intensity

			risk likelihood		
			less	More	Total
Recommended Psychotherapy	CBT	Count	3	8	11
		% within Recommended Psychotherapy	27.3%	72.7%	100.0%
	Other psychotherapies	Count	4	8	12
		% within Recommended Psychotherapy	33.3%	66.7%	100.0%
Total		Count	7	16	23
		% within Recommended Psychotherapy	30.4%	69.6%	100.0%

Fisher's Exact, p = 1.00 (Non-significant)

Summary of overall risk associations

Overall there was a greater proportion of decreased risks of depression-related negative outcomes for CBT and other psychotherapies for combined intensity depression and for high intensity depression. The proportions of decreased risks were slightly greater within the other psychotherapies group. In contrast, a greater proportion of increased risks was observed in both the CBT and other psychotherapies group for low intensity depression and the proportion of increased risk was slightly greater within the CBT than other psychotherapies group. However,

there were no statistically significant relationships between recommended psychotherapy groups and risk.

4.7 Risk associations by treatment comparator (categorical data)

Within both groups there is a significant proportion of the evidence that indicates increased risks of negative depression-related outcomes when compared with another treatment or no treatment at all. Thus it was necessary to examine treatment comparators that related to the risk associations for the psychotherapy groups.

4.7.1 Psychotherapy compared to non-active interventions

Table 4.7.1 shows the proportion of risk ratios when the recommended psychotherapies were compared with non-active interventions. Section A (Table 4.7.1) shows the proportions of risk ratios categorised as 'less' and 'more' for combined intensity depression. Within the other psychotherapies group a greater proportion of decreased risks was observed than the proportion within the CBT group (59.1% and 47.6% respectively).

Section B (Table 4.7.1) shows the proportion of risk ratios for high intensity depression. A comparable proportion of decreased risks was observed in the CBT and other psychotherapies groups (66.7% and 68.8% respectively). The reverse pattern is observed for low intensity depression (Section C of Table 4.7.1), where there is an increased risk of depression-related outcomes within both psychotherapy groups. The proportion of increased risk within the CBT group is greater than that observed within the other psychotherapies group (77.8% and 66.7% respectively).

However, the Fisher's exact tests performed revealed no significant relationships between recommended psychotherapy group and risk association when compared to non-active comparators for any treatment intensity.

Table 4.7.1 Proportion of Relative-Risk Ratios compared to non-active comparators

Section A. Combined Intensity (High and Low)

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	10	11	21
		% within Recommended Psychotherapy	47.6%	52.4%	100.0%
	Other psychotherapies	Count	13	9	22
		% within Recommended Psychotherapy	59.1%	40.9%	100.0%
Total		Count	23	20	43
		% within Recommended Psychotherapy	53.5%	46.5%	100.0%

Fisher's Exact, p = 0.55 (Non-significant)

Section B. High intensity

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	8	4	12
		% within Recommended Psychotherapy	66.7%	33.3%	100.0%
	Other psychotherapies	Count	11	5	16
		% within Recommended Psychotherapy	68.8%	31.3%	100.0%
Total		Count	19	9	28
		% within Recommended Psychotherapy	67.9%	32.1%	100.0%

Fisher's Exact, p = 1.00 (Non-significant)

Section C. Low Intensity

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	2	7	9
		% within Recommended Psychotherapy	22.2%	77.8%	100.0%
	Other psychotherapies	Count	2	4	6
		% within Recommended Psychotherapy	33.3%	66.7%	100.0%
Total		Count	4	11	15
		% within Recommended Psychotherapy	26.7%	73.3%	100.0%

Fisher's Exact, p = 1.00 (Non-significant)

4.7.2 Psychotherapy compared to active interventions

Table 4.7.2 shows the proportion of risk associations when the two groups of recommended psychotherapies were compared with an active intervention. Table 4.7.2 shows the proportion of risk associations for high intensity depression only. Within the CBT group a comparable proportion of 'more' and 'less' risk was observed (52.9% and 47.1% respectively). In contrast, within the other psychotherapies group there is a greater proportion of 'less' to 'more' risks within the other psychotherapies group (61.5% to 38.5% respectively). Only a small number of meta-analyses were performed for low intensity depression (see Appendix 11), which made these proportions of risk association difficult to interpret.

A Fisher's exact test performed to examine the relationship between psychotherapy groups and risk revealed no significant relation between the two variables when psychotherapy was compared to active interventions for any treatment intensity.

Table 4.7.2 Proportion of risk associations compared to active interventions (High intensity only)

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	8	9	17
		% within Recommended Psychotherapy	47.1%	52.9%	100.0%
	Other psychotherapies	Count	8	5	13
		% within Recommended Psychotherapy	61.5%	38.5%	100.0%
Total		Count	16	14	30
		% within Recommended Psychotherapy	53.3%	46.7%	100.0%

Fisher's Exact, p = 0.48 (Non-significant)

4.7.3 Psychotherapy compared to medication

Table 4.7.3 shows the proportion of risk associations for the two groups of recommended psychotherapies compared with medication. Within both groups

there is a greater proportion of decreased risk of negative depression-related outcomes. The proportions of decreased-risk within the CBT group and the other psychotherapies group were comparable (66.7% and 61.1% respectively). Only a small number of relative-risk meta-analyses were available within the other psychotherapies group for low intensity depression and therefore it was not possible to perform a cross-tabulation analysis.

A Fisher's exact test performed to examine the relationship between psychotherapy groups and risk revealed no significant relationship between the two variables when psychotherapies were compared to medication.

Table 4.7.3 Proportion of risk associations for psychotherapies compared to medication (high intensity only)

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	20	10	30
		% within Recommended Psychotherapy	66.7%	33.3%	100.0%
	Other psychotherapies	Count	11	7	18
		% within Recommended Psychotherapy	61.1%	38.9%	100.0%
Total		Count	31	17	48
		% within Recommended Psychotherapy	64.8%	35.4%	100.0%

Fisher's Exact, p = 0.76 (Non-significant)

4.7.4 Psychotherapies combined with medication compared to active or non-active interventions or placebo

Only high intensity meta-analyses were performed for psychotherapies and medication combined compared with active interventions/ psychotherapy or non-active comparators. There was an equal proportion of relative risk within both groups of psychotherapies. However, this was comprised of a small amount of

meta-analytic data, particularly in the other psychotherapies group, which makes this difficult to interpret.

A Fisher's exact test performed to examine the relationship between psychotherapy groups and risk revealed no significant relation between the two variables when psychotherapy combined with medication was compared to the various treatment combinations.

Table 4.7.4 Proportion of risk associations for psychotherapies and medication combined to active or non-active comparators

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	5	5	10
		% within Recommended Psychotherapy	50.0%	50.0%	100.0%
	Other psychotherapies	Count	1	1	2
		% within Recommended Psychotherapy	50.0%	50.0%	100.0%
Total		Count	6	6	12
		% within Recommended Psychotherapy	50.0%	50.0%	100.0%

Fisher's Exact, p = 1.00 (Non-significant)

4.7.5 Psychotherapy combined with medication compared to medication

Table 4.7.5 shows the proportion of risk ratios for the two groups of recommended psychotherapies combined with medication compared to medication alone. Within both the CBT and other psychotherapies groups, a greater proportion of decreased risk of depression-related outcomes was observed (74.1% and 100% respectively). The proportion of decreased risk within the other psychotherapies group was considerably larger than the proportion within the CBT group. For low intensity depression there were only two meta-analyses within the other psychotherapies

group and none performed for the CBT group, thus cross-tabulation analysis could not be performed.

A Fisher's exact test performed to examine the relationship between recommended psychotherapy groups and risk revealed no significant relationship between the two variables when recommended psychotherapies were combined with medication and compared to medication alone.

Table 4.7.5 Proportion of relative risk ratios for psychotherapies combined with medication compared to medication (High intensity only)

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	20	7	27
		% within Recommended Psychotherapy	74.1%	25.9%	100.0%
	Other psychotherapies	Count	12	0	12
		% within Recommended Psychotherapy	100.0%	0.0%	100.0%
Total		Count	32	7	39
		% within Recommended Psychotherapy	82.1%	17.9%	100.0%

Fisher's Exact, p = 0.08 (Non-significant)

4.7.6 Psychotherapy combined with medication compared to active intervention or non-active and medication or placebo

A small set of meta-analyses were performed comparing the groups of psychotherapy combined with medication to either psychotherapy or a non-active comparator combined with medication and placebo. The small amounts of meta-analyses performed were for high intensity depression (see Appendix 11). It is not possible to interpret this further due to the small numbers.

4.7.7 Psychotherapies compared to psychotherapy combined with medication or non-active comparator or placebo

A small amount of meta-analyses were performed comparing other psychotherapies with psychotherapy combined with either medication, a non-active comparator (i.e. a medication clinic) or placebo. No meta-analysis was performed for the CBT group. Therefore, cross-tabulation analysis was not performed.

Summary of risk associations by treatment comparators

A similar pattern was observed when the two recommended psychotherapy groups were compared to non-active interventions and active interventions. Within the CBT group there were marginally greater proportions of increased risks of depression-related outcomes and within the other psychotherapies group there was a greater proportion of decreased risks. However, when controlling for treatment intensity there was a greater proportion of decreased risk for both recommended psychotherapy groups within high intensity depression, which were comparable between the two groups. Within low intensity depression there was a greater proportion of increased risk for both psychotherapies, with a greater proportion in the CBT group than the other psychotherapies group.

When the psychotherapies were compared to medication both the CBT and other psychotherapies groups revealed comparable proportions of decreased risk. There were greater differences in the proportion of decreased risks within the other psychotherapies group when the two psychotherapy groups combined with medication were compared with medication alone. However, there were no statistically significant relationships between recommended psychotherapy groups and risk when controlling for specific treatment comparators.

4.8 Risk associations by quality rating

Table 4.8 shows the proportion of risk associations within the recommended psychotherapy groups by quality outcome rating. The risk associations are presented in separate sections of the table (A –C), which correspond to low, medium and high quality ratings (as per the GRADE,2004).

Section A shows a comparable proportion of decreased-risks within both the CBT and other psychotherapies groups (60.3% and 63% respectively) when the evidence was rated as 'low' quality. For 'moderate' quality evidence (Section B), a similar pattern of decreased risks within both recommended psychotherapy groups was observed, with a slightly larger proportion within the other psychotherapies group than the CBT group (65.7% and 57.1% respectively). Similarly, greater proportions of less risk were observed for 'high' quality evidence within both the CBT and the other psychotherapies group (Section C). However, this consisted of only a small number of meta-analyses within the other psychotherapies group.

Fisher's exact tests performed to examine the relationship between psychotherapy groups and risk revealed no significant relationship between two variables for any of the three quality ratings.

Table 4.8 Proportion of risk associations within recommended psychotherapies by quality rating

Section A. Low Quality Evidence

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	41	27	68
		% within Recommended Psychotherapy	60.3%	39.7%	100.0%
	Other psychotherapies	Count	34	20	54
		% within Recommended Psychotherapy	63.0%	37.0%	100.0%
Total		Count	75	47	122
		% within Recommended Psychotherapy	61.5%	38.5%	100.0%

Fisher's Exact, p = 0.85 (Non-significant)

Section B. Moderate Quality Evidence

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	15	14	29
		% within Recommended Psychotherapy	51.7%	48.3%	100.0%
	Other psychotherapies	Count	23	12	35
		% within Recommended Psychotherapy	65.7%	34.3%	100.0%
Total		Count	38	26	64
		% within Recommended Psychotherapy	59.4%	40.6%	100.0%

Fisher's Exact, p = 0.31 (Non-significant)

Section C. High Quality Evidence

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	8	3	11
		% within Recommended Psychotherapy	72.7%	27.3%	100.0%
	Other psychotherapies	Count	2	0	2
		% within Recommended Psychotherapy	100.0%	0.0%	100.0%
Total		Count	10	3	13
		% within Recommended Psychotherapy	76.9%	23.1%	100.0%

Fisher's Exact, p = 1.00 (Non-significant)

4.9 Risk association by type of risk

The type of risks measured was a factor of particular interest for dichotomous outcomes in providing context for the evidence supporting recommended psychotherapies. Within this review relative risk data were categorised into four major types: (i) depression score indicated by scores above/ below a threshold indicative of the presence of the disorder; (ii) treatment acceptability indicated by leaving study or treatment early; (iii) relapse or recurrence (indicated by clinician or self- report); and (iv) not achieving remission. The proportions of risk associations

within the two recommended psychotherapy groups for each risk type are summarised below.

4.9.1 Depression score

Table 4.9.1 shows the proportions of risk associations within CBT and other psychotherapies for the risk type depression score. Within both CBT and the other psychotherapies group there were greater proportions of decreased risk of scoring above the threshold for depression, with a greater decreased risk in the CBT group than in the other psychotherapies group (63.3% and 53.3% respectively).

A Fisher's exact test was performed to examine the relationship between psychotherapy groups and risk for depression score risk type. This revealed no significant relationship between two variables.

Table 4.9.1 Proportion of risk associations for depression scores (high-intensity only)

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	19	11	30
		% within Recommended Psychotherapy	63.3%	36.7%	100.0%
	Other psychotherapies	Count	8	7	15
		% within Recommended Psychotherapy	53.3%	46.7%	100.0%
Total		Count	27	18	45
		% within Recommended Psychotherapy	60.0%	40.0%	100.0%

Fisher's Exact, p = 0.54 (Non-significant)

4.9.2 Leaving study/ treatment early

Table 4.9.2 shows the proportion of risk associations within the recommended psychotherapy groups for the risk type leaving study/ treatment early (this risk type is thought to reflect treatment acceptability). Within section A of the table

(combined depression intensity) a greater proportion of decreased risk related to leaving the study/ treatment early is observed within the other psychotherapies group than within the CBT group (64% and 47.2% respectively).

Section B (Table 4.9.2) shows a similar pattern for high-intensity depression, where a greater proportions of 'less' risk was observed for other psychotherapies than CBT (73.7% and 56% respectively). The risk association proportions observed within the CBT group for high intensity depression indicate comparable or marginally greater 'treatment acceptability' relative to other study conditions. In contrast, the risk association proportions observed within the other psychotherapies group indicates much greater 'treatment acceptability' than other study conditions. Section C (Table 4.9.2) shows a reverse pattern for low intensity depression where greater proportions of increased risk are observed for both the CBT and other psychotherapies groups (72.7% and 66.7% respectively). This could indicate that the recommended psychotherapies (e.g. physical exercise as a psychosocial intervention) are less acceptable relative to other study conditions.

A Fisher's exact test performed to examine the relationship between psychotherapy groups and risk for 'treatment acceptability' revealed no significant relationship between the two variables.

Table 4.9.2 Proportion of risk associations for treatment acceptability (leaving study/ treatment early)

Section A. High and low-intensity combined

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	17	19	36
		% within Recommended Psychotherapy	47.2%	52.8%	100.0%
	Other psychotherapies	Count	32	18	50
		% within Recommended Psychotherapy	64.0%	36.0%	100.0%
Total		Count	49	37	86
		% within Recommended Psychotherapy	57.0%	43.0%	100.0%

Fisher's Exact, p = 0.13 (Non-significant)

Section B. High-intensity combined

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	14	11	25
		% within Recommended Psychotherapy	56.0%	44.0%	100.0%
	Other psychotherapies	Count	28	10	38
		% within Recommended Psychotherapy	73.7%	26.3%	100.0%
Total		Count	42	21	63
		% within Recommended Psychotherapy	66.7%	33.3%	100.0%

Fisher's Exact, p = 0.18 (Non-significant)

Section C. Low-intensity combined

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	3	8	11
		% within Recommended Psychotherapy	27.3%	72.7%	100.0%
	Other psychotherapies	Count	4	8	12
		% within Recommended Psychotherapy	33.3%	66.7%	100.0%
Total		Count	7	16	23
		% within Recommended Psychotherapy	30.4%	69.6%	100.0%

Fisher's Exact, p = 1.00 (Non-significant)

4.9.3 Relapse/ Recurrence

Table 4.9.3 shows the proportions of risk associations within the recommended psychotherapy groups for the risk type 'relapse/ recurrence'. Although a high proportion of 'less' risks was observed within both psychotherapy groups, this was greater within the CBT group than within other psychotherapies (82.4% to 73.9% respectively).

A Fisher's exact test performed to examine the relationship between recommended psychotherapy groups and risk for risk type 'relapse/ recurrence' revealed no significant relationship between the two variables.

Table 4.9.3 Proportion of risk associations of relapse/ recurrence (high-intensity only)

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	28	6	34
		% within Recommended Psychotherapy	82.4%	17.6%	100.0%
	Other psychotherapies	Count	17	6	23
		% within Recommended Psychotherapy	73.9%	26.1%	100.0%
Total		Count	45	12	57
		% within Recommended Psychotherapy	78.9%	21.1%	100.0%

Fisher's Exact, p = 0.52 (Non-significant)

4.9.4 Not achieving remission

Table 4.9.4 shows the proportions of risk associations within CBT and other psychotherapies for the risk type 'not achieving remission'. Within the CBT group all of the evidence available indicated 'more' risk. Although, a greater proportion of less risk was observed within the other psychotherapies group (33.3%) this was based on small amount of meta-analytic evidence (*N* = 3) and thus difficult to interpret.

A Fisher's exact test were performed to examine the relationship between psychotherapy groups and risk for the risk type 'not achieving remission', this revealed a significant relationship between the two variables. However, due to the small amount of meta-analytic evidence within the other psychotherapies group this needs to be interpreted with caution.

Table 4.9.4 Proportion of risk associations not achieving remission (High intensity only)

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	0	8	8
		% within Psychotherapy	0.0%	100.0%	100.0%
		% between Psychotherapy	0.0%	88.9%	72.7%
	Other psychotherapies	Count	2	1	3
		% within Psychotherapy	66.7%	33.3%	100.0%
		% between Psychotherapy	100.0%	11.1%	27.3%
Total		Count	2	9	11
		% within Recommended Psychotherapy	18.2%	81.8%	100.0%
		% between Psychotherapy	100.0%	100.0%	100.0%

Fisher's Exact, p = 0.05 (Significant)

Summary of risk associations by type of risk

From examination of the meta-analytic evidence controlling for risk type, a general pattern was observed of greater proportions of decreased risk within both the CBT and other psychotherapies group. In relation to depression scores and relapse/ recovery there were higher proportions of decreased risk within the CBT group compared to the proportions of decreased risk observed within the other psychotherapy groups. For risks related to 'treatment acceptability' and 'not achieving remission', higher proportions of decreased risk were observed within the other psychotherapies group, which was contrasted with higher proportions of increased risk within the CBT group.

A Fisher's exact test performed to examine the relationship between psychotherapy groups and risk ratios for various types of risk revealed no significant relationships between the two variables for specific risk-types with the exception of 'not achieving remission', which due to a small sample within the other psychotherapies group requires cautious interpretation.

5. Discussion

5.1 Chapter overview

This chapter will consider the study findings in relation to the initial research questions. It will then consider the relevance of the findings to the existing theoretical and empirical literature on psychotherapy research and guidelines. This will be followed by a consideration of the quality of this study. The clinical implications of the findings will then be considered. Finally, suggestions will be made for further research.

5.2 How strong is the evidence base for CBT compared with other psychological interventions within the NICE depression guideline?

This question was addressed through analysis of raw effect sizes and risk associations for CBT and other psychotherapies recommended by NICE. On aggregate, the evidence for CBT generally does not appear to have as much strength when compared against the collective evidence from other psychotherapies based on the analyses of effect sizes within the current project. However, there was some specific circumstances where CBT was stronger than other psychotherapies as summarised below.

5.2.1 Raw effect sizes (continuous outcomes)

The differences between the mean raw effect sizes for the two recommended psychotherapy groups, for all of the meta-analyses performed within the guideline evidence review, indicated stronger evidence within the other psychotherapies group than within the CBT group. The difference between the two groups' means was maintained when favourable raw effect size values (i.e. 0.2 or above) were

analysed separately (see Appendix 1). Although, there were no significant differences between the psychotherapy groups when high and low intensity groups were considered separately, significantly larger mean effects within the other psychotherapy group were revealed when a sub-analysis was performed on favourable raw effect size values for both high and low intensity depression (See Appendix 2).

Further post hoc analysis of treatment comparator sub-groups found that, when the psychotherapy groups were compared to medication (either individually or in combination with medication), there were significantly larger mean effect sizes for CBT-based interventions for high intensity depression. There were no other significant differences between the two groups of psychotherapy for any other treatment comparator. However, the mean effect sizes within the other psychotherapies group were larger than the CBT group for both non-active and active interventions (including psychotherapy) treatment comparators within the high intensity data set with the p-value for the latter comparator approaching significance.

Further sub-analysis of mean-ranked effect size by treatment comparator revealed these to be significantly different within the other psychotherapies group but not within the CBT group. More specifically, within the other psychotherapies group, there were larger mean effect sizes for non-active comparators than medication comparators, and larger mean differences for psychotherapy comparators than medication. These differences in effect sizes for treatment comparator could explain the larger overall mean effects observed within the other psychotherapies group. Furthermore, it is possible that less variation in effect sizes observed in the CBT group, as indicated by relatively smaller standard deviations of effect size, could explain the smaller mean effects observed.

5.2.2 Relative risk associations (dichotomous outcomes)

The mean relative risks within high intensity depression for the CBT and other psychotherapies group indicated that both groups were associated with substantially decreased risks of negative depression-related outcomes compared to comparator conditions. The overall mean relative risks for CBT-based interventions and for other psychotherapies (34% and 21% respectively) were not significantly different from one another[1]. However, it is arguably more meaningful to consider and interpret dichotomous outcome data at categorical level as opposed to its raw values.

5.3 Are there identifiable patterns of difference in the strength of evidence base for CBT and other psychotherapies?

This question was explored further through data analysis at a categorical level (i.e. applying raw effect size Cohen's conventional descriptors). Differences were observed in the proportion of effectiveness (i.e. favourable and unfavourable effects) and effect size magnitudes (i.e. small, medium, large) between the two groups of psychotherapy in respect to the overall sample and within sub-sets of the data, which are summarised below.

5.3.1 Overall effectiveness

Overall, there were equal proportions of meta-analytic data that indicated favourable effects and non-favourable effects within both CBT-based interventions and other psychotherapies groups. When effectiveness was analysed separately for depression intensity there was a slightly increased proportion of non-favourable

[1] As discussed in the methods relative risks within the psychotherapy context usually relate to attrition, relapse, recurrence etc. as opposed to harmful side effects of treatments within the medical context.

effects for high intensity depression within the two psychotherapy groups. In contrast, for low intensity depression the proportion of evidence favouring both the CBT and other psychotherapies (i.e. physical activity) groups was more pronounced, suggesting a stronger evidence base for recommended psychological and psychosocial interventions for low intensity than high intensity depression. However, the proportions of favourable effects for both psychotherapy groups were comparable for low intensity. Thus, overall the data do not indicate a pattern of difference between the two broad classes of psychotherapy.

5.3.2 Overall effect size magnitudes

Overall, contrasting patterns were found within the effect size magnitudes for CBT-based interventions and the other psychotherapies group. This consisted of increased proportions of medium and large effects within the other psychotherapies group and increased proportions of small effects within the CBT group. Moreover, there was a significant relationship between psychotherapy group (CBT and other psychotherapies) and effect size magnitude for high intensity depression, which appears to have been due to the increased proportion of small effects within the CBT group, which also made a considerable contribution to the total proportion of small effects observed for both psychotherapy groups.

5.3.3 Effect size magnitudes controlling for treatment comparators

When controlling for treatment comparators within this review, there were very few statistically significant relationships between recommended psychotherapy group and the proportions of favourable effects or effect-size magnitudes. The relationships between psychotherapy group and effect size magnitude observed in the overall sample diminished when specific classes of treatment comparators were

held constant. This suggests that specific variables could account for the differences observed within the overall effect size magnitudes between the two psychotherapy groups.

The only significant relationship between psychotherapy group and effectiveness was observed when controlling for medication in the high intensity data set. However, this was based on a restricted sample of evidence comparing the proportion of 'none' and 'small' effects (see section 4.3.2.5). A similar pattern of favourable effects for CBT was observed in the crosstab analysis that included all of the favourable effects (i.e. small, medium and large; see Section B, Table 4.3.2.5) for medication comparators. This relationship indicated that when compared to medication, CBT-based interventions had stronger evidence than the other psychotherapies group. At a categorical level this appears to be consistent with the findings from the raw post-hoc analysis of within group differences that found smaller ranked mean effect sizes for other psychotherapies compared to medication than when compared to non-active and psychotherapy treatment comparators.

As described in the previous chapter (see section 4.3.2), there was a consistent pattern of increased proportions of medium and large effect sizes within the other psychotherapies group and an increased proportion of small effect sizes within the CBT group when treatment comparators were analysed separately. However, these relationships were not significant, which may have been due to the small sample sizes within these sub-analyses (with the exception of non-active comparators), which makes it difficult to detect a statistically significant relationship.

5.3.4 Effect sizes controlling for quality

There were statistically significant relationships between recommended psychotherapy (i.e. CBT and other psychotherapies) and effect for evidence rated as

moderate in quality. This consisted of a clear pattern within the other psychotherapies group of increased proportions of favourable effects relative to the proportion observed within the CBT group. There was also a clear pattern of increased proportions of medium and large effects than small effect sizes within the other psychotherapies for moderate quality evidence. In contrast, a reverse pattern was observed within the CBT group both in respect to increased proportions of non-favourable effects and smaller effect size magnitudes for moderate quality evidence.

Although a similar pattern was observed for high quality evidence, the relationship between recommended psychotherapy group and effect was not significant. This could have been due to the small sample size for high quality evidence.

5.3.5 Relative Risk Associations

Overall, there were increased proportions of meta-analytic data that indicated decreased risk of depression related outcomes within both CBT-based interventions and other psychotherapies. Within the data analyses at a categorical level, there was a general pattern of the majority of risk associations favouring the recommended psychotherapies (i.e. indicating a reduced risk of depression related outcome). However, there were no significant relationships between recommended psychotherapy group and risk for the overall evidence. There was a general pattern of greater proportions of decreased risk for high intensity depression and greater proportions of increased risk for low intensity depression. However, the proportions of risk were comparable across psychotherapy group and this was maintained when the risk associations for different treatment comparators were analysed separately. Generally there were no statistically significant relationships when controlling for risk type except for not achieving remission, which was based on a small amount of data and thus requires a cautious interpretation. There was no significant

relationship between psychotherapy and relative risk for either of the three quality groups.

5.4 Relevance of findings to the existing theoretical and empirical literature

The process of exploring guideline evidence highlighted a number of issues in relation to the evidence base that bear both a direct and indirect relevance to the current project's findings. These issues include the guideline's approach of data synthesis, the selection of outcomes from primary studies, and interpretation of evidence as a basis for recommending psychotherapies. The findings will be considered in relation to these issues and in reference to the NICE depression guideline's summary of the evidence, and the interpretations and recommendations reported within the guideline's psychological interventions chapter (NCCMH, 2010). The findings will also be considered in relation to relevant aspects of the wider psychotherapy research literature.

5.4.1 How should psychotherapy effect sizes be interpreted within the guideline evidence reviews?

The two findings, 1) the aggregate raw effect-sizes were larger within the 'other psychotherapies' group than within the CBT group; 2) overall there were larger effect size magnitudes within the other psychotherapies group, can be interpreted in different ways depending on how effect size magnitudes are conceptualised.

The NICE guidelines place emphasis on the magnitude of effect through taking the view that effect sizes of 0.5 and above and relative risks of 0.81 are important outcomes (Pilling, 2009). On this basis one interpretation of this study's findings is that the 'other psychotherapies' group collectively generates stronger effect sizes than those of a single class of psychotherapy, namely CBT-based interventions. The

larger mean effect size within the other psychotherapies group could be due to a wider range of treatments that are more representative of the general effect of psychotherapy (Wampold, 2001).

However, the wider psychotherapy research literature indicates that there are a number of factors associated with effect size magnitude that guideline reviews do not seem to take into account or make adjustments for within their own meta-analytic outcomes. Moreover, it has been suggested that smaller effect sizes are associated with more robust evidence (Cuijpers, van Straten, Bohlmeijer, Hollon and Andersson, 2010). The fact that the majority of evidence generated within the guideline review was rated as low quality (an issue that I consider in more detail later in this chapter), and that lack of statistical power to detect small effects (i.e. $d = 0.2$) was common, presents a problem where effect size magnitudes are taken as important outcomes.

As the 'other psychotherapies' group consists of diverse psychotherapies it is likely to contain greater variance of client, therapist and study characteristics that could result in greater variance in effect sizes amongst this group. Such variation from diverse theoretical orientations could translate into better outcomes, particularly when one considers the potential for pluralistic approaches to cater for a range of client and therapist factors known to influence outcome. This is important considering that it is widely reported within the psychotherapy research literature that 25% of the variance between treatment interventions is due to study characteristics (Wampold, 2001). Furthermore, this is consistent with findings from Cuijpers et al. (2008) that demonstrated associations between effect size magnitudes and various depression study characteristics including intervention, treatments aimed at specific populations, type of recruitment and type of control.

The relationship between psychotherapy group and effect size magnitude is interesting. It is important to consider differences in the volume of meta-analytic

data available for the two groups. Within the CBT group the number of meta-analyses indicating a favourable effect (i.e. small, medium, large) was almost double that of the other psychotherapies group for high intensity depression. Statistical theory (i.e. random and fixed effect models) would suggest that more accurate effect size estimations would occur with an increase in both the participant and study sample size (see Hedges and Vevea, 1998). Therefore, despite the limitations of the guidelines approach to data synthesis it is possible that the smaller effects on aggregate across the larger sample of CBT meta-analyses performed provide a more accurate reflection of this psychotherapy's effect.

5.4.2 The construction of relative risk outcomes within guideline evidence

The relative-risk reductions on specific risk types observed within the correlational analysis are worth further consideration. The dichotomous outcomes relapse, recurrence and not achieving remission were frequently based upon cut-off scores on depression scales, and thus were proxy measures of continuous outcomes. This could explain the more pronounced reductions in risk for CBT based on relapse and recovery within the NICE guideline evidence. Twice as many meta-analyses within the CBT group were based on these proxy outcomes compared to the other psychotherapies group, where a modest proportion of increased risk on these outcomes was observed. It could be argued that such dichotomous outcomes are a duplicate of the continuous measure; particularly as it is unclear on a psychometric level how rigorously these proxy measures have been tested for construct validity.

5.4.3 Equivalence and superiority as conceptual approaches within guideline evidence reviews

Within recent years clearer distinctions have been drawn between studies designed to address psychotherapy equivalence (i.e. non-inferiority study designs) and superiority (i.e. relative efficacy trials; see Greene, Morland, Durkalski and Freuh, 2008; Piaggio et al.,2006). On review of the primary study characteristics that form the basis of the guideline's psychotherapy recommendations (NCCMH, 2010), the vast majority of primary comparative outcome studies reviewed followed a superiority design. The guideline's evidence review makes interpretations about the equivalence of psychological interventions. This is contrary to how the guideline interpreted previous meta-analytic findings that found no large differences in effects between different classes of psychotherapy (e.g. Cuijpers et al, 2008), where it draws a distinction between trials designed to test differences in efficacy and trials designed to establish equivalence, stating that "Cuijpers and colleagues (2008a) had failed to find such differences rather than establishing that no differences existed" (NCCMH, 2010, p. 161). However, NICE's meta-analytic strategy makes no clear attempt to establish equivalence between psychotherapies in head-to-head comparisons of the data. Instead, they adopt a similar approach to other meta-analyses that assess differences in efficacy and generally find no important differences between psychotherapies' effectiveness.

On review of the NICE guideline evidence, an equivalence (or non-inferiority) approach is conceptually appropriate for making recommendations on low intensity interventions due to the availability of evidence for individual and group-based approaches for CBT and physical activity. Such a conceptual approach is also understandable *within* specific classes of high intensity treatments where the evidence is available for individual and group formats, such as mindfulness-based cognitive therapy. However, considering the differences in underpinning theories and mechanisms of change and the lack of parity in the use of outcome assessments

consistent with underpinning theories (discussed l), neither a non-inferiority nor superiority conceptual approach appears appropriate to reviewing psychological interventions within guidelines.

5.4.4 CBT-based interventions and bona fide psychotherapies

The findings of this review provide general support for prior meta-analyses that have generally found no difference between CBT and bona fide psychotherapies (Wampold, 1997; Wampold et al., 2002). Most of the treatments within the other psychotherapies group would meet the definition of bona fide psychotherapies[2] . On review of the study characteristics used within NICE's evidence review, it was apparent that CBT-based interventions were directly compared to all of the treatments within the other psychotherapies group. However, there was a lack of direct comparisons between treatments within the other psychotherapies group with the exception of comparisons between short-term psychodynamic psychotherapy and behavioural activation, and couples therapy and IPT[3]. As CBT was the only intervention to be compared with all of the other recommended psychotherapies, this aspect of the review arguably says more about the relative strength of evidence for CBT compared to other psychotherapies than psychotherapies more generally. There were no clinically important differences between CBT and other psychotherapies irrespective of whether or not they were designed for treating depression.

[2] Bona fide psychotherapies are defined as those that are intended to be therapeutic as opposed to those that are merely designed to act as a 'non-specific' control.
[3] Counselling was also compared to another bona fide psychotherapy, Emotion Focussed Therapy, which was not recommended within the guideline.

5.4.5 Inconsistencies within the evidence base

This review highlights the inconsistency of the evidence relating to the effects of recommended psychotherapies, which is an inherent part of the guideline's evidence base. Such inconsistencies are likely fuelled by the vast amount of meta-analytic data generated through the guideline development process. This reflects the guideline's attempt to address a range of clinical questions relevant to its broad scope. However, in doing so it leaves the evidence exposed to further bias and confusion.

Jauhar, McKenna and Laws (2016) highlighted similar issues in relation to the more recent update of NICE guidance on psychological treatments for bipolar disorder. They argued that the large amount of meta-analytic data based upon single studies was inappropriate and defeated the objective of data synthesis. It was also noted that NICE meta-analyses often consisted of related outcomes (e.g. outcomes based on multiple measures such as self and clinician scales from the same set of participants). Furthermore, attempts to conduct multiple meta-analyses on related outcomes have drawn criticism for not dealing with statistical problems (i.e. false positives) that emerge from multiple comparisons (Jauhar et al., 2016).

In addition to the multitude of meta-analyses performed, further sources of inconsistencies cause confusion within guideline meta-analytic data including the different end points (end of study, 6 month follow-up, etc.) and, as mentioned before, the use of continuous outcome cut-off scores as a proxy for dichotomous outcomes. This creates a considerable challenge to interpretation of the evidence due to multiple indices and inconsistent outcomes.

One of the challenges for a guideline development programme is balancing the review of evidence on outcomes that reflect the broad scope of stakeholders in a coherent way. Kendall et al. (2016) assert that this challenge is considerable as it involves comparisons of multiple interventions with one another on multiple

outcomes and multiple points in time. However, it is debatable whether such a multiplicity of outcomes is useful, particularly where this involves the construction of outcomes from other outcomes as discussed earlier in relation to continuous measures converted into dichotomous outcomes for remission and relapse.

The guideline development programme's broad scope essentially leads to the evidence base being extracted out of its original context to answer questions that the primary studies were not designed to address. This is similar to what Moreira (2007) describes as the parallel processes of 'disentanglement', where data is extracted from its original situation, and 'qualification' or providing the data with further qualities such as precision and a perception that they are unbiased. Moreira argues that these processes are fundamental to evidence based guidance establishing 'persuasive power'. However, as illustrated in this review, such an approach could seriously compromise the credibility of the guidance.

5.4.6 Biased interpretations of the evidence base for CBT

This study found comparable strength of evidence for CBT and other psychotherapies. This is broadly consistent with the guideline's review of CBT and individual treatments within the other psychotherapies (i.e. IPT, counselling etc.). However, as has been argued previously (e.g. Holmes, 2002; Winter, 2010), the guideline's evidence review chapter frames its interpretations of the evidence in a way that seems to subtly shape it in favour of CBT. For example, within the summary of CBT's relative efficacy to other psychological therapies (NCCMH, 2010), different conclusions are drawn about its possible equivalence with IPT than short-term dynamic psychotherapy (both recommended as treatment options, although the latter with some limits and further qualification). On reporting evidence for no clinically important differences in decreasing depression between comparisons of CBT and IPT, NICE states that "evidence although limited suggests that IPT might be

as effective as CBT in the treatment of depression" (p.235), whereas when short-term psychodynamic therapy is considered as a comparison, there is no mention of the possibility of equivalence within the statement "it is not possible to draw any clear conclusions about the relative efficacy of the treatments" (p.235).

The findings of this review seem to confirm that certain types of evidence hold considerably more weight than others, namely the relative efficacy of psychotherapy compared to medication. This strongly favours CBT-based interventions for high intensity depression over psychotherapies aggregated within the 'other psychotherapies' group. The absolute efficacy of anti-depressants compared to placebo interventions is estimated to be 0.31 (Kirsch et al., 2008; Turner, 2004). As antidepressant therapy is widely used in the treatment of depression the efficacy of psychotherapies of course needs to be considered relative to that of antidepressant therapy in treatment guidelines. Furthermore, it could be argued that this approach is consistent with the aims of the guideline development programme, namely weighing up evidence with 'clinical judgement' and taking broader factors of the service context in which the guidance is to be applied into account (Pilling, 2008).

5.4.7 Quality of guideline evidence

The relationship between psychotherapy effects and the quality rating assigned to them makes the guideline evidence for recommended psychotherapies difficult to interpret. Fifty seven percent of the meta-analytic data (effect size and relative risk) was assigned a low quality outcome. The combined impact of a lack of methodological rigour at primary study level and secondary data synthesis level as assessed by Oxman & GRADE (2004) is not well understood. Wampold (2001) suggested that quality has a relatively minimal impact on the general effect of psychotherapy, whereas Cuijpers et al. (2010) suggested that poor quality studies have led to an overestimation of treatment effects in studies of adult depression.

These contrasting findings could be due to different approaches to assessing quality and statistical techniques employed (e.g. statistical corrections or regression analyses.

Within the current study when moderate quality evidence was analysed separately there was a statistically significant association between psychotherapy group and both effectiveness and effect size magnitude. More specifically, there was a pattern of increased proportions of favourable effects and larger effect size magnitudes were also observed within the other psychotherapies group. In contrast, there were increased proportions of non-favourable effects and small effect size magnitudes within the CBT group. Furthermore, these patterns were consistent with the significantly larger mean effect sizes found for other psychotherapies group compared to CBT for moderate and high quality evidence.

The pattern within the other psychotherapies group was not consistent with the findings of Cuijpers et al. (2010) that smaller effect sizes (d = 0.22) for high quality studies differed significantly from the mean effect sizes found within lower quality studies. However, the methods used by Cuijpers and colleagues to assess quality were more robust and directly relevant to psychological interventions than the GRADE system adopted within NICE. This consisted of combining an assessment of treatment delivery as set out by empirically supported therapies review (Chambless & Hollon, 1998) with the criteria proposed by the Cochrane Collaboration for the methodological validity of studies (Higgins & Green, 2008). Thus, there may be specific quality factors that CBT and other psychotherapies' evidence differed on that are obscured within a generic system (i.e. GRADE) of rating the overall quality of evidence. Such factors could account for the differences in effect size magnitudes that were observed for moderate and high quality evidence between CBT and other psychotherapies.

Within the depression guideline, low quality evidence is frequently cited as clinically important outcomes for CBT and other psychotherapies. Although the depression guidance acknowledges the limitations of the quality of its clinically important outcomes and presents the quality assessment outcomes alongside its summary evidence, this arguably only adds to the confusion within the evidence. This is further complicated by the separate use of statistical significance from factors that contribute to quality outcome, thus resulting in a scenario where an effect size that favours a treatment over a comparator condition is reported as statistically significant (i.e. confidence interval does not include zero) but also rated as low quality evidence. Furthermore, CBT-based interventions appear to be afforded more leeway than other psychological treatments for positive interpretations of low quality evidence.

In response to similar criticisms of NICE bipolar disorder guidelines, Kendall et al. (2016) argued that limiting recommendations of psychotherapies (and pharmacological therapies) to high quality evidence, which is scarce, would result in a restriction of recommendations of psychotherapy. Kendall and colleagues refer to the process of clinical decision-making within the guideline recommendations and the views of stakeholders about the benefits of psychotherapies that are not captured by meta-analysis of RCT evidence. Although this argument has its merits, it runs into difficulty when one considers the emphasis that NICE places on guidelines recommendations with a robust evidence base.

5.4.8 Psychotherapy evidence reviewed within a medical context

This review's findings highlight the extent of the guideline's reliance upon symptom scales as primary outcome measures in the evaluation of effectiveness of psychological interventions. The Hamilton Depression Rating Scale (HDRS; Hamilton, 1960) and the Beck Depression Inventory (BDI; Beck et al., 1961) are the most

frequently used measures within the guideline. The Hamilton Depression Rating Scale (1960) was primarily developed to assess depression severity and change during antidepressant therapy, and thus focuses on psychiatric symptoms. It has been criticised as inadequate for capturing the full spectrum of depression as a construct (Bagby et al., 2004) and for running the risk of placing 'all depressions in the same basket' (Demyttenaere & De Fruyt, 2003). Although symptoms are an indicator of pathological processes, i.e. depression at a diagnostic level, they are not necessarily linked to the underlying problem, e.g. problems in an individual's wider social environment (Roth and Fonagy, 2005).

The Beck Depression Inventory (BDI; Beck, 1961) is the most widely used self-report measure, and although it provides an assessment of a psychological representation of depression it represents only one psychological domain, and is highly reactive to cognitive interventions, which is likely to be indicating greater degrees of success than would be found using alternative assessments. From reviewing the study characteristics (NCCMH, 2010, appendix 17b) it was of note that, for the other psychotherapies group, the use of psychologically oriented measures was far less common compared to medically oriented outcome measures of symptomatology such as the Hamilton Depression Rating Scale (HDRS) and the Montgomery-Asberg Depression Scale (MDRS; Montgomery and Asberg, 1979). It would also be expected that CBT-based interventions would do better on such measures through the interventions' direct targeting of depression symptoms (e.g. negative thought patterns), relative to the diverse treatments within the 'other psychotherapies' group.

Since the other psychotherapies group was assessed by outcome measures that were not theoretically coherent with the therapies concerned, they were more disadvantaged in demonstrating their effectiveness in relation to symptom and diagnostic-level outcomes. This is also relevant where depression scores are used as

proxy measures of risk outcomes. Unsurprisingly the BDI is frequently used within CBT trials. The widespread use of the BDI as a psychological assessment in the construction of guideline evidence also runs the risk of placing all psychotherapies into the same basket! This presents a general challenge to evaluation of psychotherapies within a medical context.

The guideline excluded other outcome measures relevant to important areas of interpersonal, social and occupational outcomes that were available within primary studies for review (see NCCMH, 2010, appendix 17b). It has been previously highlighted that such outcome measures should be given more prominence within guideline reviews (Pilling, 2009). However, the medical context in which psychotherapy evidence is considered has resulted in no uptake of broader outcomes within the guideline development programme. This unfortunately perpetuates a predominant EBM culture within psychotherapy research, focussed on treating symptoms within the narrow construct of psychiatric diagnosed disorder. When one considers the complex combination of social, psychological and biological factors that cause depression, its broad impact on individuals' lives and its relationship to the wider societal context (Cuijpers, Beekman and Reynolds, 2012), such a narrow approach seems regressive.

Psychotherapies differ in terms of their underpinning theories of specific disorders and psychotherapeutic mechanisms of change (including factors common to all therapies). This illustrates a fundamental challenge when evidence is reviewed and psychotherapy recommendations formed within a purely medical context. Attempts to scrutinise their efficacy on the basis of a common medical or psychological outcome measure create an uneven playing field. A social constructionist perspective would suggest that different psychotherapies address multiple realties or meanings of psychological distress for different individuals, families, communities, etc. Thus, reviews of psychotherapy evidence within a medical context that

discounts broader based theories within its evaluation of outcome would be viewed as insufficient.

5.5 Study limitations

5.5.1 Categorisation of psychotherapies

For the purposes of this review two psychotherapy 'super-classes' (CBT-based interventions and other psychotherapies) were essentially constructed and data from separate classes of psychotherapy were pooled into these broader categories. This grouping of meta-analytic data for different psychotherapies contains considerable heterogeneity. For instance, the CBT-based interventions group included evidence on variants such as mindfulness-based cognitive therapy, group-based interventions and CBT for specific patient populations. The other psychotherapies group consisted of distinct and diverse classes of psychotherapies. Furthermore, some of the treatments included in the other psychotherapies group, namely behavioural activation and couples therapy (based on behavioural principles), were identified as sharing a common theoretical base with CBT (NCCMH, 2010).

It could be argued that such an approach of categorisation results in loss of meaning in the comparison of the evidence. However, this approach was based on my primary research question as to 'whether or not differences exist in the strength of NICE's evidence-base for CBT and other psychotherapies?' This differs from a question concerned with the relative efficacy of specific psychotherapies, where such grouping of evidence would be more problematic. Furthermore, such an approach has precedence in systematic and meta-analytic reviews in psychotherapy (e.g. Butler, Chapman and Forman, 2006; Gaffan, Tsaousis & Kemp-Wheeler, 1995; Robinson, Berman and Neimeyer, 1990; Smith and Glass, 1980; Tolin, 2010;

Wampold et al., 2002). Therefore, this categorisation of recommended psychotherapies was considered an appropriate pragmatic approach in reference to the research question.

5.5.2 Pooling of different types of relative risk outcomes

It was necessary to pool risk types together into larger groups for the purposes of sub-group correlational analyses. This meant that some types of risk outcomes, namely, 'participants leaving treatment early' and 'participants who leave study early', were not differentiated within this review. It is likely that there is a real distinction between these two sub-types of patient attrition, which became undifferentiated through pooling. These could have been distinguished further by the reason for attrition, such as 'due to side effects' or 'for any reason'. It is possible that the evidence for risk type leaving treatment early is more representative of 'treatment acceptability' than the risk type leaving study early. However, within this review the decision was taken to pool together these two types of participant attrition due to how the data are utilised within NICE guidelines and more generally within the evaluation of psychotherapy (Roth and Fonagy, 2005) as proxy measures for treatment acceptability. This is an inherent problem in making interpretations of treatment acceptability, particularly as data is likely to be comprised of diverse study variables that make it difficult to attribute participant attrition purely to treatment.

5.5.3 Coding of effect sizes as 'none'

Analysis of the evidence at a categorical level required coding numerical effect sizes descriptively, which presented a challenge for interpretation. The code 'none' is ambiguous as it obscured evidence that either favoured the intervention but was below the minimal benchmark of Cohen's conventional descriptors (i.e. below 0.2);

and effect sizes that indicated a negative effect relative to its comparator. Assigning a general 'none' code was helpful in distinguishing these effect size values from those indicating positive effects (i.e. coded as per Cohen's conventional descriptions) and enabled an overview of the strengths of guideline evidence for CBT and other psychotherapies. However, this code potentially concealed important features about the weaker meta-analytic evidence for the two groups of psychotherapy and this potentially limits our understanding of the evidence presented.

5.5.4 Coding of treatment comparators

Treatment comparators were coded and further analysed as superordinate classes of comparator (e.g. non-active/ placebo interventions, active/ psychotherapy comparator, etc.) as opposed to specific comparators (e.g. treatment as usual or placebo, etc.). Within the non-active treatment comparator group the inclusion of placebo conditions, GP care and treatment as usual illustrates the heterogeneity that typifies study synthesis. Although this broad approach to coding enabled further analyses of the recommended psychotherapies' strengths in relation to different treatment comparators, these broad codes also require cautious interpretations of the data analysis.

5.5.5 Statistical significance of evidence

This project did not include a sub-group analysis to control for statistical significance in tests of differences of raw effect sizes and risk associations within the recommended psychotherapies group or the association analysis of data at a categorical level. This was a pragmatic decision taken to enable a succinct overview of the data. Controlling for the statistical significance of effect sizes and relative associations would not necessarily have provided more meaningful analyses. On

review of the evidence profiles, it was apparent that the vast majority of confidence intervals accompanying effect sizes and relative risks were non-significant irrespective of whether or not they favoured the psychological intervention. This was unsurprising given the very small number of studies included within many of the NICE guideline meta-analyses. However, not examining significance within the evidence meant that it was not possible to determine the extent that effect size and risk association values were due to random fluctuations. As non-significant findings are not in themselves evidence of no effect (Kendall, 2016), more attention was paid to the clinical significance of evidence than to its statistical significance. Moreover, statistical significance was a component of the GRADE quality ratings.

5.5.6 Specific quality of evidence

Although there were some associations reported between study quality and effectiveness/ effect size magnitude, these findings did not examine specific components of the GRADE system such as those components related to primary studies or to the guideline's own meta-analytic evidence. Therefore, it was not possible to determine the impact of specific quality components on guideline evidence for CBT or the other psychotherapies group.

As mentioned earlier, a large set of effect-size and relative risk data was generated through the guideline review process. GRADE quality outcomes use an aggregated system, and it would not have been possible within this project's scope to isolate specific quality factors that contributed to overall reductions in GRADE quality outcome ratings for each meta-analysis. Considering the secondary nature of guideline evidence, it would not have been practical to use a more robust method such as that proposed by Cuijpers and colleagues (2008) that consisted of combining the criteria of empirically supported therapies review (Chambless & Hollon, 1998) with the criteria proposed by the Cochrane Collaboration for the methodological

validity of studies (Higgins & Green, 2008). However, further research examining the relationship between specific quality factors and guideline evidence would improve understanding of this area.

5.5.7 Guideline evidence not reviewed in this study

In order to contain the scope of this project, evidence was not reviewed on a range of areas. This included evidence for health-economic modelling, subthreshold depression and service delivery.

For sub-threshold depression (i.e. dysthymia), all classes of psychotherapy where evidence was available were reviewed collectively. Although this was due to the limited amount of evidence available, interestingly it also arguably mirrors a common factors approach (Wampold, 2001) and the approach to aggregation of treatments within the other psychotherapies group adopted in the current study. Thus, it was decided that further review of this evidence was not relevant to the research question.

Guidance on service delivery is an important area that has considerable influence on mental health service provisions. Potentially, service delivery and psychotherapy provisions are not mutually exclusive when one considers the depression pathways of stepped care models, IAPT services and secondary care services. Psychotherapies are delivered within a healthcare context and it would have been interesting to explore whether the strength of the two sets of evidence coheres. The decision not to review service delivery evidence within this project was due to the breadth of this area and the need to contain the scope of this project to the area of psychotherapy.

Health economic evidence was integral to the recommendations for low intensity, and economic modelling was performed for high intensity CBT. However, this was only considered if there was adequate data from primary studies, which was not the

case for treatments that comprised the other psychotherapies. As a result the majority of guidance on specific treatments and the general guidance on psychotherapy delivery were based upon the treatment manuals used within primary studies. Therefore, separate analysis of the health economic evidence would not have informed this study any further.

There were other potentially interesting evidence characteristics that were not included in the sub-analysis of effect sizes and risk associations for the two psychotherapy groups. Treatment follow-up was one potential variable of interest that was pooled with treatment endpoint data and was not controlled for within the data analysis. Prior reviews of psychotherapy focus primarily on treatment endpoint and have typically excluded follow-up data. However, it is acknowledged that not controlling for this variable is a study limitation.

5.6 Clinical implications of research findings

5.6.1 CBT's status as the frontline intervention for depression

The findings of this review confirm that in relation to the overall strength and quality of evidence, the evidence presented within the guideline for CBT-based interventions and other psychotherapies is not very robust. This finding questions the NICE guidance for CBT's status as a frontline psychological intervention for depression based upon the evidence available (NCCMH, 2010). From the evidence presented within this project, it is argued that despite the large body of evidence available for CBT-based interventions, a broader set of psychological therapies have as strong, if not a stronger, collective evidence base. The exception to this is the stronger evidence available for CBT compared with medication for high intensity depression. As antidepressant medication is the treatment standard for depression within a medical healthcare system, CBT's status as frontline psychotherapy appears to be based primarily on its performance relative to pharmacotherapy. However,

when CBT-based interventions' large body of research evidence is considered next to aggregated evidence of other psychotherapies it is argued that more qualifiers and restrictions should be placed upon CBT recommendations.

Within more recent NICE guidelines for borderline personality disorder (NCCMH, 2009) the limited evidence base for any particular psychotherapy resulted in no recommendations for a specific class of psychotherapy. Instead, the guidance focused on what therapists should take into account when considering psychological therapy for an individual. Amongst these considerations were the choice and preference of the service user; the degree of impairment and severity of the disorder; the person's willingness to engage with therapy and their motivation to change; the person's ability to remain within the boundaries of a therapeutic relationship; and the availability of personal and professional support (NCCMH, 2009; recommendations 1.3.4.1). Such recommendations seem more consistent with the other core skills of clinical psychologists and psychotherapy practitioners within the NHS in terms of assessment, formulation and intervention irrespective of the diagnosed disorder.

Furthermore, broader recommendations of this kind accommodate multiple constructs of psychological distress and enable a broad-based approach to both psychotherapy practice and evidencing psychotherapy outcome (discussed below). The current approach to recommendations within the NICE depression guidelines restricts patient choice to a range of therapeutic methods that in my view do not reflect the breadth of theoretical integration necessary within psychotherapeutic practice.

5.6.2 Other psychotherapies' empirical journeys

Other psychotherapies are yet to journey as far along the 'empirical road' (NCCMH, 2010, pp. 160) as set out by the EBM movement that predominates in guideline development programmes. It has been highlighted within this review and within the wider literature (Pilling, 2007) that a narrow range of diagnostic and symptom-focussed outcomes are utilised to construct guideline evidence. Despite this, the collective strength of evidence for other psychotherapies suggests that there is the potential for a broader base of psychotherapies to progress further along this particular empirical road. This is irrespective of whether or not this is the ideal road for psychotherapies to take.

An implication of guidelines prematurely declaring 'winners' amongst psychological therapies is that they could inadvertently become an obstacle to other psychotherapies which currently have more modest amounts of empirical support from progressing any further. To illustrate this, the research recommendations for high-intensity psychological interventions (NCCMH, 2010; 8.12) set a research agenda and requirements for further evidence in order for other psychotherapies to be recommended. This involves comparisons against CBT on specific outcomes (e.g. relapse prevention, cost effectiveness, etc.) through the use of non-inferiority designs. One possible implication of this, which seems apparent from the current study's findings, is that higher effect sizes may be necessary for other psychotherapies to be recommended within the guidelines than cognitive behavioural therapies.

5.6.3 The reach of guideline recommendations

NICE's approach to constructing evidence for psychotherapy recommendations has broader implications due to 'the reach' that psychotherapy recommendations for depression have in respect to other guidelines. These are often referenced in order

to address issues of co-morbid depression with other mental disorders. A recent example of this can be found in the Bipolar Guideline (CG185; NICE, 2014), which recommends CBT, IPT and behavioural couples therapy for bipolar depression based on the evidence reviewed within the depression guideline (Jauhar et al., 2016). Thus, questions about the strength of evidence for the NICE depression guideline highlighted in this review are also relevant to the insertion of this psychotherapy guidance into subsequent guidelines. Furthermore, the findings of this review have implications for the future updates of the guideline, the next of which is currently in preparation and planned for publication in November 2017.

These findings also have implications for the wider service context where NICE recommendations are perceived as the blueprint for best clinical practice. Jauhar et al. (2016) assert that via NICE quality standards, guideline recommendations are able to penetrate the decisions health commissioners in the UK make about the services that are to be commissioned.

This reach of guidelines further extends to a significant role in the education and training of health care professionals including the doctoral training courses in clinical psychology. For instance the British Psychological Society's (BPS) standards for accrediting doctoral courses are informed by evidence-based practice guidelines such as NICE. The BPS acknowledges that such guidance "should not be applied in any formulaic fashion – especially pertinent when dealing with the complexity and co-morbidity for which training needs to prepare clinical psychologists" (BPS, 2015; p, 15- 16). However, the reach of guidelines and their narrative of the evidence-based psychotherapy recommendations will continue to bear down heavily upon training providers, psychotherapy practitioners and psychological services that are left to negotiate the expectations of commissioners, interdisciplinary colleagues and service users based on over simplistic aspects of guideline evidence review.

5.7 Suggestions for further research

5.7.1 Assessing the evidence strength in other NICE guidelines

Further research should focus on whether similar patterns of the strength of evidence for CBT and other psychotherapies exist within other guidelines. Expansion of this question into other guidelines would assist in understanding whether or not the inadequacies reported here in respect to the guideline approach to reviewing psychotherapy evidence are generalizable. Furthermore, it would assist in understanding the extent to which CBT's status as a frontline intervention in other guidelines is evidence-based. This would help to determine the generalisability of these findings and to ascertain whether or not the approach used within this study to reviewing guideline evidence is useful for other problem-based guidelines.

5.7.2 Incorporating broader outcomes measures relevant to psychotherapies

Further research should focus on reviewing broader outcomes for CBT-based interventions and other psychotherapies recommended for depression. This is important in addressing whether broader outcomes that are theoretically relevant to different psychotherapies can be incorporated into the evidence base for guideline recommendations. Such primary outcome measures would include those of interpersonal and intrapersonal experiences, social, and occupational functioning, for which strong arguments for their inclusion have been previously highlighted (Barkham et al., 2001; Pilling & Price, 2006). Psychotherapy guidance within a medical context that is primarily symptom-focussed limits the capturing of the full range of benefits that psychotherapy has for service users. Widening the scope of outcomes, including outcome measures specific to the mode of therapy that were often excluded from NICE's review, could enable more accurate evaluation of diverse treatments based on what psychotherapy practitioners within a healthcare context are actually targeting in therapy.

Psychotherapy guidance would further benefit from research incorporating alternative practice-based evidence that involve the use of more generic measures of psychological distress such as the Clinical Outcomes for Routine Evaluation (CORE; Barkham et al., 2001). The CORE is pan-theoretical (i.e., not associated with a school of therapy), pan-diagnostic (i.e. not focused on a single presenting problem) and draws upon the views of what practitioners consider to be the most important generic aspects of psychological wellbeing. Such an approach to measuring outcome has the potential to assist in avoiding the narrow notions of psychotherapy outcome currently used, as previously discussed (see Guy et al., 2012), and has demonstrated utility in assessing a range of psychological interventions.

5.7.3 Continued scrutiny of evidence based on guideline characteristics and evidence quality

NICE guideline evidence for psychotherapies should continue to be scrutinised rigorously through appraisals of the evidence base utilising both quantitative approaches and through more qualitative approaches that Pearce (2015) argues require the self-reflection of practitioners on 'the inbuilt assumptions within evidence architectures' (p. 6). For instance, further quantitative research should focus on re-analyses of guideline evidence using more robust meta-analytic techniques to correct for factors known to lead to overestimations of psychotherapy effects, as demonstrated by Cuijpers and colleagues (2008). Research focussed on re-analyses of CBT evidence within the NICE guidance for schizophrenia and bipolar guidelines by Jauhar and colleagues are encouraging steps. Furthermore, NICE's current method of addressing a wide scope of clinical questions through multiple meta-analyses, often based on small samples and individual studies, should be revised as this amounts to stretching the data further than it will go. Future research in this area would also benefit from further narrative reviews of primary

studies and how these are interpreted as a basis for guideline evidence and psychotherapy recommendations (e.g. Nel, 2014).

5.7.4 Examining the role of investigator allegiance in psychotherapy guidance

In the initial planning phases of this project I was interested in whether well documented allegiance effects in comparative outcome studies, meta-analyses and reviews (e.g. Robinson et al., 1990) can also predict which psychotherapies are most likely to be recommended by NICE. However, preliminary attempts to estimate allegiances of the entire guideline development group (GDG) through review of existing NICE documents were unsuccessful. This was due to there being a large proportion of GDG members who were not clinical academics or psychotherapy practitioners.

Such allegiance bias has previously been tested experimentally during a guideline development process to assess its impact on the interpretation of evidence used for guidance on psychotherapy and counselling (see DoH, 2001). This demonstrated that therapeutic allegiance influenced the interpretations of psychotherapy practitioners oriented to cognitive-behavioural therapies to a greater degree than others (see DoH, 2001).

Within the wider literature there is a further suggestion that factors associated with researcher allegiance play a role in the guideline development process. Such factors include professional affiliations, diversity of opinions, vested interests and broader world view (Atkins et al., 2013). In addition to this Pagliari and Grimshaw (2002) have suggested that there is a marked effect of professional role and status on the contribution of group members to the guideline develop process. Furthermore, Pagliari, Grimshaw and Eccles (2001) suggest a range of psychosocial factors that can influence the process by which guideline development groups interact, make

decisions and achieve consensus. They argue that as multidisciplinary groups are also multi-status groups, there is ample opportunity for psychosocial factors to intervene in key aspects of the guideline development process including systematic appraisal of the evidence base used. Small group processes are considered to be core psychosocial processes within guideline development and ones in which an interplay between minority influence and investigator allegiance could occur within the interpretation of the evidence.

Therefore, it is expected that a combination of personal views, multi-professional interests, small group processes and investigator allegiance are likely to have influenced the supposedly evidence based recommendations within the NICE depression guideline. Moreover, research into the guideline development process has found such factors to correlate strongly with recommendations and have a comparable impact on them to that of the research evidence base (Raine et al, 2004; Burgers et al, 2002). Pagliari and colleagues suggest that this could have important implications for the validity and the reliability of the recommendations produced. Thus, a further recommendation for future research is to measure theoretical or therapeutic allegiance within guideline panels and to explore its relationship with interpretation of the evidence.

5.8 Returning to my social constructionist stance

When I initially shared my stance as a researcher (see section 3.5), I stated my intention to be reflective and reflexive concerning my role as a social 'actor' in the construction and interpretation of the evidence. Throughout this process I have attempted to hold in mind my actions as a researcher alongside the findings of this review and the potential link between the two.

My social constructionist stance encouraged a deconstruction of taken-for-granted knowledge in respect to NICE's approach to recommending psychotherapies for depression. Guideline evidence was re-examined through this lens and this guided many of my choices throughout this project. For instance, there is a critical thread throughout this review, which connects the quantitative findings with wider elements of the guideline development process (e.g. the selection of primary outcomes) that sit neatly within a medical context. Within this thread I recognise myself as a clinician interested in the wider spectrum of psychotherapy evidence and a breadth of approaches available. I also recognise my tendency as a therapist to search for and develop alternative, less privileged narratives, which, in the age of evidence-based practice and NICE, other approaches to psychotherapy far too often reflect. Thus, it is difficult to disentangle myself from the conduct and the outcome of this review.

This also prompts me to reflect upon the role my perspective as a clinical psychologist (in training) has had on the process. I believe the core skills of our profession extend much further than guideline recommendations are able to go. Court, Cooke and Scrivener (2016) suggest that this is a common belief within the profession in relation to clinical practice guidelines. I share concerns similar to those reported by clinical psychologists interviewed within their study, namely about NICE guideline's potential to restrict the full range of professional skills that offer flexibility to clinicians and choice to service users. In turn, these beliefs partly contribute to the lack of implementation of guidelines amongst practitioners through active choice. I was mindful throughout of a sense of responsibility to challenge the current status quo and advocate for the practitioners' perspective that is often unvoiced. I suspect that this would have contributed to further attempts to deconstruct the guideline evidence.

As an actor, holding such beliefs it is hard to judge the extent to which I shaped the direction of this review and its findings. These were checked at each stage with my supervisors, although they too were not without their own biases! However, as difficult as this is to disentangle, I think it is important to acknowledge the potential bias that stems from my choices and interpretations of the evidence related to my stance. It is particularly important to acknowledge this when one considers the unusual marriage between philosophy and method. However, from a social constructionist position I would argue that bias is inherent within all approaches to research including the more traditional stances to evidence reviews and outcome within guideline programmes that are based upon empirical traditions. This is similar to Cooper's (2008) argument that within the broader field of comparative outcome research it is how one slices the cake that typically determines outcome. This is apparent in the contrasting findings from psychotherapy researcher outlined earlier in this review. It is also evident in researcher's with opposing theoretical allegiances and findings from different research paradigms (e.g. specific and common factors), which utilise a range of different research designs and techniques. I think this represents a dilemma that all researchers face but often goes unacknowledged and I hope my endeavours to remain clear about my position and consider its impact on outcome and interpretation have helped maintain a balance within this review.

5.9 Conclusions

This project sought to better understand the evidence base constructed for psychotherapies within NICE guidelines for depression in adults. This project utilised an exploratory data analysis approach to examine secondary evidence generated within the NICE evidence review process. This enabled a focus on the strength of the

evidence that supports recommendations for CBT and other psychological therapies within the guideline.

This project's exploration of the evidence for CBT and other psychotherapies questions the strength of the evidence base for CBT as a front line psychological intervention for the treatment of depression, particularly when compared against the strength of the collective evidence for other psychotherapies. Furthermore, the relationship between evidence and quality for the two psychotherapy groups suggests that larger amounts of low quality evidence are associated with favourable effects for CBT than other psychotherapies. The limitations of the project's findings have been considered. However, this project makes an important contribution to understanding how guideline evidence for psychological treatments for depression is constructed to fit within a medical context, placing other psychotherapies available to clients and practitioners at a disadvantage from enjoying similar levels of support.

6. References

Alderson, P., & Tan, T. (2011). The use of Cochrane Reviews in NICE clinical guidelines. *The Cochrane database of systematic reviews*(9), 10.1002/14651858.ED000032.

Alvesson, M., & Skoldberg, K. (2000). *Reflexive Methodology: New Vistas for Qualitative Research*. London: SAGE Publications Ltd

Andersson, G., & Cuijpers, P. (2009). 'Psychological treatment' as an umbrella term for evidence-based psychotherapies? *Nordic Psychology,61*(2), 4-15.

Atkins, L., Smith, J. A., Kelly, M. P., & Michie, S. (2013). The process of developing evidence-based guidance in medicine and public health: a qualitative study of views from the inside. *Implementation Science*, *8*(1), 1.

Bagby, R. M., Ryder, A. G., Schuller, D. R., & Marshall, M. B. (2004). The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *American Journal of Psychiatry, 161*(12), 2163-2177.

Barker, C., & Pistrang, N. (2015). *Research methods in clinical psychology: An introduction for students and practitioners.* Chichester: John Wiley & Sons.

Barker, C., Pistrang, N. & Elliott, R. (1994). *Research Methods in Clinical and Counseling Psychology.* Chichester: John Wiley & Sons

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., ... & McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*, *69*(2), 184.

Barkham, M., & Parry, G. (2008). Balancing rigour and relevance in guideline development for depression: The case for comprehensive cohort studies. *Psychology and Psychotherapy: Theory, Research and Practice, 81*(4), 399-417.

Barlow, D. H. (2004). Psychological treatments. *American psychologist, 59*(9), 869-878.

Beck, A. T. (1961). A systematic investigation of depression. *Comprehensive psychiatry, 2*(3), 163-170.

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods, 2*(2), 131-160.

Bell, S., Clark, D. M., Layard, R., Meacher, M. & Priebe, S. (2006). The depression report: A new deal for depression and anxiety disorders. *Health Policy*: London

Boltanski, L., & Thévenot, L. (1999). The sociology of critical capacity. *European journal of social theory, 2*(3), 359-377.

British Psychological Society (2015). *Standards for the accreditation of Doctoral programmes in counselling psychology*. Retrieved 23 November 2016 from, http://www.bps.org.uk/system/files/Public%20files/PaCT/counselling_accreditation_2015_web.pdf

Bronfenbrenner, U. and Morris, P. A. (2006). The Bioecological Model of Human Development. *Handbook of Child Psychology, 1*(14), 795-825. DOI/10.1002/9780470147658.chpsy0114.Burgers, J. S., Bailey, J. V., Klazinga, N. S., Van der Bij, A. K., Grol, R., & Feder, G. (2002). Inside Guidelines Comparative analysis of recommendations and evidence in diabetes guidelines from 13 countries. *Diabetes Care, 25*(11), 1933-1939.

Burr, V. (2002). *An Introduction to Social Constructionism.* London: Taylor & Francis.

Burr, V. (2004). *Social Constructionism.* London: Taylor & Francis.

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clinical psychology review, 26*(1), 17-31.

Castonguay, L. G., Muran, J. C, Strauss, B., (2010). A brief introduction to psychotherapy research. In Castonguay, L. G., Muran, J. C., Angus, L. E., Hayes, J. A., Ladany, N. E., & Anderson, T. E., (Eds.). *Bringing psychotherapy research to life: Understanding change through the work of leading clinical researchers.* Washington, DC: American Psychological Association.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of consulting and clinical psychology, 66*(1), 7-18.

Chiles, J. A., Lambert, M. J., & Hatch, A. L. (1999). The impact of psychological interventions on medical cost offset: A meta-analytic review. *Clinical Psychology: Science and Practice, 6*(2), 204-220.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) Hillsdale, NJ: Lawrence Erlbaum Associates.

Consumer Reports (1995). Mental health: does therapy help? 734-739.

Cooper, M. (2008). *Essential research findings in counselling and psychotherapy: The facts are friendly.* London: Sage.

Court, A. J., Cooke, A., & Scrivener, A. (2016). They're NICE and Neat, but Are They Useful? A Grounded Theory of Clinical Psychologists' Beliefs About and Use of NICE Guidelines. *Clinical Psychology & Psychotherapy.* DOI: 10.1002/cpp.2054

Cuijpers, P., Beekman, A. T., & Reynolds, C. F. (2012). Preventing depression: a global priority. *Jama*, *307*(10), 1033-1034.

Cuijpers, P., Driessen, E., Hollon, S. D., van Oppen, P., Barth, J., & Andersson, G. (2012). The efficacy of non-directive supportive therapy for adult depression: a meta-analysis. *Clinical psychology review, 32*(4), 280-291.

Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychological medicine, 40*(02), 211-223.

Cuijpers, P., Van Straten, A., Warmerdam, L., & Smits, N. (2008). Characteristics of effective psychological treatments of depression: a metaregression analysis. *Psychotherapy Research, 18*(2), 225-236.

Danziger, K. (1997). The Varieties of Social Construction. *Theory & Psychology, 7*(3), 399-416. Demyttenaere, K., & De Fruyt, J. U. R. (2003). Getting what you ask for: on the selectivity of depression rating scales. *Psychotherapy and psychosomatics, 72*(2), 61-70.

Department of Health (1997). *The new NHS modern. dependable*. Retrieved 2 November 2016 from, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/266003/newnhs.pdf

Department of Health (2001) *Treatment Choices in Psychological Therapies and Counselling*. London: Department of Health.

Egger, M., Davey-Smith, G., & Altman, D. (2008). *Systematic reviews in health care: meta-analysis in context.* Chichester: John Wiley & Sons.

Elliott, R. (2010). Psychotherapy change process research: Realizing the promise. *Psychotherapy research, 20*(2), 123-135.

Elliott, R., Fischer, C. T., & Rennie, D. L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British journal of clinical psychology, 38*(3), 215-229.

Field, M. J., & Lohr, K. N. (Eds.). (1990). *Clinical practice guidelines: directions for a new program* (Vol. 90, No. 8). National Academies Press.

Frank, J. D., & Frank, J. B. (1993). *Persuasion and healing: A comparative study of psychotherapy.* Baltimore: Johns Hopkins University Press.

Gaffan, E., Tsaousis, J., & Kemp-Wheeler, S. (1995). Researcher allegiance and meta-analysis: the case of cognitive therapy for depression. *Journal of consulting and clinical psychology, 63*(6), 966.

Gray, J. A. M. (2001). *Evidence-based healthcare.* Elsevier Health Sciences.

Greene, C. J., Morland, L. A., Durkalski, V. L., & Frueh, B. C. (2008). Noninferiority and equivalence designs: issues and implications for mental health research. *Journal of traumatic stress, 21*(5), 433-439.

Guy, A., Thomas, R., Stephenson, S., & Loewenthal, D. (2012). NICE under scrutiny: The impact of the National Institute for Health and Clinical Excellence guidelines on the provision of psychotherapy in UK. *The Journal of Psychological Therapies in Primary Care, 1*(1-2), 77-102.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery & Psychiatry, 23*(1), 56-62.

Hansen, N. B., Lambert, M. J. & Forman, E. M. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9*(3), 329-343.

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological methods, 3*(4), 486.

Higgins, J. P., & Green, S. (Eds.). (2008). *Cochrane handbook for systematic reviews of interventions* (Vol. 5). Chichester: Wiley-Blackwell.

Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: a review of meta-analyses. *Cognitive therapy and research, 36*(5), 427-440.

Holmes, J. (2002). All you need is cognitive behaviour therapy?. *British Medical Journal, 324*(7332), 288-294.

Hunot, V., Moore, T. H., Caldwell, D. M., Furukawa, T. A., Davies, P., Jones, H., . . . Churchill, R. (2013). Third wave' cognitive and behavioural therapies versus other psychological therapies for depression. *The Cochrane Library*.

Jauhar, S., McKenna, P. J., Radua, J., Fung, E., Salvador, R., & Laws, K. R., (2014). Cognitive-behavioural therapy for the symptoms of schizophrenia: systematic review and meta-analysis with examination of potential bias. *The British Journal of Psychiatry, 204*(1), 20-29.

Jauhar, S., McKenna, P. J., & Laws, K. R. (2016). NICE guidance on psychological treatments for bipolar disorder: searching for the evidence. *The Lancet Psychiatry, 3*(4), 386-388.

Karasu, T. B. (1986). The specificity versus nonspecificity dilemma: Toward identifying therapeutic change agents. *The American Journal of Psychiatry, 143*, 687-695.

Kendall, T., Morriss, R., Mayo-Wilson, E., Meyer, T. D., Jones, S. H., Oud, M., & Baker, M. R. (2016). NICE guidance on psychological treatments for bipolar disorder. *The Lancet Psychiatry*, *3*(4), 317-320.

King, M. W., & Resick, P. A. (2014). Data mining in psychological treatment research: A primer on classification and regression trees. *Journal of consulting and clinical psychology, 82*(5), 895.

Kirsch, I., Deacon, B. J., Huedo-Medina, T. B., Scoboria, A., Moore, T. J., & Johnson, B. T. (2008). Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med, 5*(2), e45. https://doi.org/10.1371/journal.pmed.0050045.

Kopta, S. M., Howard, K. I., Lowry, J. L., & Beutler, L. E. (1994). Patterns of symptomatic recovery in psychotherapy. *Journal of consulting and clinical psychology*, 62(5), 1009-1016.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of general psychiatry, 59*(10), 877-883.

Lambert, M. J. (2004). *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th edn.). Chicago: John Wiley & Sons.

Lambert, M. J. & Ogles, B. M. (2004). 'The efficacy and effectiveness of psychotherapy supervision', in C. E. Watkins (ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th edn). Chicago: John Wiley& Sons, pp. 139-93.

Lambert, M. J. (2013a). *Bergin and Garfield's handbook of psychotherapy and behavior change.* Chichester: John Wiley & Sons.

Lambert, M. J. (2013b). Outcome in psychotherapy: The past and important advances. *Psychotherapy* 50(1), 42–51

Laska, K. M., Gurman, A. S., & Wampold, B. E. (2014). Expanding the lens of evidence-based practice in psychotherapy: A common factors perspective. *Psychotherapy, 51*(4), 467-481.

Leng, G. (2009). NHS Evidence: raising standards. *Prescriber, 20*(20), 8-10.

Lépine, J.-P., & Briley, M. (2011). The increasing burden of depression. *Neuropsychiatr Dis Treat, 7*(Suppl 1), 3-7.

Lock, A., & Strong, T. (2010). *Social Constructionism: Sources and Stirrings in Theory and Practice.* Cambridge, UK: Cambridge University Press.

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

Mollon, P. (2009). The NICE guidelines are misleading, unscientific, and potentially impede good psychological care and help. *Psychodynamic Practice, 15*(1), 9-24.

Montgomery, S. A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *The British journal of psychiatry, 134*(4), 382-389.

Moreira, T. (2007). Entangled evidence: knowledge making in systematic reviews in healthcare. *Sociology of Health & Illness, 29*(2), 180-197.

NCCMH (2009). *Borderline Personality Disorder: Treatment and Management*. Leicester and London: The British Psychological Society and the Royal College of Psychiatrists. [Full guideline]

NCCMH (2010). *Depression: the Treatment and Management of Depression in Adults (Update).* Updated edition. Leicester and London: The British Psychological Society and the Royal College of Psychiatrists. [Full guideline]

Nel, P. W. (2014). The NICE guideline on the treatment of child and adolescent depression: a meta-review of the evidence for individual CBT. *European Journal of Psychotherapy & Counselling, 16*(3), 267-287.

NICE (2009) *Schizophrenia: Core Interventions in the Treatment and Management of Schizophrenia in Adults in Primary and Secondary Care.* NICE clinical guideline 82. Available at www.nice.org.uk/CG82 [NICE guideline]

NICE (2009) *The guidelines manual.* London: National Institute for Health and Clinical Excellence. Available from: www.nice.org.uk

NICE (2014) *The guidelines manual*. London: National Institute for Health and Clinical Excellence. Available from: www.nice.org.uk

Nilsson, C. (2010). A phenomenological approach to practical knowledge in psychotherapy. *Santalka: Filologija, Edukologija,* (3), 70-80.

Norcross, J. C. (2002). *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients*. Oxford University Press.

Oxman, A. D., & GRADE Working Group. (2004). Grading quality of evidence and strength of recommendations. *British Medical Journal, 328*(19), 1490-1494.

Pagliari, C., & Grimshaw, J. (2002). Impact of group structure and process on multidisciplinary evidence-based guideline development: an observational study. *Journal of evaluation in clinical practice, 8*(2), 145-153.

Pagliari, C., Grimshaw, J., & Eccles, M. (2001). The potential influence of small group processes on guideline development. *Journal of evaluation in clinical practice, 7*(2), 165-173.

Parry, G. (1992). Improving psychotherapy services: applications of research, audit and evaluation. *British Journal of Clinical Psychology*, 31(1), 3-19.

Parry, G., Cape, J., & Pilling, S. (2003). Clinical practice guidelines in clinical psychology and psychotherapy. *Clinical Psychology & Psychotherapy, 10*(6), 337-351.

Pearce, W., Raman, S., & Turner, A. (2015). Randomised trials in context: practical problems and social aspects of evidence-based medicine and policy. *Trials, 16*(1), 1-7.

Piaggio, G., Elbourne, D. R., Altman, D. G., Pocock, S. J., Evans, S. J., & Group, C. (2006). Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *Jama, 295*(10), 1152-1160.

Pilling, S. (2008). History, context, process, and rationale for the development of clinical guidelines. *Psychology and psychotherapy: Theory, research and practice, 81*(4), 331-350.

Pilling, S. (2009). Developing evidence-based guidance–implications for systemic interventions. *Journal of Family Therapy, 31*(2), 194-205.

Pilling, S., & Price, K. (2006). Developing and implementing clinical guidelines: lessons from the NICE schizophrenia guideline. *Epidemiologia e Psichiatria Sociale, 15*(02), 109-116.

Pokorny, D. (2015). Quantitative Data Analysis in Psychotherapy Process Research: Structures and Procedures *Psychotherapy Research*, 229-245: Springer Vienna.

Raine, R., Sanderson, C., Hutchings, A., Carter, S., Larkin, K., & Black, N. (2004). An experimental study of determinants of group judgments in clinical guideline development. *The Lancet, 364*(9432), 429-437.

Reynolds, S. (2008). *Evidence-based practice: a critical appraisal.* Trinder L. (Ed.). John Wiley & Sons.

Robinson, L. A., Berman, J. S., & Neimeyer, R. A. (1990). Psychotherapy for the treatment of depression: a comprehensive review of controlled outcome research. *Psychological bulletin, 108*(1), 30-49.

Roth, A., & Fonagy, P. (2005). *What works for whom?: A critical review of psychotherapy research.* (2nd ed.) New York: Guilford Publications.

Ryle, A. (1982). *Psychotherapy, a cognitive integration of theory and practice*. Academic Press.

Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal, 312*(7023), 71-72.

Shadish, W. R., Matt, G. E., Navarro, A. M. and Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: a meta-analysis. *Psychological Bulletin, 126*(4): 512-529

Slade, M., & Priebe, S. (2001). Are randomised controlled trials the only gold that glitters? *The British Journal of Psychiatry, 179*(4), 286-287.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American psychologist, 32*(9), 752.

Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Johns Hopkins University Press.

Stiwne, D., & Abrandt Dahlgren, M. (2004). Challenging evidence in evidence-based practice. In Higgs, J., Richardson, B., & Abrandt Dahlgren, M. (2004). *Developing practice knowledge for health professionals*. Butterworth & Heinemann.

Strupp, H. H. (1978). The therapist's theoretical orientation: An overrated variable. *Psychotherapy: Theory, Research & Practice, 15*(4), 314-317.

Tolin, D. F. (2010). Is cognitive–behavioral therapy more effective than other therapies?: A meta-analytic review. *Clinical psychology review, 30*(6), 710-720.

Trinder, L. (2000). A critical appraisal of evidence-based practice. *Evidence-based practice: A critical appraisal*, 212-241.

Trinder, L., & Reynolds, S. (2008). *Evidence-Based Practice: A Critical Appraisal.* Chichester: John Wiley & Sons.

Tufford, L., & Newman, P. (2012). Bracketing in qualitative research. *Qualitative Social Work, 11*(1), 80-96.

Tukey, J. W. (1980). We need both exploratory and confirmatory. *The American Statistician, 34*(1), 23-25.

Turner, E. H. (2004). A taxpayer-funded clinical trials registry and results database. *PLoS Med, 1*(3), e60. https://doi.org/10.1371/journal.pmed.0010060

Vigni, M. L., Durante, C., & Cocchi, M. (2013). Exploratory data analysis. *Chemometrics in food chemistry, 28*, 55-126.

Wampold, B. E. (2001). *The Great Psychotherapy Debate: Models, Methods, and Findings.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Wampold, B. E. (2007). Psychotherapy: The humanistic (and effective) treatment. *American Psychologist, 62*, 857– 873.

Wampold, B. E., Minami, T., Baskin, T. W., & Tierney, S. C. (2002). A meta-(re) analysis of the effects of cognitive therapy versus 'other therapies' for depression. *Journal of affective disorders, 68*(2), 159-165.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. N. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empiricially,"all must have prizes.". *Psychological bulletin, 122*(3), 203.

Weiner, I. B., Freedheim, D. K., Schinka, J. A., & Velicer, W. F. (2003). *Handbook of Psychology, Research Methods in Psychology.* Chichester: John Wiley & Sons.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. *Psychological bulletin, 130*(4), 631.

White, M., & Epston, D. (1990). *Narrative means to therapeutic ends*. WW Norton & Company.

Williams, C. (2015). Improving access to psychological therapies (IAPT) and treatment outcomes: epistemological assumptions and controversies. *Journal of psychiatric and mental health nursing, 22*(5), 344-351.

Winter, D. (2010). Editorial - allegiance revisited. *European Journal of Psychotherapy and Counselling, 12*(1), 3-9. 10.1080/13642531003637726

7. Appendices

7. Appendices

7.1 Appendix 1

[subsection 4.2] Overall test of difference between the CBT and other psychotherapies group (treatment intensity combined)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	130	-.1634	.31378	.02752
	Non-CBT	93	-.2947	.49830	.05167

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	22.549	.000	2.412	221	.017	.13135	.05447	.02401	.23869
	Equal variances not assumed			2.244	143.369	.026	.13135	.05854	.01563	.24707

[subsection 4.2] Test of difference between recommended psychotherapy groups for favourable effects sizes (0.2 and above)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	63	-.4151	.21490	.02707
	Non-CBT	46	-.6893	.37480	.05526

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	26.309	.000	4.827	107	.000	.27427	.05682	.16163	.38691
	Equal variances not assumed			4.457	66.419	.000	.27427	.06154	.15142	.39712

[Subsection 4.2] Test of difference between recommended psychotherapy groups by low quality outcome rating (combined treatment intensity)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	72	-.0874	.29405	.03465
	Non-CBT	57	-.0777	.31019	.04109

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	.241	.624	-.181	127	.857	-.00964	.05341	-.11534	.09605
	Equal variances not assumed			-.179	117.227	.858	-.00964	.05375	-.11609	.09680

[Subsection 4.2] Test of difference between recommended psychotherapy groups by low quality outcome (controlling for treatment intensity)

T-Test

Group Statistics

Guideline Code		Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
CG90(H)	Raw Effect Size	CBT	56	-.0584	.28909	.03863
		Non-CBT	36	-.0544	.26629	.04438
CG90(L)	Raw Effect Size	CBT	16	-.1887	.29790	.07448
		Non-CBT	21	-.1176	.37772	.08242

Independent Samples Test

Guideline Code			Levene's Test for Equality of Variances		t-test for Equality of Means						
			F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
										Lower	Upper
CG90(H)	Raw Effect Size	Equal variances assumed	2.470	.120	-.066	90	.948	-.00395	.05991	-.12297	.11507
		Equal variances not assumed			-.067	79.197	.947	-.00395	.05884	-.12106	.11317
CG90(L)	Raw Effect Size	Equal variances assumed	.348	.559	-.620	35	.539	-.07113	.11474	-.30407	.16181
		Equal variances not assumed			-.640	34.938	.526	-.07113	.11109	-.29666	.15440

[Subsection 4.2] Test of difference between recommended psychotherapy groups by moderate quality outcome (combined treatment intensity)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	39	-.2013	.27257	.04365
	Non-CBT	30	-.5977	.56762	.10363

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	18.662	.000	3.830	67	.000	.39638	.10349	.18982	.60295
	Equal variances not assumed			3.525	39.258	.001	.39638	.11245	.16898	.62379

[Subsection 4.2] Test of difference between recommended psychotherapy groups by moderate quality outcome (controlling for treatment intensity)

T-Test

Group Statistics

Guideline Code		Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
CG90(H)	Raw Effect Size	CBT	29	-.1131	.22612	.04199
		Non-CBT	18	-.3761	.54415	.12826
CG90(L)	Raw Effect Size	CBT	10	-.4570	.23856	.07544
		Non-CBT	12	-.9300	.43706	.12617

Independent Samples Test

Guideline Code			Levene's Test for Equality of Variances		t-test for Equality of Means						
										95% Confidence Interval of the Difference	
			F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
CG90(H)	Raw Effect Size	Equal variances assumed	19.185	.000	2.312	45	.025	.26301	.11374	.03393	.49208
		Equal variances not assumed			1.949	20.695	.065	.26301	.13495	-.01790	.54391
CG90(L)	Raw Effect Size	Equal variances assumed	3.353	.082	3.056	20	.006	.47300	.15478	.15014	.79586
		Equal variances not assumed			3.218	17.533	.005	.47300	.14700	.16357	.78243

[Subsection 4.2] Test of difference between recommended psychotherapy groups by high quality outcome (combined treatment intensity)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	19	-.3737	.36711	.08422
	Non-CBT	6	-.8417	.41945	.17124

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	.859	.364	2.636	23	.015	.46798	.17753	.10073	.83523
	Equal variances not assumed			2.452	7.588	.041	.46798	.19083	.02374	.91223

[Subsection 4.2] Test of difference between recommended psychotherapy groups by high quality outcome (controlling for treatment intensity)

T-Test

Group Statistics

Guideline Code		Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
CG90(H)	Raw Effect Size	CBT	14	-.3164	.27385	.07319
		Non-CBT	1	-1.3500	.	.
CG90(L)	Raw Effect Size	CBT	5	-.5340	.56492	.25264
		Non-CBT	5	-.7400	.37736	.16876

Independent Samples Test

			Levene's Test for Equality of Variances		t-test for Equality of Means						
										95% Confidence Interval of the Difference	
Guideline Code			F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
CG90(H)	Raw Effect Size	Equal variances assumed	.	.	3.646	13	.003	1.03357	.28346	.42119	1.64595
		Equal variances not assumed			.	.	.	1.03357	.	.	.
CG90(L)	Raw Effect Size	Equal variances assumed	.421	.535	.678	8	.517	.20600	.30382	-.49461	.90661
		Equal variances not assumed			.678	6.977	.520	.20600	.30382	-.51290	.92490

7.2 Appendix 2

[Subsection 4.2.1.1] Test of difference between recommended psychotherapy groups for high intensity depression

T-Test

Guideline Code = CG90(H)

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	99	-.1109	.28120	.02826
	Non-CBT	55	-.1833	.43341	.05844

a. Guideline Code = CG90(H)

Independent Samples Test[a]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	9.546	.002	1.254	152	.212	.07236	.05770	-.04163	.18636
	Equal variances not assumed			1.115	79.805	.268	.07236	.06492	-.05683	.20156

a. Guideline Code = CG90(H)

[4.2.1.1] Test of difference between recommended psychotherapy groups for low intensity depression

Guideline Code = CG90(L)

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	31	-.3310	.35630	.06399
	Non-CBT	38	-.4561	.54595	.08856

a. Guideline Code = CG90(L)

Independent Samples Test[a]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	7.132	.009	1.098	67	.276	.12508	.11389	-.10224	.35241
	Equal variances not assumed			1.145	64.153	.257	.12508	.10926	-.09319	.34336

a. Guideline Code = CG90(L)

[Subsection 4.2.1.1] Test of difference between recommended psychotherapy groups based on favourable effects sizes 0.2 and above (controlling for treatment intensity)

T-Test

Group Statistics

Guideline Code		Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
CG90(H)	Raw Effect Size	CBT	41	-.3795	.14018	.02189
		Non-CBT	21	-.6243	.33337	.07275
CG90(L)	Raw Effect Size	CBT	22	-.4814	.30305	.06461
		Non-CBT	25	-.7440	.40492	.08098

Independent Samples Test

Guideline	Code			Levene's Test for Equality of Variances		t-test for Equality of Means						
											95% Confidence Interval of the Difference	
				F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
CG90(H)	Raw Effect Size	Equal variances assumed		32.992	.000	4.073	60	.000	.24477	.06009	.12457	.36497
		Equal variances not assumed				3.222	23.689	.004	.24477	.07597	.08787	.40168
CG90(L)	Raw Effect Size	Equal variances assumed		4.210	.046	2.489	45	.017	.26264	.10552	.05010	.47517
		Equal variances not assumed				2.535	43.934	.015	.26264	.10360	.05384	.47144

7.3 Appendix 3

T-tests repeated (post-hoc) controlling for treatment comparators

[Subsection 4.2.1.2] Non-active interventions comparator (high and low intensity combined)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	37	-.3195	.36241	.05958
	Non-CBT	44	-.4977	.52813	.07962

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	7.997	.006	1.737	79	.086	.17827	.10262	-.02600	.38253
	Equal variances not assumed			1.793	76.128	.077	.17827	.09944	-.01978	.37632

[Subsection 4.2.1.2] Non-active interventions comparator (high intensity)

T-Test

Guideline Code = CG90(H)

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	10	-.1500	.34679	.10967
	Non-CBT	18	-.4617	.49556	.11680

a. Guideline Code = CG90(H)

Independent Samples Test[a]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	1.815	.190	1.757	26	.091	.31167	.17735	-.05289	.67622
	Equal variances not assumed			1.945	24.387	.063	.31167	.16022	-.01873	.64206

a. Guideline Code = CG90(H)

[Subsection 4.2.1.2] Non-active interventions comparator (low intensity)

Guideline Code = CG90(L)

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	27	-.3822	.35368	.06807
	Non-CBT	26	-.5227	.55783	.10940

a. Guideline Code = CG90(L)

Independent Samples Test[a]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	7.369	.009	1.099	51	.277	.14047	.12779	-.11608	.39702
	Equal variances not assumed			1.090	42.043	.282	.14047	.12884	-.11954	.40048

a. Guideline Code = CG90(L)

[Subsection 4.2.1.2] Active interventions comparator (high and low intensity depression combined)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	29	.0138	.22188	.04120
	Non-CBT	23	-.1478	.35604	.07424

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	2.025	.161	2.005	50	.050	.16162	.08061	-.00029	.32353
	Equal variances not assumed			1.903	35.029	.065	.16162	.08491	-.01074	.33398

[Subsection 4.2.1.2] Active interventions comparator (high intensity depression)

T-Test

Guideline Code = CG90(H)

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	25	.0136	.23909	.04782
	Non-CBT	17	-.1571	.31396	.07615

a. Guideline Code = CG90(H)

Independent Samples Test[a]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	.509	.480	1.999	40	.052	.17066	.08536	-.00186	.34317
	Equal variances not assumed			1.898	28.186	.068	.17066	.08992	-.01347	.35479

a. Guideline Code = CG90(H)

[Subsection 4.2.1.2] Active interventions comparator (low intensity depression)

Guideline Code = CG90(L)

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	4	.0150	.04655	.02327
	Non-CBT	6	-.1217	.49114	.20051

a. Guideline Code = CG90(L)

Independent Samples Test[a]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	2.468	.155	.544	8	.601	.13667	.25131	-.44285	.71618
	Equal variances not assumed			.677	5.134	.528	.13667	.20185	-.37816	.65150

a. Guideline Code = CG90(L)

[Subsection 4.2.1.2] Medication comparator (high and low intensity depression combined)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	23	-.1239	.30167	.06290
	Non-CBT	14	-.0471	.38778	.10364

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	.034	.855	-.674	35	.505	-.07677	.11398	-.30816	.15462
	Equal variances not assumed			-.633	22.535	.533	-.07677	.12124	-.32785	.17431

[Subsection 4.2.1.2] Medication comparator (high intensity depression)

T-Test

Guideline Code = CG90(H)

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	23	-.1239	.30167	.06290
	Non-CBT	11	.1191	.13057	.03937

a. Guideline Code = CG90(H)

Independent Samples Test[a]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	8.254	.007	-2.544	32	.016	-.24300	.09552	-.43757	-.04843
	Equal variances not assumed			-3.275	31.857	.003	-.24300	.07421	-.39419	-.09182

a. Guideline Code = CG90(H)

Guideline Code = CG90(L)

Warnings

No statistics are computed for a split file in the Independent Samples table. The split file is: Guideline Code=CG90(L).

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	0[b]	.	.	.
	Non-CBT	3	-.6567	.42771	.24694

a. Guideline Code = CG90(L)

b. t cannot be computed because at least one of the groups is empty.

[Subsection 4.2.1.2] Recommended psychotherapy and medication combined vs psychotherapy

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	18	-.0128	.19152	.04514
	Non-CBT	1	.0400	.	.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	.	.	-.268	17	.792	-.05278	.19677	-.46792	.36237
	Equal variances not assumed			.	.	.	-.05278	.	.	.

[Subsection 4.2.1.2] Recommended psychotherapy and medication combined vs psychotherapy (High intensity)

T-Test

Guideline Code = CG90(H)

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	18	-.0128	.19152	.04514
	Non-CBT	1	.0400	.	.

a. Guideline Code = CG90(H)

Independent Samples Test[a]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	.	.	-.268	17	.792	-.05278	.19677	-.46792	.36237
	Equal variances not assumed			.	.	.	-.05278	.	.	.

a. Guideline Code = CG90(H)

[Subsection 4.2.1.2] Recommended psychotherapy and medication combined vs medication (high and low intensity depression combined)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	16	-.2719	.28713	.07178
	Non-CBT	10	-.0690	.47864	.15136

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	1.758	.197	-1.358	24	.187	-.20287	.14945	-.51131	.10556
	Equal variances not assumed			-1.211	13.106	.247	-.20287	.16752	-.56448	.15873

[Subsection 4.2.1.2] Recommended psychotherapy and medication combined vs medication (high intensity depression)

T-Test

Guideline Code = CG90(H)

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	16	-.2719	.28713	.07178
	Non-CBT	7	.0500	.40776	.15412

a. Guideline Code = CG90(H)

Independent Samples Test[a]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	1.910	.181	-2.178	21	.041	-.32187	.14781	-.62927	-.01448
	Equal variances not assumed			-1.893	8.721	.092	-.32187	.17001	-.70836	.06461

a. Guideline Code = CG90(H)

[Subsection 4.2.1.2] Recommended psychotherapy and medication combined vs medication (high intensity depression)

Guideline Code = CG90(L)

Warnings

No statistics are computed for a split file in the Independent Samples table. The split file is: Guideline Code=CG90(L).

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	0[b]	.	.	.
	Non-CBT	3	-.3467	.60575	.34973

a. Guideline Code = CG90(L)

b. t cannot be computed because at least one of the groups is empty.

[Subsection 4.2.1.2] Recommended psychotherapy and medication combined vs psychotherapy comparator and medication (high and low intensity depression combined)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	1	-.3900	.	.
	Non-CBT	1	-.8000	.	.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	.	.	.	0	.	.41000	.	.	.
	Equal variances not assumed			.	.	.	.41000	.	.	.

[Subsection 4.2.1.2] Recommended psychotherapy and medication combined vs psychotherapy comparator and medication combined (high intensity depression)

T-Test

Guideline Code = CG90(H)

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	1	-.3900	.	.
	Non-CBT	1	-.8000	.	.

a. Guideline Code = CG90(H)

Independent Samples Test[a]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Effect Size	Equal variances assumed	.	.	.	0	.	.41000	.	.	.
	Equal variances not assumed			.	.	.	.41000	.	.	.

a. Guideline Code = CG90(H)

[Subsection 4.2.1.2] Recommended psychotherapy vs vs non-active comparator/medication (High and low intensity depression)

T-Test

Warnings

The Independent Samples table is not produced.

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	6	-.3333	.11219	.04580
	Non-CBT	0[a]	.	.	.

a. t cannot be computed because at least one of the groups is empty.

[Subsection 4.2.1.2] Recommended psychotherapy compared to non-active comparator/medication (High intensity depression)

T-Test

Warnings

No statistics are computed for a split file in the Independent Samples table. The split file is: Guideline Code=CG90(H).

Guideline Code = CG90(H)

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Effect Size	CBT	6	-.3333	.11219	.04580
	Non-CBT	0[b]	.	.	.

a. Guideline Code = CG90(H)

b. t cannot be computed because at least one of the groups is empty.

7.4 Appendix 4

4.2.1.3 Kruskal-Wallis test for treatment comparators (combined intensity)

Kruskal-Wallis Test

Ranks

	Comparator Re-coded	N	Mean Rank
Raw Effect Size	vs non-active comparator	81	79.16
	vs active/ psychotherapy comparator	52	133.36
	vs medication	37	131.39
	+ medication vs medication and/ or psychotherapy	47	115.85
	Total	217	

Test Statistics[a,b,c]

	Raw Effect Size
Chi-Square	31.392
df	3
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: Comparator Re-coded

c. Some or all exact significances cannot be computed because there is insufficient memory.

4.2.1.3 Mann Whitney Tests for pairwise treatment comparators (combined intensity)

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Raw Effect Size	217	-.2150	.41014	-1.58	.80
Comparator Re-coded	217	2.23	1.168	1	4

Mann-Whitney Test

Ranks

	Comparator Re-coded	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs non-active comparator	81	53.83	4360.50
	vs active/ psychotherapy comparator	52	87.51	4550.50
	Total	133		

Test Statistics[a]

	Raw Effect Size
Mann-Whitney U	1039.500
Wilcoxon W	4360.500
Z	-4.918
Asymp. Sig. (2-tailed)	.000
Exact Sig. (2-tailed)	.000
Exact Sig. (1-tailed)	.000
Point Probability	.000

a. Grouping Variable: Comparator Re-coded

Mann-Whitney Test

Ranks

	Comparator Re-coded	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs non-active comparator	81	51.15	4143.50
	vs medication	37	77.77	2877.50
	Total	118		

Test Statistics[a]

	Raw Effect Size
Mann-Whitney U	822.500
Wilcoxon W	4143.500
Z	-3.922
Asymp. Sig. (2-tailed)	.000
Exact Sig. (2-tailed)	.000
Exact Sig. (1-tailed)	.000
Point Probability	.000

a. Grouping Variable: Comparator Re-coded

Mann-Whitney Test

Ranks

	Comparator Re-coded	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs non-active comparator	81	56.17	4550.00
	+ medication vs medication and/ or psychotherapy	47	78.85	3706.00
	Total	128		

Test Statistics[a]

	Raw Effect Size
Mann-Whitney U	1229.000
Wilcoxon W	4550.000
Z	-3.335
Asymp. Sig. (2-tailed)	.001
Exact Sig. (2-tailed)	.001
Exact Sig. (1-tailed)	.000
Point Probability	.000

a. Grouping Variable: Comparator Re-coded

Mann-Whitney Test

Ranks

	Comparator Re-coded	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs active/ psychotherapy comparator	52	44.74	2326.50
	vs medication	37	45.36	1678.50
	Total	89		

Test Statistics[a]

	Raw Effect Size
Mann-Whitney U	948.500
Wilcoxon W	2326.500
Z	-.112
Asymp. Sig. (2-tailed)	.910
Exact Sig. (2-tailed)	.913
Exact Sig. (1-tailed)	.456
Point Probability	.002

a. Grouping Variable: Comparator Re-coded

Mann-Whitney Test

Ranks

	Comparator Re-coded	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs active/ psychotherapy comparator	52	54.11	2813.50
	+ medication vs medication and/ or psychotherapy	47	45.46	2136.50
	Total	99		

Test Statistics[a]

	Raw Effect Size
Mann-Whitney U	1008.500
Wilcoxon W	2136.500
Z	-1.497
Asymp. Sig. (2-tailed)	.135
Exact Sig. (2-tailed)	.135
Exact Sig. (1-tailed)	.068
Point Probability	.000

a. Grouping Variable: Comparator Re-coded

Mann-Whitney Test

Ranks

	Comparator Re-coded	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs medication	37	46.26	1711.50
	+ medication vs medication and/ or psychotherapy	47	39.54	1858.50
	Total	84		

Test Statistics[a]

	Raw Effect Size
Mann-Whitney U	730.500
Wilcoxon W	1858.500
Z	-1.253
Asymp. Sig. (2-tailed)	.210
Exact Sig. (2-tailed)	.212
Exact Sig. (1-tailed)	.106
Point Probability	.001

a. Grouping Variable: Comparator Re-coded

7.5 Appendix 5

[Sub-section 4.2.1.4] Kruskal-Wallis and Mann-Whitney tests for treatment comparators (controlling for treatment intensity)

Guideline Code = CG90(H)

Kruskal-Wallis Test

Ranks[a]

	Comparator Re-coded	N	Mean Rank
Raw Effect Size	vs non-active comparator	28	53.00
	vs active/ psychotherapy comparator	42	83.27
	vs medication	34	86.71
	+ medication vs medication and/ or psychotherapy	44	70.38
	Total	148	

a. Guideline Code = CG90(H)

Test Statistics[a,b,c,d]

	Raw Effect Size
Chi-Square	11.971
df	3
Asymp. Sig.	.007

a. Guideline Code = CG90(H)

b. Kruskal Wallis Test

c. Grouping Variable: Comparator Re-coded

d. Some or all exact significances cannot be computed because there is insufficient memory.

Guideline Code = CG90(L)

Kruskal-Wallis Test

Ranks[a]

	Comparator Re-coded	N	Mean Rank
Raw Effect Size	vs non-active comparator	53	32.52
	vs active/ psychotherapy comparator	10	49.60
	vs medication	3	23.67
	+ medication vs medication and/ or psychotherapy	3	41.50
	Total	69	

a. Guideline Code = CG90(L)

Test Statistics[a,b,c,d]

	Raw Effect Size
Chi-Square	7.381
df	3
Asymp. Sig.	.061

a. Guideline Code = CG90(L)

b. Kruskal Wallis Test

c. Grouping Variable: Comparator Re-coded

d. Some or all exact significances cannot be computed because there is insufficient memory.

[sub-section 4.2.1.4)] Mann-Whitney (pair-wise comparisons) (High intensity)

Guideline Code = CG90(H)

Descriptive Statistics[a]

	N	Mean	Std. Deviation	Minimum	Maximum
Raw Effect Size	148	-.1288	.34772	-1.35	.67
Comparator Re-coded	148	2.64	1.101	1	4

a. Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Comparator Re-coded	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs non-active comparator	28	26.91	753.50
	vs active/ psychotherapy comparator	42	41.23	1731.50
	Total	70		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	347.500
Wilcoxon W	753.500
Z	-2.884
Asymp. Sig. (2-tailed)	.004
Exact Sig. (2-tailed)	.004
Exact Sig. (1-tailed)	.002
Point Probability	.000

a. Guideline Code = CG90(H)

b. Grouping Variable: Comparator Re-coded

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Comparator Re-coded	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs non-active comparator	28	24.32	681.00
	vs medication	34	37.41	1272.00
	Total	62		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	275.000
Wilcoxon W	681.000
Z	-2.844
Asymp. Sig. (2-tailed)	.004
Exact Sig. (2-tailed)	.004
Exact Sig. (1-tailed)	.002
Point Probability	.000

a. Guideline Code = CG90(H)

b. Grouping Variable: Comparator Re-coded

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Comparator Re-coded	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs non-active comparator	28	30.77	861.50
	+ medication vs medication and/ or psychotherapy	44	40.15	1766.50
	Total	72		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	455.500
Wilcoxon W	861.500
Z	-1.855
Asymp. Sig. (2-tailed)	.064
Exact Sig. (2-tailed)	.064
Exact Sig. (1-tailed)	.032
Point Probability	.000

a. Guideline Code = CG90(H)

b. Grouping Variable: Comparator Re-coded

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Comparator Re-coded	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs active/ psychotherapy comparator	42	37.44	1572.50
	vs medication	34	39.81	1353.50
	Total	76		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	669.500
Wilcoxon W	1572.500
Z	-.465
Asymp. Sig. (2-tailed)	.642
Exact Sig. (2-tailed)	.646
Exact Sig. (1-tailed)	.323
Point Probability	.002

a. Guideline Code = CG90(H)

b. Grouping Variable: Comparator Re-coded

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Comparator Re-coded	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs active/ psychotherapy comparator	42	47.61	1999.50
	+ medication vs medication and/ or psychotherapy	44	39.58	1741.50
	Total	86		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	751.500
Wilcoxon W	1741.500
Z	-1.491
Asymp. Sig. (2-tailed)	.136
Exact Sig. (2-tailed)	.137
Exact Sig. (1-tailed)	.069
Point Probability	.001

a. Guideline Code = CG90(H)

b. Grouping Variable: Comparator Re-coded

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Comparator Re-coded	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs medication	34	44.49	1512.50
	+ medication vs medication and/ or psychotherapy	44	35.65	1568.50
	Total	78		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	578.500
Wilcoxon W	1568.500
Z	-1.709
Asymp. Sig. (2-tailed)	.088
Exact Sig. (2-tailed)	.088
Exact Sig. (1-tailed)	.044
Point Probability	.000

a. Guideline Code = CG90(H)

b. Grouping Variable: Comparator Re-coded

7.6 Appendix 6

[Subsection 4.1.2.5] Kruskal-Wallis test for treatment comparators controlling for recommended psychotherapy group (High intensity)

Recommended Psychotherapy = CBT

Descriptive Statistics[a]

	N	Mean	Std. Deviation	Minimum	Maximum
Raw Effect Size	99	-.1109	.28120	-.90	.67
Comparator Code Re-grouped	99	3.04	1.584	0	7

a. Recommended Psychotherapy = CBT

Kruskal-Wallis Test

Ranks[a]

	Comparator Code Re-grouped	N	Mean Rank
Raw Effect Size	vs non-active comparator	10	44.80
	vs active/ psychotherapy	25	57.50
	vs medication	23	45.43
	+ medication vs medication and/ or psychotherapy	35	41.16
	Total	93	

a. Recommended Psychotherapy = CBT

Test Statistics[a,b,c,d]

	Raw Effect Size
Chi-Square	5.571
df	3
Asymp. Sig.	.134

a. Recommended Psychotherapy = CBT

b. Kruskal Wallis Test

c. Grouping Variable: Comparator Code Re-grouped

d. Some or all exact significances cannot be computed because there is insufficient memory.

Recommended Psychotherapy = Non-CBT

Descriptive Statistics[a]

	N	Mean	Std. Deviation	Minimum	Maximum
Raw Effect Size	55	-.1833	.43341	-1.35	.59
Comparator Code Re-grouped	55	1.87	1.479	0	4

a. Recommended Psychotherapy = Non-CBT

Kruskal-Wallis Test

Ranks[a]

	Comparator Code Re-grouped	N	Mean Rank
Raw Effect Size	vs non-active comparator	18	17.67
	vs active/ psychotherapy	17	27.32
	vs medication	11	42.41
	+ medication vs medication and/ or psychotherapy	9	32.33
	Total	55	

a. Recommended Psychotherapy = Non-CBT

Test Statistics[a,b,c,d]

	Raw Effect Size
Chi-Square	17.086
df	3
Asymp. Sig.	.001

a. Recommended Psychotherapy = Non-CBT

b. Kruskal Wallis Test

c. Grouping Variable: Comparator Code Re-grouped

d. Some or all exact significances cannot be computed because there is insufficient memory.

[sub-section 4.2.1.5] Kruskal-Wallis test for treatment comparators controlling for psychotherapy (low intensity)

Recommended Psychotherapy = CBT

Descriptive Statistics[a]

	N	Mean	Std. Deviation	Minimum	Maximum
Raw Effect Size	31	-.3310	.35630	-1.54	.39
Comparator Code Re-grouped	31	.26	.682	0	2

a. Recommended Psychotherapy = CBT

Kruskal-Wallis Test

Ranks[a]

	Comparator Code Re-grouped	N	Mean Rank
Raw Effect Size	vs non-active comparator	27	14.30
	vs active/ psychotherapy	4	27.50
	Total	31	

a. Recommended Psychotherapy = CBT

Test Statistics[a,b,c]

	Raw Effect Size
Chi-Square	7.353
df	1
Asymp. Sig.	.007
Exact Sig.	.003
Point Probability	.001

a. Recommended Psychotherapy = CBT

b. Kruskal Wallis Test

c. Grouping Variable: Comparator Code Re-grouped

Recommended Psychotherapy = Non-CBT

Descriptive Statistics[a]

	N	Mean	Std. Deviation	Minimum	Maximum
Raw Effect Size	38	-.4561	.54595	-1.58	.80
Comparator Code Re-grouped	38	.87	1.379	0	4

a. Recommended Psychotherapy = Non-CBT

Kruskal-Wallis Test

Ranks[a]

	Comparator Code Re-grouped	N	Mean Rank
Raw Effect Size	vs non-active comparator	26	18.19
	vs active/ psychotherapy	6	25.33
	vs medication	3	15.33
	+ medication vs medication and/ or psychotherapy	3	23.33
	Total	38	

a. Recommended Psychotherapy = Non-CBT

Test Statistics[a,b,c]

	Raw Effect Size
Chi-Square	2.794
df	3
Asymp. Sig.	.425
Exact Sig.	.[d]
Point Probability	.

a. Recommended Psychotherapy = Non-CBT

b. Kruskal Wallis Test

c. Grouping Variable: Comparator Code Re-grouped

d. Numerical difficulties prevented calculation.

[4.2.1.5] Mann-Whitey Tests for pairwise comparisons of treatment comparators in the other psychotherapies group (High intensity)

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Comparator Code Re-grouped	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs non-active comparator	10	14.90	149.00
	vs active/ psychotherapy	25	19.24	481.00
	Total	35		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	94.000
Wilcoxon W	149.000
Z	-1.133
Asymp. Sig. (2-tailed)	.257
Exact Sig. [2*(1-tailed Sig.)]	.270[c]
Exact Sig. (2-tailed)	.266
Exact Sig. (1-tailed)	.133
Point Probability	.004

a. Guideline Code = CG90(H)

b. Grouping Variable: Comparator Code Re-grouped

c. Not corrected for ties.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Comparator Code Re-grouped	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs non-active comparator	18	14.22	256.00
	vs active/ psychotherapy	17	22.00	374.00
	Total	35		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	85.000
Wilcoxon W	256.000
Z	-2.245
Asymp. Sig. (2-tailed)	.025
Exact Sig. [2*(1-tailed Sig.)]	.025[c]
Exact Sig. (2-tailed)	.024
Exact Sig. (1-tailed)	.012
Point Probability	.001

a. Guideline Code = CG90(H)

b. Grouping Variable: Comparator Code Re-grouped

c. Not corrected for ties.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Comparator Code Re-grouped	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs non-active comparator	18	10.72	193.00
	vs medication	11	22.00	242.00
	Total	29		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	22.000
Wilcoxon W	193.000
Z	-3.463
Asymp. Sig. (2-tailed)	.001
Exact Sig. [2*(1-tailed Sig.)]	.000[c]
Exact Sig. (2-tailed)	.000
Exact Sig. (1-tailed)	.000
Point Probability	.000

a. Guideline Code = CG90(H)

b. Grouping Variable: Comparator Code Re-grouped

c. Not corrected for ties.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Comparator Code Re-grouped	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs non-active comparator	18	11.72	211.00
	+ medication vs medication and/ or psychotherapy	9	18.56	167.00
	Total	27		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	40.000
Wilcoxon W	211.000
Z	-2.109
Asymp. Sig. (2-tailed)	.035
Exact Sig. [2*(1-tailed Sig.)]	.035[c]
Exact Sig. (2-tailed)	.034
Exact Sig. (1-tailed)	.017
Point Probability	.001

a. Guideline Code = CG90(H)

b. Grouping Variable: Comparator Code Re-grouped

c. Not corrected for ties.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Comparator Code Re-grouped	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs active/ psychotherapy	17	10.62	180.50
	vs medication	11	20.50	225.50
	Total	28		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	27.500
Wilcoxon W	180.500
Z	-3.107
Asymp. Sig. (2-tailed)	.002
Exact Sig. [2*(1-tailed Sig.)]	.001[c]
Exact Sig. (2-tailed)	.001
Exact Sig. (1-tailed)	.001
Point Probability	.000

a. Guideline Code = CG90(H)

b. Grouping Variable: Comparator Code Re-grouped

c. Not corrected for ties.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Comparator Code Re-grouped	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs active/ psychotherapy	17	12.71	216.00
	+ medication vs medication and/ or psychotherapy	9	15.00	135.00
	Total	26		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	63.000
Wilcoxon W	216.000
Z	-.728
Asymp. Sig. (2-tailed)	.467
Exact Sig. [2*(1-tailed Sig.)]	.491[c]
Exact Sig. (2-tailed)	.482
Exact Sig. (1-tailed)	.241
Point Probability	.008

a. Guideline Code = CG90(H)

b. Grouping Variable: Comparator Code Re-grouped

c. Not corrected for ties.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Comparator Code Re-grouped	N	Mean Rank	Sum of Ranks
Raw Effect Size	vs medication	11	11.91	131.00
	+ medication vs medication and/ or psychotherapy	9	8.78	79.00
	Total	20		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	34.000
Wilcoxon W	79.000
Z	-1.180
Asymp. Sig. (2-tailed)	.238
Exact Sig. [2*(1-tailed Sig.)]	.261[c]
Exact Sig. (2-tailed)	.252
Exact Sig. (1-tailed)	.126
Point Probability	.007

a. Guideline Code = CG90(H)

b. Grouping Variable: Comparator Code Re-grouped

c. Not corrected for ties.

7.7 Appendix 7

[Subsection 4.2.1.6] Kruskal-Wallis (Actual psychotherapies) (High intensity)

Guideline Code = CG90(H)

Descriptive Statistics[a]

	N	Mean	Std. Deviation	Minimum	Maximum
Raw Effect Size	154	-.1368	.34374	-1.35	.67
Actual Therapy	154	1.14	1.708	0	5

a. Guideline Code = CG90(H)

Kruskal-Wallis Test

Ranks[a]

	Actual Therapy	N	Mean Rank
Raw Effect Size	Behavioural Activation	6	35.67
	Couples therapy	10	24.25
	Interpersonal	19	19.55
	Couselling	8	37.38
	Short-term psychodynamic	12	34.42
	Total	55	

a. Guideline Code = CG90(H)

Test Statistics[a,b,c,d]

	Raw Effect Size
Chi-Square	11.876
df	4
Asymp. Sig.	.018

a. Guideline Code = CG90(H)

b. Kruskal Wallis Test

c. Grouping Variable: Actual Therapy

d. Some or all exact significances cannot be computed because there is insufficient memory.

[subsection 4.2.1.6] Kruskal-Wallis (Actual psychotherapies) (Low intensity)

Guideline Code = CG90(L)

Descriptive Statistics[a]

	N	Mean	Std. Deviation	Minimum	Maximum
Raw Effect Size	69	-.3999	.47129	-1.58	.80
Actual Therapy	69	3.30	3.006	0	6

a. Guideline Code = CG90(L)

Kruskal-Wallis Test

Ranks[a]

	Actual Therapy	N	Mean Rank
Raw Effect Size	Physical activity	38	19.50
	Total	38[b]	

a. Guideline Code = CG90(L)

b. There is only one non-empty group. Kruskal-Wallis Test cannot be performed.

[From section number 4.2.1.6] Mann-Whitney pairwise comparisons (Actual psychotherapies) (High intensity)

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Actual Therapy	N	Mean Rank	Sum of Ranks
Raw Effect Size	Behavioural Activation	6	11.25	67.50
	Couples therapy	10	6.85	68.50
	Total	16		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	13.500
Wilcoxon W	68.500
Z	-1.791
Asymp. Sig. (2-tailed)	.073
Exact Sig. [2*(1-tailed Sig.)]	.073[c]
Exact Sig. (2-tailed)	.077
Exact Sig. (1-tailed)	.039
Point Probability	.005

a. Guideline Code = CG90(H)

b. Grouping Variable: Actual Therapy

c. Not corrected for ties.

Warnings

There are not enough valid cases to perform the Mann-Whitney Test for Raw Effect Size * Actual Therapy (Behavioural Activation, Interpersonal) in split file Guideline Code=CG90(L). No statistics are computed.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Actual Therapy	N	Mean Rank	Sum of Ranks
Raw Effect Size	Behavioural Activation	6	19.08	114.50
	Interpersonal	19	11.08	210.50
	Total	25		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	20.500
Wilcoxon W	210.500
Z	-2.324
Asymp. Sig. (2-tailed)	.020
Exact Sig. [2*(1-tailed Sig.)]	.017[c]
Exact Sig. (2-tailed)	.018
Exact Sig. (1-tailed)	.009
Point Probability	.001

a. Guideline Code = CG90(H)

b. Grouping Variable: Actual Therapy

c. Not corrected for ties.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Actual Therapy	N	Mean Rank	Sum of Ranks
Raw Effect Size	Behavioural Activation	6	6.83	41.00
	Counselling	8	8.00	64.00
	Total	14		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	20.000
Wilcoxon W	41.000
Z	-.517
Asymp. Sig. (2-tailed)	.605
Exact Sig. [2*(1-tailed Sig.)]	.662[c]
Exact Sig. (2-tailed)	.637
Exact Sig. (1-tailed)	.318
Point Probability	.020

a. Guideline Code = CG90(H)

b. Grouping Variable: Actual Therapy

c. Not corrected for ties.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Actual Therapy	N	Mean Rank	Sum of Ranks
Raw Effect Size	Behavioural Activation	6	9.00	54.00
	Short-term psychodynamic	12	9.75	117.00
	Total	18		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	33.000
Wilcoxon W	54.000
Z	-.281
Asymp. Sig. (2-tailed)	.779
Exact Sig. [2*(1-tailed Sig.)]	.820[c]
Exact Sig. (2-tailed)	.820
Exact Sig. (1-tailed)	.410
Point Probability	.035

a. Guideline Code = CG90(H)

b. Grouping Variable: Actual Therapy

c. Not corrected for ties.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Actual Therapy	N	Mean Rank	Sum of Ranks
Raw Effect Size	Couples therapy	10	17.40	174.00
	Interpersonal	19	13.74	261.00
	Total	29		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	71.000
Wilcoxon W	261.000
Z	-1.102
Asymp. Sig. (2-tailed)	.271
Exact Sig. [2*(1-tailed Sig.)]	.286[c]
Exact Sig. (2-tailed)	.281
Exact Sig. (1-tailed)	.141
Point Probability	.005

a. Guideline Code = CG90(H)

b. Grouping Variable: Actual Therapy

c. Not corrected for ties.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Actual Therapy	N	Mean Rank	Sum of Ranks
Raw Effect Size	Couples therapy	10	7.40	74.00
	Couselling	8	12.13	97.00
	Total	18		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	19.000
Wilcoxon W	74.000
Z	-1.867
Asymp. Sig. (2-tailed)	.062
Exact Sig. [2*(1-tailed Sig.)]	.068[c]
Exact Sig. (2-tailed)	.064
Exact Sig. (1-tailed)	.032
Point Probability	.003

a. Guideline Code = CG90(H)

b. Grouping Variable: Actual Therapy

c. Not corrected for ties.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Actual Therapy	N	Mean Rank	Sum of Ranks
Raw Effect Size	Couples therapy	10	9.10	91.00
	Short-term psychodynamic	12	13.50	162.00
	Total	22		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	36.000
Wilcoxon W	91.000
Z	-1.583
Asymp. Sig. (2-tailed)	.114
Exact Sig. [2*(1-tailed Sig.)]	.123[c]
Exact Sig. (2-tailed)	.123
Exact Sig. (1-tailed)	.061
Point Probability	.008

a. Guideline Code = CG90(H)

b. Grouping Variable: Actual Therapy

c. Not corrected for ties.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Actual Therapy	N	Mean Rank	Sum of Ranks
Raw Effect Size	Interpersonal	19	11.37	216.00
	Couselling	8	20.25	162.00
	Total	27		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	26.000
Wilcoxon W	216.000
Z	-2.657
Asymp. Sig. (2-tailed)	.008
Exact Sig. [2*(1-tailed Sig.)]	.007[c]
Exact Sig. (2-tailed)	.006
Exact Sig. (1-tailed)	.003
Point Probability	.000

a. Guideline Code = CG90(H)

b. Grouping Variable: Actual Therapy

c. Not corrected for ties.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Actual Therapy	N	Mean Rank	Sum of Ranks
Raw Effect Size	Interpersonal	19	13.37	254.00
	Short-term psychodynamic	12	20.17	242.00
	Total	31		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	64.000
Wilcoxon W	254.000
Z	-2.029
Asymp. Sig. (2-tailed)	.043
Exact Sig. [2*(1-tailed Sig.)]	.043[c]
Exact Sig. (2-tailed)	.042
Exact Sig. (1-tailed)	.021
Point Probability	.001

a. Guideline Code = CG90(H)

b. Grouping Variable: Actual Therapy

c. Not corrected for ties.

Guideline Code = CG90(H)

Mann-Whitney Test

Ranks[a]

	Actual Therapy	N	Mean Rank	Sum of Ranks
Raw Effect Size	Couselling	8	10.50	84.00
	Short-term psychodynamic	12	10.50	126.00
	Total	20		

a. Guideline Code = CG90(H)

Test Statistics[a,b]

	Raw Effect Size
Mann-Whitney U	48.000
Wilcoxon W	126.000
Z	.000
Asymp. Sig. (2-tailed)	1.000
Exact Sig. [2*(1-tailed Sig.)]	1.000[c]
Exact Sig. (2-tailed)	1.000
Exact Sig. (1-tailed)	.515
Point Probability	.030

a. Guideline Code = CG90(H)

b. Grouping Variable: Actual Therapy

c. Not corrected for ties.

7.8 Appendix 8

[Subsection 4.5] Overall test of difference between the CBT and other psychotherapies group for dichotomous outcomes (treatment intensity combined)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Risk Ratio	CBT	108	1.3396	2.18876	.21061
	Non-CBT	91	1.2119	1.51550	.15887

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Risk Ratio	Equal variances assumed	.049	.825	.470	197	.639	.12776	.27191	-.40846	.66398
	Equal variances not assumed			.484	190.197	.629	.12776	.26381	-.39261	.64814

[Subsection 4.5] Overall test of difference between the CBT and other psychotherapies group for dichotomous outcomes (high intensity)

T-Test

Guideline Code = CG90(H)

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Risk Ratio	CBT	97	1.1721	1.99733	.20280
	Non-CBT	79	1.1586	1.54721	.17407

a. Guideline Code = CG90(H)

Independent Samples Test[a]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Risk Ratio	Equal variances assumed	.250	.618	.049	174	.961	.01345	.27422	-.52778	.55469
	Equal variances not assumed			.050	173.594	.960	.01345	.26726	-.51405	.54096

a. Guideline Code = CG90(H)

[Subsection 4.5] Overall test of difference between the CBT and other psychotherapies group for dichotomous outcome (low intensity)

Guideline Code = CG90(L)

Group Statistics[a]

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Risk Ratio	CBT	11	2.8173	3.20759	.96712
	Non-CBT	12	1.5625	1.28933	.37220

a. Guideline Code = CG90(L)

Independent Samples Test[a]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Risk Ratio	Equal variances assumed	4.151	.054	1.251	21	.225	1.25477	1.00270	-.83045	3.33999
	Equal variances not assumed			1.211	12.924	.248	1.25477	1.03627	-.98530	3.49484

a. Guideline Code = CG90(L)

7.9 Appendix 9

[Subsection 4.5] Tests of difference between recommended psychotherapy groups based on low quality outcome rating for dichotomous outcomes (combined treatment intensity)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Risk Ratio	CBT	68	1.0906	1.14431	.13877
	Non-CBT	54	1.3576	1.84987	.25174

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Risk Ratio	Equal variances assumed	3.933	.050	-.978	120	.330	-.26700	.27296	-.80744	.27343
	Equal variances not assumed			-.929	83.971	.356	-.26700	.28745	-.83863	.30462

[Subsection 4.5] Tests of difference between recommended psychotherapy groups for dichotomous outcomes based on low quality outcome rating (controlling for treatment intensity)

T-Test

Group Statistics

Guideline Code		Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
CG90(H)	Raw Risk Ratio	CBT	63	1.0073	.88811	.11189
		Non-CBT	48	1.2521	1.87307	.27036
CG90(L)	Raw Risk Ratio	CBT	5	2.1400	2.86743	1.28235
		Non-CBT	6	2.2017	1.52726	.62350

Independent Samples Test

Guideline Code			Levene's Test for Equality of Variances		t-test for Equality of Means						
										95% Confidence Interval of the Difference	
			F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
CG90(H)	Raw Risk Ratio	Equal variances assumed	4.269	.041	-.912	109	.364	-.24478	.26832	-.77659	.28703
		Equal variances not assumed			-.837	63.077	.406	-.24478	.29259	-.82947	.33991
CG90(L)	Raw Risk Ratio	Equal variances assumed	1.277	.288	-.046	9	.964	-.06167	1.34724	-3.10933	2.98600
		Equal variances not assumed			-.043	5.853	.967	-.06167	1.42590	-3.57205	3.44872

[Subsection 4.5] Tests of difference between recommended psychotherapy groups for dichotomous outcomes based on moderate quality outcome rating (combined intensity)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Risk Ratio	CBT	29	1.4821	1.87720	.34859
	Non-CBT	35	1.0220	.80626	.13628

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Risk Ratio	Equal variances assumed	.860	.357	1.313	62	.194	.46007	.35046	-.24049	1.16063
	Equal variances not assumed			1.229	36.511	.227	.46007	.37428	-.29864	1.21878

[Subsection 4.5] Tests of difference between recommended psychotherapy groups for dichotomous outcomes based on moderate quality outcome rating (controlling for treatment intensity)

T-Test

Group Statistics

Guideline Code		Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
CG90(H)	Raw Risk Ratio	CBT	24	1.0383	.55398	.11308
		Non-CBT	29	1.0424	.85169	.15815
CG90(L)	Raw Risk Ratio	CBT	5	3.6120	4.00629	1.79167
		Non-CBT	6	.9233	.58667	.23951

Independent Samples Test

Guideline Code			Levene's Test for Equality of Variances		t-test for Equality of Means						
										95% Confidence Interval of the Difference	
			F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
CG90(H)	Raw Risk Ratio	Equal variances assumed	3.815	.056	-.020	51	.984	-.00408	.20215	-.40992	.40176
		Equal variances not assumed			-.021	48.512	.983	-.00408	.19442	-.39489	.38672
CG90(L)	Raw Risk Ratio	Equal variances assumed	5.816	.039	1.641	9	.135	2.68867	1.63882	-1.01859	6.39593
		Equal variances not assumed			1.487	4.143	.209	2.68867	1.80760	-2.26238	7.63971

[Subsection 4.5] Tests of difference between recommended psychotherapy groups for dichotomous outcomes based on high quality outcome rating (combined intensity)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Risk Ratio	CBT	11	2.5036	5.53616	1.66921
	Non-CBT	2	.6000	.28284	.20000

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Risk Ratio	Equal variances assumed	.718	.415	.469	11	.648	1.90364	4.05816	-7.02831	10.83559
	Equal variances not assumed			1.132	10.268	.283	1.90364	1.68115	-1.82900	5.63627

[Subsection 4.5] Tests of difference between recommended psychotherapy groups for dichotomous outcomes based on high quality outcome rating (controlling for treatment intensity)

T-Test

Warnings

No statistics are computed for a split file in the Independent Samples table. The split file is: Guideline Code=CG90(L).

Group Statistics

Guideline Code		Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
CG90(H)	Raw Risk Ratio	CBT	10	2.5310	5.83484	1.84514
		Non-CBT	2	.6000	.28284	.20000
CG90(L)	Raw Risk Ratio	CBT	1	2.2300	.	.
		Non-CBT	0[a]	.	.	.

a. t cannot be computed because at least one of the groups is empty.

Independent Samples Test[a]

			Levene's Test for Equality of Variances		t-test for Equality of Means						
										95% Confidence Interval of the Difference	
Guideline Code			F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
CG90(H)	Raw Risk Ratio	Equal variances assumed	.824	.385	.450	10	.662	1.93100	4.28827	-7.62387	11.48587
		Equal variances not assumed			1.040	9.201	.325	1.93100	1.85595	-2.25348	6.11548

a. No statistics are computed for one or more split files

7.10 Appendix 10

[Subsection 4.5] Tests of difference between recommended psychotherapy groups for dichotomous outcomes based on reduced risk of negative outcome (combined treatment intensity)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Risk Ratio	CBT	64	.6494	.20104	.02513
	Non-CBT	59	.5854	.23972	.03121

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Risk Ratio	Equal variances assumed	2.714	.102	1.607	121	.111	.06395	.03978	-.01481	.14271
	Equal variances not assumed			1.596	113.619	.113	.06395	.04007	-.01543	.14333

[Subsection 4.5] Tests of difference between recommended psychotherapy groups for dichotomous outcomes based on reduced risk of negative outcome (controlling for treatment intensity)

T-Test

Group Statistics

Guideline Code		Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
CG90(H)	Raw Risk Ratio	CBT	61	.6467	.20308	.02600
		Non-CBT	55	.5896	.23825	.03213
CG90(L)	Raw Risk Ratio	CBT	3	.7033	.17673	.10203
		Non-CBT	4	.5275	.29067	.14534

Independent Samples Test

			Levene's Test for Equality of Variances		t-test for Equality of Means						
										95% Confidence Interval of the Difference	
Guideline Code			F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
CG90(H)	Raw Risk Ratio	Equal variances assumed	2.035	.156	1.393	114	.166	.05708	.04099	-.02412	.13829
		Equal variances not assumed			1.381	106.710	.170	.05708	.04133	-.02485	.13902
CG90(L)	Raw Risk Ratio	Equal variances assumed	1.347	.298	.916	5	.402	.17583	.19199	-.31769	.66936
		Equal variances not assumed			.990	4.900	.368	.17583	.17758	-.28345	.63512

[Subsection 4.5] Tests of difference between recommended psychotherapy groups for dichotomous outcomes based on increased risk of negative outcome (combined treatment intensity)

T-Test

Group Statistics

	Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
Raw Risk Ratio	CBT	44	2.3436	3.18138	.47961
	Non-CBT	32	2.3669	2.10629	.37234

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw Risk Ratio	Equal variances assumed	.248	.620	-.036	74	.971	-.02324	.64635	-1.31112	1.26464
	Equal variances not assumed			-.038	73.446	.970	-.02324	.60718	-1.23322	1.18674

[Subsection 4.5] Tests of difference between recommended psychotherapy groups for dichotomous outcomes based on increased risk of negative outcome (controlling for treatment intensity)

T-Test

Group Statistics

Guideline Code		Recommended Psychotherapy	N	Mean	Std. Deviation	Std. Error Mean
CG90(H)	Raw Risk Ratio	CBT	36	2.0622	3.09442	.51574
		Non-CBT	24	2.4625	2.33170	.47596
CG90(L)	Raw Risk Ratio	CBT	8	3.6100	3.47214	1.22759
		Non-CBT	8	2.0800	1.28759	.45523

Independent Samples Test

			Levene's Test for Equality of Variances		t-test for Equality of Means						
										95% Confidence Interval of the Difference	
Guideline Code			F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
CG90(H)	Raw Risk Ratio	Equal variances assumed	.097	.757	-.539	58	.592	-.40028	.74229	-1.88613	1.08557
		Equal variances not assumed			-.570	57.042	.571	-.40028	.70180	-1.80558	1.00502
CG90(L)	Raw Risk Ratio	Equal variances assumed	5.771	.031	1.169	14	.262	1.53000	1.30928	-1.27812	4.33812
		Equal variances not assumed			1.169	8.890	.273	1.53000	1.30928	-1.43741	4.49741

7.11 Appendix 11

Subsection 4.7.2 Proportion of relative risk ratios for recommended psychotherapy compared to active interventions (low intensity depression)

Section C. Low Intensity

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	1	1	2
		% within Recommended Psychotherapy	50.0%	50.0%	100.0%
	Other psychotherapies	Count	0	1	1
		% within Recommended Psychotherapy	0.0%	100.0%	100.0%
Total		Count	1	2	3
		% within Recommended Psychotherapy	33.3%	66.7%	100.0%

Fischer Exact, p = 1.00 (Non-significant)

Table 4.7.6 Proportion of relative risk ratios for psychotherapies combined with medication compared to active or non-active comparators combined with medication (High intensity depression)

			risk likelihood		
			less	more	Total
Recommended Psychotherapy	CBT	Count	0	1	1
		% within Recommended Psychotherapy	0.0%	100.0%	100.0%
	Other psychotherapies	Count	3	2	5
		% within Recommended Psychotherapy	60.0%	40.0%	100.0%
Total		Count	3	3	6
		% within Recommended Psychotherapy	50.0%	50.0%	100.0%

Fischer Exact, p = 1.00 (Non-significant)

8. Glossary of terms used within this project

Below is a selected list of terms relevant to this project. This is not intended to provide an exhaustive or definitive list of definitions; rather it is aimed at providing clarity of technical and idiosyncratic terminology used within the review.

Actual psychotherapies – refers to the specific classes of psychotherapies recommended by NICE that were categorised within the current project into the 'other psychotherapies' group (see also other psychotherapies).

Association analysis – within this project this term refers to examining the relationship between the variables recommended psychotherapy (i.e. CBT and other psychotherapies) and outcome at a categorical data level (see also correlational analysis).

Bias – in research, this term refers to occurrence of systematic error that is introduced into sampling or testing by selecting or encouraging one outcome or answer over others. There are numerous types of research bias that can occur at various phases of the research process, including study design or data collection, as well as in the process of data analysis and publication. Within this project the topic of bias is central to the exploration of the NICE evidence base, which occurs at both a primary and secondary level of data abstraction. Bias is equally relevant to this project's process of reviewing NICE's evidence and the interpretation of its findings.

Bona fide psychotherapies – within this project this term is used as previously defined by Wampold (2001) as treatments that are intended to be therapeutic and are based on a clear rationale for a particular problem. These are in contrast to minimal and 'intent-to-fail' psychological interventions that are sometimes implemented to control for common factors.

Categorical data – refers to data or variables that are separated into different categories that are mutually exclusive from one another. Within this project categorical data mainly consisted of effect size magnitudes (i.e. 'small' or 'medium/ large'), overall effectiveness (i.e. 'none' or 'any magnitude') and risk likelihood (i.e. 'less' or 'more').

CBT-based interventions – within this project this refers to treatments that were reviewed together within the depression guideline on the basis of containing both cognitive and behavioural components of treatment.

Clinical Practice Guidelines – refers to 'systematic statements to guide decisions about appropriate health care for specific and clinical purposes' (Field and Lohr, 1990, p. 38)

Clinical significance – this term refers to the practical importance of a treatment effect and whether it is likely to make a meaningful difference to a patient's life. Within guideline development this is usually determined by the magnitude of treatment effect.

Cohen's d effect size – refers to the difference between two means divided by a standard deviation for the data (see also definition of effect size).

Contingency tables - a table showing the distribution of one variable in rows and another in columns, used to study the correlation between the two variables. Within this project this term and cross tabulation are used interchangeably (see also definition for cross tabulation).

Continuous outcome – this term refers to evidence based upon measurement of change on a numerical scale, the results usually being summarised as the mean differences between the two psychotherapy groups being compared.

Controlling for variables – in statistics this term refers to controlling for a variable in an attempt to reduce the effect of confounding variables on an observational study. Controlling for a variable involves holding a specific variable (e.g. quality rating, treatment comparator etc.) constant for calculations made about the effect of the independent variable (i.e. recommended psychotherapy group) on the dependent variable (e.g. effect size magnitude).

Correlational analysis - is a method of statistical evaluation used to study the strength of a relationship between two variables (e.g. psychotherapy and outcome). This term was used interchangeably with the terms association analysis (as defined earlier) and cross tabulation analysis (see definitions below).

Cross tabulation analysis – this refers to a statistical tool that is used to analyse categorical data. Cross tabulation helps to understand how two different variables are related to each other.

Dichotomous outcome – this term refers to outcomes in which there are only two possibilities (e.g. achieved remission or did not achieve remission). Dichotomous outcomes are categorical variables usually based upon cut-off scores from data on continuous outcomes.

Descriptive statistics – this term refers to description of basic features of the data in the study. They provide simple summaries about the sample and the measures.

Effectiveness – within this project this term refers to the effect size values of 0.2 and above, which is indicative of a small effect. Within this project this term is used interchangeably with overall effectiveness. The use of this term here is not to be confused with 'effectiveness' in the general use of psychotherapy research, which refers to how an interventions works in clinical populations.

Effect size – this refers to a calculation for the size of the difference between two groups. Within this project this refers to the routine use of effect sizes as a technique in meta-analysis for combining and comparing estimates from different studies. (See also earlier definition of Cohen's d effect size)

Effect size magnitudes – within this project this refers to the scale of effect sizes within the guideline evidence base i.e. 'small' for values of at least 0.2 and below 0.5, 'medium' for values of at least 0.5 and below 0.8 and 'large' for values of 0.8 and above.

Evidence-based practice – this is a generic term for core principles of evidence-based medicine within wider disciplines of healthcare including psychological interventions (Trinder & Reynolds, 2000).

Exploratory data analysis – refers to an approach to analysing data sets to summarise their main characteristics. This is primarily used for seeing what the data can tell us beyond the formal tasks of statistical modelling or hypothesis testing.

Guideline characteristics/ variables – within this project this term refers to specific features of the meta-analyses conducted within NICE's evidence review. Within this project these were also referred to as meta-analytic and review characteristics.

High-intensity depression – this term refers to the use of psychosocial interventions for depressions diagnosed as moderate to severe.

Low-intensity depression – this term refers to the use of psychosocial interventions for depressions diagnosed as mild to moderate.

Meta-analytic characteristics – see guideline characteristics/ variables

None effect – within this project this term was used to describe effect size values that were below 0.2. Such values could be indicative of either equivalence or inferiority of the recommended treatment depending on their specific value however such distinctions were not made within the current project.

Other psychotherapies – within this project this term refers to psychological interventions that were reviewed by NICE separately from their review of evidence for cognitive and behavioural therapies.

Overall effectiveness – see effectiveness

Proportions of effect size/ risk associations – this term refers to descriptive data at a categorical level where 2 x 2 contingency tables were used to express the percentage of effect size magnitudes, overall effectiveness, or risk likelihood.

Psychological interventions – see psychotherapy

Psychosocial interventions – refers to interventions that emphasize psychological or social factors rather than biological factors (Ruddy and House, 2005). This can include health education, as well as interventions that focus on social aspects, such as social support, and interventions that include a physiological component, such as physical exercise. Interventions with organization of care as the main focus were not considered for this review. See also psychotherapy.

Psychotherapy – the use of psychological theory and methods to help a person change or overcome problems in desired ways. Psychotherapy aims to improve well-being and mental health, to resolve or mitigate troublesome behaviours, beliefs, compulsions, thoughts, or emotions, and to improve relationships and social skills. Psychotherapy usually involves interaction between therapist and client and can be delivered in group or individual formats. Certain psychotherapies are considered evidence-based for treating some diagnosed mental disorders. Within this project the term psychotherapy is used interchangeably with psychosocial interventions and psychological interventions.

Quality assessment outcome – this term refers to the rating given to the quality of evidence through the use of the GRADE system adopted by NICE (i.e. high, medium, low). The GRADE system incorporates both primary study factors (e.g. randomization, trial treatment protocols, etc.) and factors relating to the secondary data analysis (e.g. statistical power, statistical significance, etc.).

Recommended psychotherapy – within this project this general term was used when collectively referring to all the psychological interventions recommended by NICE (i.e. both CBT and other psychotherapies group).

Relative effectiveness – within this project this term refers to the evidence of effectiveness when the recommended psychotherapies were compared against various treatment comparators (e.g. active interventions, medication). This was relevant to data at both raw and categorical level.

Review characteristics – see guideline characteristics/ variables

Risk associations – within this project this term refers to the rate (or proportion) of risk of a negative outcome in the treatment group relative to that of the control (or treatment comparison) group. This term is used interchangeably with risk ratios.

Risk likelihood – this term refers to the direction of the risk ratio value as to whether it suggests decreased (i.e. less) or increased (i.e. more) risk of a depression related outcome.

Relative risk ratios – see risk associations

Standardised mean difference – this term refers to a summary statistic used in meta-analysis when all the studies assess the same outcome but measure it in a variety of ways (.e.g. different psychometric scales for depression). Within this project this term is used interchangeably with effect size and Cohen's d.

The guideline – within this project this term is often used when referring to the full version of the NICE guidelines for depression in the treatment of depression in adults. When other guidelines are being referred to this is clearly stated within the report.

Treatment intensity – within this project this term refers collectively to low and high intensity interventions. This term represents an umbrella term for high and low intensity treatments, which was a variable that was frequently controlled as part of the exploratory data analysis.