Psychotherapy for depression works! But how? : Investigating the effects and mechanisms of cognitive therapy vs. interpersonal psychotherapy for major depressive disorder

Document status and date:
Published: 01/01/2015

DOI:
10.26481/dis.20150918ll

Document Version:
Publisher's PDF, also known as Version of record

Psychotherapy for Depression Works! But How?

Investigating the Effects and Mechanisms of Cognitive Therapy

vs. Interpersonal Psychotherapy for Major Depressive Disorder

Psychotherapy for Depression Works! But How?

Investigating the Effects and Mechanisms of Cognitive Therapy

vs. Interpersonal Psychotherapy for Major Depressive Disorder

Proefschrift

door

Lotte Hubertina Josephina Mathias Lemmens

Table of contents

Preface

Je wilt wel opstaan, maar het gaat niet
Je wilt gewoon naar buiten, maar het lukt niet
De deur uit, de straat op, maar je durft niet
Je zou me bellen, maar je belt niet

Ik vraag waarom, je zegt: 'het ging niet'
'Kom op, we gaan wat drinken', maar je wilt niet
Je wilt alleen maar slapen, want je slaapt niet

De luiken dicht
Met je hoofd onder het kussen

.....

Gooi die luiken open
Zwaai met je haren naar de wind

Wat is er gebeurd?
Je zegt: *'ik weet het niet'*

Maarten van Roozendaal – Luiken

1

General Introduction

This dissertation concerns the effectiveness and mechanisms of change of Cognitive Therapy (CT) and Interpersonal Psychotherapy (IPT) for Major Depressive Disorder (MDD). It contains results of a large randomized controlled trial (RCT) that was conducted in an effort to gain more insight into the potential differential effects between these two treatments and their underlying mechanisms. In this first chapter, a general introduction to the contents of this dissertation will be provided. The chapter contains an overview of the diagnostic criteria of MDD and discusses prevalence rates and burden of disease associated with depression. Furthermore, theoretical underpinnings, rationales and techniques of CT and IPT are presented, together with an overview of their similarities and differences. In addition, the rationale for the research presented in this dissertation is explained, and the RCT is introduced. The chapter ends with the aims and outline of the dissertation.

Meet Anna

Anna is a 46 year-old married woman. She is a mother of two teenage children and has a full-time job as a lawyer. Anna has always been a high achiever. She has very high standards for herself and can be very self-critical when she fails to meet them. For the past few weeks, Anna has felt unusually fatigued and found it increasingly difficult to concentrate at work. She is consumed by negative thoughts and significant feelings of worthlessness and shame due to her inability to perform as well as she always has in the past. Her co-workers have noticed that she is often irritable and withdrawn, which is quite different from her typically upbeat and friendly disposition. She has called in sick on several occasions. On those days she stays in bed all day, watching TV or sleeping. Anna's husband has noticed changes as well. She is moody, has had difficulties falling asleep at night, and has shown little interest in sex. Although Anna has never considered suicide, she has found herself increasingly dissatisfied with her life. She has been having frequent thoughts of wishing she was dead. She gets frustrated with herself because she feels like she has every reason to be happy, yet she cannot seem to shake the sense of doom and gloom that has been clouding each day as of late[1].

Depression

Feeling sad from time to time is part of life. Sadness can be seen as a normal reaction to challenges in everyday life. However, when feelings of intense sadness are present (nearly) every day, keep a person from normal functioning, and also include other feelings such as helplessness, hopelessness, and worthlessness, it might well be Major Depressive Disorder (MDD). Major Depression is a psychiatric disorder characterized by depressed mood and markedly diminished interest or pleasure in almost all activities (anhedonia). Other symptoms include significant weight loss/gain or decreased/increased appetite, insomnia/hypersomnia, psychomotor agitation or retardation, fatigue or loss of energy, feelings of worthlessness or excessive or inappropriate guilt, concentration problems or

indecisiveness, and recurrent thoughts of death or suicidal ideation (American Psychiatric Association (APA), 2000). According to the fourth revised version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; APA, 2000), criteria for MDD are met when five or more of the symptoms mentioned above are present for (nearly) every day during a two-week period (with at least one of them being depressed mood or loss of interest/pleasure). Furthermore, the symptoms have to cause clinically significant distress or impairment in social-, occupational-, or other important areas of functioning. A complete overview of the diagnostic criteria for MDD according to the DSM-IV-TR can be found in Table 1. Depending on the number and severity of symptoms, an episode of MDD can be categorized as being mild, moderate or severe.

Prevalence

With lifetime prevalence estimates ranging from 15 to 20%, MDD is one of the most prevalent mental illnesses worldwide (de Graaf, ten Have, van Gool, & van Dorsselaer, 2012; Kessler et al., 2003; Simon, Goldberg, von Korff, & Üstün, 2002). According to the World Health Organization (WHO), globally more than 350 million people of all ages suffer from depression (WHO, 2012). MDD is present in all layers of society, regardless of social-economic status, ethnicity or profession, and rates are almost twice as high among women (Kessler et al., 2003; Velde van de, Bracke, & Levecque, 2010). A cross-national epidemiological study that compared life time prevalence estimates of major depressive episodes worldwide showed that the highest rates were found in high-income countries including France, the USA and the Netherlands (Bromet et al., 2011). The latest screening of the Dutch population (Nemesis-2 study) showed that over 500.000 adults (5.2% of the total Dutch population) suffer from MDD every year (de Graaf, ten Have, & van Dorsselaer, 2010). Research has shown that approximately 50% of those who recover from a first episode of depression will have at least one more episode later in life (Eaton et al., 2008). For patients with a history of two episodes this is as high as 80% (e.g., Frank et al., 1990; Kupfer et al., 1992; Prien & Kupfer, 1986). Because of its highly recurrent nature depression is more and more considered to be a chronic disorder.

Burden of Disease

Depression has an enormous impact on physical, social and emotional functioning and well-being (Bijl & Ravelli, 2000; Kessler et al., 2003; Kruijshaar, Hoeymans, Bijl, Spijker, & Essink-Bot, 2003) and is associated with high economic and societal costs (de Graaf, Tuithof, van Dorsselaer, & ten Have, 2011; Luppa, Heinrich, Angermeyer, König, & Riedel-Heller, 2007). MDD was ranked the third largest contributor to the burden of disease worldwide in 2004 and is even expected to be the leading global cause of years of health lost due to disease in middle- and high-income countries by 2030 (Mathers,

Ma Fat, & Boerma, 2008). Recent estimations by the Dutch National Institute for Public Health and the Environment (RIVM) show that depression yearly costs 168.000 Disability Adjusted Life Years (Daly's). This number indicates that the negative effects of depression on health in the Netherlands are even larger than those of e.g. dementia, diabetes or lung cancer (RIVM, 2013). The total annual costs of depression in the Netherlands (direct and indirect) are estimated at almost three billion euro (de Graaf et al., 2011; Slobbe, Smit, Groen, Poos, & Kommer, 2011).

Treatment of Depression

Given the increasing number of depressed patients, and the large impact of the disease, there is a need for effective and efficient treatments. And they do exist! Over the years, a wide range of interventions has shown to be effective in decreasing depressive symptoms in the acute phase. According to national and international guidelines for depression (National Institute for Health and Care Excellence (NICE), 2009; Spijker et al., 2013), MDD should be treated with pharmacotherapy or psychotherapy, or their combination. Pharmacological treatments for depressed outpatients include Tricyclic Antidepressants (TCA's), Selective Serotonin Reuptake Inhibitors (SSRI's), Selective Serotonin and Norepinephrine Reuptake Inhibitors (SNRI's), mirtazapine, or bupropion (Spijker et al., 2013). Psychological treatments for depression embrace a large assortment of interventions (see Cuijpers et al., 2014 for an overview), with Cognitive Therapy (CT) and Interpersonal Psychotherapy (IPT) being the two most commonly-practiced, well-studied and empirically validated (Cuijpers et al., 2013; Cuijpers et al., 2011; Hollon, Thase, & Markowitz, 2002). Over the years, numerous studies have shown that CT and IPT are well-standardized, efficacious interventions for the acute treatment of MDD (see meta-analyses of e.g., Barth et al., 2013 and Cuijpers et al., 2013 for an overview). In addition, CT and IPT are believed to produce enduring effects that reduce the risk for symptom return following treatment termination. Both treatments are therefore currently considered the psychological treatment of choice for MDD in the Netherlands (Spijker et al., 2013). Since CT and IPT play a central role in the current dissertation, their theoretical background, rational and content will be discussed below.

Cognitive Therapy (CT)

CT has its roots in Beck's (1967) cognitive theory of depression. According to cognitive theory, depressed mood is caused and maintained by maladaptive information processing strategies and dysfunctional beliefs. In other words, depressed people differ from non-depressed people with regard to both the content and process of thinking. Beck believes that the inner life of depressed patients is dominated by a set of silent assumptions. These assumptions stem from underlying schemas: cognitive core beliefs that are part of the filter people use to view themselves, the world and the future. Beck identified three main themes of core beliefs in depressed patients: 'I am worthless or inadequate', 'the world

Table 1. DSM-IV-TR Diagnostic Criteria for Major Depressive Disorder (MDD).

A.		Presence of a single Major Depressive Episode (see below).
	A.	Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning. At least one of the symptoms is (1) depressed mood or (2) loss of interest or pleasure.
		- Depressed mood most of the day, nearly every day, as indicated by either subjective report or observation made by others.
		- Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day.
		- Significant weight loss when not dieting or significant gain, or decrease or increase in appetite nearly every day.
		- Insomnia or hypersomnia nearly every day.
		- Psychomotor agitation or retardation nearly every day.
		- Fatigue or loss of energy nearly every day
		- Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
		- Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
		- Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or specific plan for committing suicide.
	B.	The symptoms do not meet the criteria for a mixed episode.
	C.	The symptoms cause clinically significant distress or impairment in social, occupational or other important areas of functioning.
	D.	The symptoms are not due to the direct physiological effects of a substance or a general medical condition.
	E.	The symptoms are not better accounted for by bereavement, i.e. after the loss of a loved one; the symptoms persist for longer than 2 months or are characterized by marked functional impairment, morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms or psychomotor retardation.
B.		The Major Depressive Episode is not better accounted for by Schizoaffective Disorder and is not superimposed on Schizophrenia, Schizophreniform Disorder, Delusional Disorder, or Psychotic Disorder Not Otherwise Specified.
C.		There has never been a Manic Episode, a Mixed Episode, or a Hypomanic Episode (*note*: this exclusion does not apply if all of the manic-like, mixed-like, or hypomanic-like episodes are substance or treatment induced or are due to the direct physiological effects of a general medical condition).

Source: American Psychiatric Association (2000, p.349–351 & 369–370).

is unfair, and 'the future is hopeless' (cognitive triad). These schemas are formed early in life but are supposed to only come to the surface in periods of stress. Once activated, they influence the content of conscious cognitions and responses to stressful life circumstances, thereby maintaining the symptoms (Beck, 1987; Clark, Beck, & Alford, 1999). According to the cognitive model, CT exerts its beneficial effects by altering the function, content and structure of cognitions and schemas associated with negative affect (Beck, Rush, Shaw, & Emery, 1979). The treatment therefore mainly focuses on identification and modification of distorted thinking patterns associated with depressive mood.

Over the course of treatment, patients are guided through several structured learning experiences. Initial treatment sessions are spent providing psycho-education, socialising the patient with the treatment model, and establishing a positive working alliance. Furthermore, the patient learns to monitor mood and behaviour to examine which activities lead to an increase of positive feelings, and which do not. Subsequently, the patient is encouraged to increase pleasant activities in order to (re)experience the rewarding effects of these activities.

The vast majority of sessions however, are spent identifying and modifying patterns of dysfunctional cognitions. Patients are asked to monitor negative thoughts and mental images in order to recognize the association between thoughts, feelings, physiology and behaviour. Furthermore, they learn to evaluate the validity and utility of these cognitions, and to replace unrealistic thoughts with alternatives that reflect reality more closely (Beck et al., 1979). In the beginning, cognitive interventions mainly involve monitoring and challenging negative automatic thoughts that are considered to be more state dependent and therefore more accessible. As therapy progresses, the focus shifts to the modification of underlying assumptions and core beliefs that may have predisposed someone to depression and make them vulnerable to future episodes. Cognitive therapists use a variety of techniques to help patients uncover and examine their thoughts including psycho-education, guided discovery, Socratic questioning, role playing, imagery, and behavioural experiments. Furthermore, patient and therapist focus on (re)learning the use of adaptive coping skills. The final sessions are used to build relapse prevention skills and discuss termination issues.

Each session has a similar structure. The session starts with a brief check on mood and symptoms and a short review of the previous session. After that, the agenda for today's session is set. The core of the session includes reviewing homework and discussing issues on the agenda. The session is concluded with a summary, setting of new homework and feedback for the therapist.

Interpersonal Psychotherapy (IPT)

IPT, developed by Klerman, Weissman, Rounsaville, and Chevron (1984), is based on the idea that depression often occurs in the context of social and interpersonal events, such as a dispute that threatens an important relationship, the death of a loved one, or life

changes that require role adaptations, e.g. a break-up, (threat of) dismissal, retirement, becoming a parent, or moving to a new neighbourhood (Weissman, Markowitz, & Klerman, 2000; Weissman, Markowitz, & Klerman, 2007). Once depressed, symptoms of the disorder further compromise interpersonal functioning, causing a downward spiral. IPT tries to understand the social and interpersonal context in which the depressive symptoms arose and investigates how they relate to the current social and personal context. The theorized mechanism is that if the patient can solve the interpersonal problem, or is able to change the negative emotions related to this problem, the depressive symptoms will resolve as well (Markowitz & Weissman, 2004). More specifically, Lipsitz and Markowitz (2013) propose that resolving the interpersonal problem might lead to symptom reduction through four specific mechanisms of change including enhancement of social support, decreasing interpersonal stress, facilitating emotional processing, and improving interpersonal skills.

IPT consists of three phases: a beginning, middle and end. Each phase has its specific tasks and functions. The initial phase (sessions 1–3) focuses on establishing a good working alliance and facilitating initial symptom relief. The therapist provides psycho-education about depression and the interpersonal context in which it presents itself. Furthermore, the patient is given the 'sick role': a temporary status recognizing that depression is a medical illness that keeps the patient from functioning at full capacity. In addition, patient and therapist elicit a thorough review of current and past relationships, called the 'interpersonal inventory'. This inventory is then used as a basis to define the interpersonal problem area that will serve as a primary treatment focus in the middle phase (sessions 4–13). IPT distinguishes four treatment foci. The focus 'grief/complicated bereavement' will be chosen when the depressive symptoms are strongly associated with the loss of a loved one. The goal here is to facilitate the mourning process, and help the patient re-establish interests and relationships. When the depression shows a strong association with a conflict with a significant other, the focus 'role dispute' might be an appropriate choice. In this focus, patient and therapist identify the dispute, explore and define a plan of action, and focus on the modification of faulty or non-reciprocal role expectations related to the dispute. The third focus, 'role transition', can be used when the depression seems to be related to important life changes. Primary goal here is facilitating mourning and acceptance of the loss of the old role. After that, the therapist helps the patient to see the new role in a more positive light, hereby restoring self-esteem. When a problem in the first three categories cannot be defined, and the problem primarily seems caused and maintained by poor social skills for making and maintaining relationships, the fourth focus, 'interpersonal deficits', will be selected. This focus is aimed at improving social skills and encouraging the patient to form new relationships hereby reducing the patient's social isolation. Even though each focus uses specific interventions to reach the treatment goals, all foci address the patients' ability to assert his/her needs and wishes in interpersonal encounters, and pull for affect. In the final phase (sessions 14–16) the patient is ac-

commodated to the fact that treatment is temporary and termination is nearing. Therapist and patient review the progress of the previous sessions to help the patient to feel more capable and independent (Weissman et al., 2000; Weissman et al., 2007).

Similarities and Differences between CT and IPT

CT and IPT share several common features. For example, both interventions are time-limited (16–20 sessions), symptom-targeted and present-focused (i.e. they address present thinking, behaviour and communication, rather than past experiences), and encourage the patient to regain control of mood and functioning. Furthermore, CT and IPT both try to increase the patient's activity level and pay special attention to the identification of expectations and assumptions using interventions such as exploration and clarification (Ablon & Jones, 1999; Willemse & Trijsburg, 2005). Other shared characteristics include offering a specific formulation of the individual's problem, openly sharing the model of therapy with the client, a rational use of techniques in a logical sequence, emphasis on skill development and transfer of learning outside of therapy sessions. To conclude, both therapies emphasize the importance of other (non-specific) factors such as structure, motivation and alliance.

There are also some important differences between the two interventions. First of all, CT seeks to identify, evaluate and change dysfunctional patterns of thinking in order to resolve depressive symptoms, whereas IPT focuses more on introspection and affect (feeling states). Second, the IPT therapist is a supportive ally, who stimulates healthy interpersonal skills as a cheerleader, and offers empathy when the patient doesn't succeed in an interpersonal situation. The CT therapist, on the other hand, is more directive and guides the patient through the process. A third important difference is the use of homework assignments. In CT there is a central role for homework assignments, whereas in IPT no formal homework is assigned.

Challenges in Psychotherapy Research

Even though the effectiveness of CT and IPT has been well established (Cuijpers et al., 2011; Cuijpers, van Straten, Andersson, & van Oppen, 2008; Hollon & Ponniah, 2010; Hollon et al., 2002), and both interventions are currently considered to be the treatment of choice for MDD, there is room for improvement. A recent study by Cuijpers et al. (2014), comparing the effects of 19 types of psychotherapy for depression, showed that overall 66% of patients treated with C(B)T and 57% of patients treated with IPT no longer met criteria for MDD at the end of treatment. Although encouraging, these remission indicate that approximately 40% of patients suffering from depression do not (or insufficiently) respond to initial treatment. Furthermore, even when treated effectively in the acute phase chances of relapse (episode of MDD after remission) and recurrence (episode of

MDD after recovery) are high (Keller & Boland, 1998; Paykel, 2008). The challenge in contemporary depression research is therefore to improve treatments to increase acute response rates and prevent relapse and recurrence in the long term.

Treatment improvement starts with a full understanding of the effects and mechanisms of interventions. By knowing how and why treatments work, certain components of interventions can be added, strengthened or removed to make them more efficient and (cost) effective. Furthermore, information on how treatments compare–and whether this differs for certain subgroups of patients–can be used to select the best available treatment for (individual) patients, hereby also increasing therapy effects. Although our knowledge about psychotherapy for depression has increased tremendously over the past few decades, psychological treatments – including CT and IPT – are not yet fully understood. There are several unresolved issues that need further examination. For instance, it is not clear yet whether one therapy outperforms the other with regard to severity and course of the disorder, especially on the long-term. Furthermore, it is still largely unknown how CT and IPT work and – more specifically – whether they work for reasons hypothesized in their respective theoretical backgrounds. These questions were the leading ground for conducting the large randomized controlled trial that is described in this dissertation. They will be further introduced below.

Relative effectiveness of CT vs. IPT

Numerous treatment studies have examined the effects of CT and IPT for depression. However, direct comparisons of the two treatment modalities are scarce. So far only three trials have compared individual CT and IPT head-to-head (Elkin, Shea, Watkins, & Imber, 1989; Luty et al., 2007; Quilty, McBride, & Bagby, 2008). And even though these studies indicate that both treatments are overall equally effective in treating depressive symptoms, it is too early to reject the possibility that they differ in effectiveness. Additional research is necessary to definitely settle this question. This has several reasons. First of all, a recent study by Jakobsen, Hansen, Simonsen, Simonsen, and Gluud (2012) has shown that so far the data collected in randomized comparisons of CT vs. IPT ($n = 526$) are insufficient to reliably decide whether the two therapies have differential treatment effects. Second, the three head-to-head comparisons are classified as having high risk of bias indicating that they may overestimate benefits and underestimate harm (Jakobsen et al., 2012). Third, none of these randomized comparisons included a waiting-list control (WLC) condition. This is remarkable since it has been frequently argued in the field of psychotherapy research that the effects of treatment studies cannot be attributed to the intervention(s) that are studied unless a placebo or no-treatment group is included (Klein, 1990). Fourth, in spite of general equality of treatments, individual patients might respond differently to different therapies. Therefore, it might be the case that one treatment is superior to the other for specific (subgroups of) patients, such as patients with severe depression. Unfortunately, these treatment moderators are still largely unknown.

Fifth, therapist effects are still poorly understood, and need to be examined. The sixth and final issue is related to enduring effects of the interventions. In addition to the fact that CT and IPT have shown to be efficacious in the acute phase, they are also believed to reduce the risk of relapse and recurrence in the long term. However, the evidence for this claim is limited and differs for CT and IPT. Research in CT has shown that cognitive interventions indeed have enduring effects that reduce the risk of relapse and recurrence (see Hollon, Stewart, & Strunk, 2006 for an overview). Research in IPT is less extensive. Even though IPT has shown to prevent relapse and recurrence when continued (Klerman, Dimascio, Weissman, Prusoff, & Paykel, 1974) or delivered in a less frequent maintenance format (Frank et al., 1990; Reynolds et al., 1999), the effects after treatment termination have not been tested yet. Therefore, it remains unclear whether IPT also has an enduring effect that remains after treatment is finished, and whether one of the two treatments is superior to the other in preventing relapse and recurrence. Additional research in randomized comparisons of CT and IPT including WLC conditions and follow-up assessments is necessary to answer these questions.

Mechanisms of Change

More and more researchers share the opinion that knowledge about the mechanisms that can explain therapeutic change is the key factor in the process of therapy improvement (Kazdin & Nock, 2003; Kraemer, Wilson, Fairburn, & Agras, 2002). Unfortunately, these mechanisms are still largely unknown. The fact that there still is no evidence-based explanation of how CT and IPT for depression work cannot be contributed to a lack of interest. The past decades, several research groups worldwide have conducted numerous studies aimed at identifying the active ingredients of therapy. Findings are promising, and definitely contribute to our knowledge on mechanisms, but do not fully explain why and how therapy leads to an improvement in depressive symptoms. Comparing recent work (e.g., Coleman, Cole, & Wuest, 2010; Warmerdam, van Straten, Jongsma, Twisk & Cuijpers, 2010) to work of 20 years ago (e.g., DeRubeis et al., 1990; Whisman, 1993), indicates that progress in this area has been slow. Mechanism research is very complicated and has to cope with theoretical, methodological and statistical difficulties (Kazdin, 2007; 2009). Several difficulties relevant to the topics of this dissertation will be addressed here.

First of all, difficulties arise because of conflicting hypotheses about what needs to be studied. Different research groups have different theories about change, resulting in a wide variety of rival hypotheses about mechanisms. The different views on change can be roughly divided into two groups. On the one hand there is the group that focuses on so-called specific factors; treatment specific elements marked in the theory of the therapy. CT and IPT stem from different theoretical backgrounds and claim to target depression through different therapy specific key-processes. The cognitive model assumes that interventions aimed at altering the function, content and structure of maladaptive patterns of thinking lead to symptom improvement in CT (Beck et al., 1979). According to

the interpersonal model, reduction of depressive symptoms is caused by improvement of interpersonal functioning by identifying and linking interpersonal stressors to symptoms of depression (Klerman et al., 1984). However, these assumptions have been rarely put to the test empirically, and it is therefore still unclear whether treatments actually work according to reasons specified in theory. Furthermore, so far empirical evidence for the notion that the mechanism should be specific for one type of treatment is lacking. Contrary to the view that treatments exert their beneficial effects through their own (specific) theorized mechanisms, is the idea that they work through change in factors common to all psychotherapies, such as expectations, motivational issues and the quality of the therapeutic alliance (Oei & Shuttlewood, 1996). This idea is mainly driven by the fact that various forms of psychotherapy that theoretically differ from each other and show differences in therapeutic approach (Ablon & Jones, 1999), have repeatedly shown to be equally effective in treating depression; a phenomenon known as the Dodo bird verdict. Unfortunately, because studies often only include either specific or common factors, there is no consensus yet about the order of significance and potential collaboration between theorized processes (DeRubeis et al., 1990; Garrat, Ingram, Rand, & Sawalani, 2007).

Even in situations in which it is clear which processes need to be studied (for example because one wants to test a specific theory), mechanisms research is difficult because of conceptual issues. Not only because core elements of these constructs can be defined in many different ways, but also because measurement instruments might not be sensitive enough (Johansson & Høglend, 2007). Furthermore, it is not clear yet whether changes achieved in therapy are best reflected by explicit or implicit measures of psychopathology (Clark et al., 1999). Explicit measures strongly depend on introspection, and a disadvantage is that it is known that people do not have access to all of the mechanisms that underlie their behaviour. To the extent that implicit measures reflect uncontrollable, unaware, fast mechanisms, they could provide insight beyond that from explicit measures. More specifically, they might potentially serve as a proxy of underlying schemata, thereby contributing to research on mechanisms. However, even though the interest for the use of implicit measures in research on mechanisms for depression has grown, and various measurement procedures have been developed to obtain implicit measures of self-esteem, current knowledge on mechanisms is still mainly based on research using explicit measures. This can be explained by the fact that procedures for obtaining implicit measures are relatively young and the best way to obtain implicit measures is still unclear. As a result, implicit processes are not well understood.

Second, difficulties arise in designing and conducting appropriate studies. Process research requires specific features of study designs. According to the latest standards, therapeutic change can only be examined properly in a theoretically well planned RCT with carefully spaced repeated measurements over the course of treatment, sufficient power and an appropriate control group (Kazdin, 2007; Kazdin & Nock, 2003; Kraemer et al., 2002; Laurenceau, Hayes, & Feldman, 2007). Furthermore, researchers should use uniform, state-of-the-art analytic techniques (Collins & Graham, 2002; Haaga & Stiles, 2000; Haubert &

Dobson, 2007; Kraemer et al., 2002; Laurenceau et al., 2007; MacKinnon, Fairchild, & Fritz, 2007). Unfortunately, up until now, most RCTs do not meet the criteria for reputable process research, mainly because they do not include repeated measures of clinical outcomes and process measures over the course of treatment (Kazdin, 2007, 2009). Without repeated measures it is impossible identify the exact shape of change, and to discern the temporal relationships necessary to identify mechanisms of change. A clear view of the shape of change is furthermore important because previous studies on change patterns in psychotherapy for depression have shown that a substantial portion of a patient's total reduction in symptom intensity often occurs suddenly (i.e. between two sessions) rather than gradually over the course of treatment (Elkin et al., 1989; Hollon et al., 1992; Tang & DeRubeis, 1999). These large and sudden improvements in symptoms – known as Sudden Gains – seem to have clinical advantages, but may also assist in the verification of mechanisms since they mark a critical point of change in treatment. Furthermore, traditional methods to monitor (change in) depressive symptoms might be inexhaustive. Usually, depression diagnostic instruments such as the BDI-II are analysed using a total score (i.e. the sum of the individual item scores). A drawback of this method is that it assumes the symptoms of depression to be statistically independent. This is remarkable, because theories of the treatment of depression focus specifically on the relation between symptoms when describing their proposed mechanisms of change. For example, Beck et al. (1979) state that change in cognitive processes (negative thinking), leads to change in symptoms such as affect. Recent research therefore suggests that, in addition to studying (change in) the mean level of total scores, researchers should also focus on the relation between individual symptoms as they unfold over time to gain more insight in the mechanisms of change. Unfortunately, research in this field is still in its infancy (partly because intensive longitudinal data in which a set of symptoms is measured frequently across time is scarce). In addition, studies often use different (out-dated) statistical analyses to examine the underlying mechanisms of change complicating comparisons across studies and integration of findings into broader knowledge. To conclude, process research often has to deal with small sample sizes. Underpowered studies are extra troubling in a field in which there often are only subtle differences between the intervention groups, such as CT and IPT.

Aims of this dissertation

It may be clear that there is a need for additional and methodologically solid research examining the (long-term) effects of CT vs. IPT and their underlying mechanisms. The general aim of this dissertation is therefore twofold. The first objective is to provide more insight in the relative effectiveness of CT vs. IPT, both in the acute phase and after treatment termination. This is done by evaluating and comparing the clinical effects of CT, IPT, and – for the short-term effects – a waiting-list control condition in the context of a large RCT. The second objective is to broaden our knowledge on the underlying processes that facilitate symptom change. Towards this end, we have critically reviewed the exist-

ing literature on mechanisms, and conducted four empirical studies in which we examined various aspects of psychotherapeutic change. With this we aim to shed more light on the quality of existing process research, the exact shape of change over the course of treatment and the extent to which CT and IPT differ in this regard, the relation between individual symptoms of depression, and the extent to which treatments actually work for reasons hypothesized in their respective theoretical backgrounds. In addition, by exploring the value of a newly developed implicit self-associative measure for depression, we aim to contribute to the knowledge on the use of implicit measures in psychotherapy process research. The specific aims and hypotheses per study are outlined below.

Data used in this dissertation

All data used in this dissertation were collected within one research project: a large single-centre RCT (parallel group design) examining the effectiveness and mechanisms of change of individual CT and IPT for adult depression (STEPd; ISRCTN67561918). A total of 182 depressed outpatients were randomly assigned to one of three conditions: (a) CT ($n=76$), (b) IPT ($n=75$), or (c) a 2-month WLC condition followed by treatment of choice ($n=31$). Participants were recruited during regular intakes at the Maastricht Community Mental Health Centre (RIAGG). Follow-up data were gathered within a 24 month time frame. In reporting the follow-up of this study, three phases can be distinguished: (1) the 'Treatment Phase' (baseline to 7 months); the period when therapy was delivered; (2) the 'Trial Follow-Up Phase' (month 8–12) in which depression severity was measured monthly; and (3) the 'Long-term Follow-Up Phase' (month 12–24); the period covered by retrospective assessment at 24 months. The data presented in this dissertation concern data collected up until 12 months (Treatment Phase and Trial Follow-Up Phase). Figure 1 gives a schematic overview of the study design and indicates which data are used within the chapters of this dissertation. More information about the design of the study, recruitment procedure and data collection within this project can be found in chapter 2 (study design).

Outline of this dissertation

This dissertation comprises 9 chapters. **Chapter 2** presents the design of our RCT. It contains a rationale for the trial, an overview of research questions, and an extensive description of the procedure, measures and statistical analyses. **Chapter 3** focuses on the clinical effectiveness of CT vs. IPT in the acute phase and reports on the short-term follow-up. We tested whether active treatment was superior to waiting-list control after 2 months, and examined whether one of the treatments outperformed the other in decreasing depression severity and improving quality of life at post-treatment (7 months) and up to 5-months after treatment termination. It was hypothesized that active treatments would be more effective than waiting-list control, but that no differences would emerge between the two active interventions.
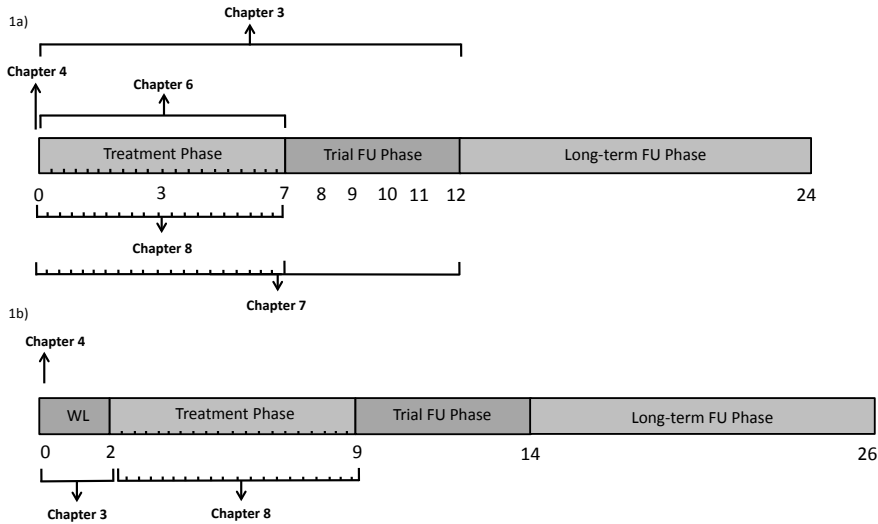
Figure 1. Study design of the RCT and overview of the data used in each chapter. Panel 1a: Active conditions Cognitive Therapy (CT) and Interpersonal Psychotherapy (IPT). Panel 1b: Waiting-List Control (WLC) condition.

Chapters 4 to 8 examine the various questions related to the potential mechanisms of change. The study presented in **Chapter 4** aims to contribute to our knowledge on the use of implicit measures in mechanisms research by examining the value of a newly developed implicit self-associative measure specifically related to core beliefs of depression. By comparing scores on implicit and explicit measures of self-esteem of depressed patients and healthy controls, the hypothesis was tested that depressed patients, who were about to start psychotherapy, would show more negative self-associations than would healthy control participants. Furthermore, we examined whether the implicit and explicit self-associative measures were associated with each other and with depressive symptoms, and investigated the effect of a discrepancy between the implicit and explicit measure.

Chapters 5 and 6 focus on mediation analysis, the most commonly used method to examine mechanisms of change. In **Chapter 5**, a systematic overview and critical evaluation of the empirical literature on psychological mediators in psychotherapy for depression is provided. We describe characteristics and results of 32 studies that were identified in a systematic literature search, and report on the extent to which they meet the up-to-date requirements for appropriate mechanism research. **Chapter 6** reports the result of an empirical study aimed at testing the theoretical models of CT and IPT, and the common factor model. We explored change in five (specific and non-specific) candidate working mechanisms over the course treatment, and examined whether this differed for CT and IPT. Furthermore, the temporal relationships between change in potential mechanisms and change in depressive symptoms were examined. In addition, we investigated whether the-

orized processes of change mediated the relation between treatment and outcome. It was hypothesized that scores on the process measures would change in an adaptive direction in both conditions. Furthermore, based on theoretical underpinnings we expected the effects of treatment to be mediated by theorized factors such as cognitive change in CT and interpersonal change in IPT. However, based on previous empirical studies, we did not rule out the possibility that the treatments would work through similar mechanisms.

The empirical study reported in **Chapter 7** aims to shed light on the occurrence and clinical impact of sudden gains – large symptom improvements in a single between-session interval marking a critical point of change. We identified the patients who met criteria for sudden gains and examined the relation between sudden gain status and end-of-treatment symptom severity, as well as symptom severity at five-months follow-up. In addition, we explored baseline patient characteristics that might predict the occurrence of sudden gains, both within each treatment as well as across the two treatments. We expected sudden gains to appear at a similar rate as in other studies of treatment for depression. Furthermore, we expected that patients with sudden gains would experience superior outcomes as compared to those without sudden gains. With regard to the pre-treatment factors associated with sudden gains, it was hypothesized that lower levels of depressive symptomatology would be predictive of sudden gains.

Chapter 8 describes an empirical study of the dynamic (i.e. session-to-session) relations between individual items of the BDI-II over the course of treatment as an alternative to the commonly used total score. We explored the network representing the session-to-session relations between the 21 BDI-II symptoms across 20 weeks of treatment, and examined whether this differed for CT and IPT. Also, the centrality of symptoms and the community structure of the BDI-II (i.e. the presence of specific clusters of symptoms that are more strongly interconnected with each other than with symptoms in the network) were examined. We expected the various symptoms of depression to be strongly connected to each other. Furthermore, it was hypothesized that BDI-II items that were most closely related to the main symptoms of depression (anhedonia and depressed mood) would be the most central in the network. In addition, given the fact that previous cross-sectional research has identified specific clusters of symptoms within the BDI-II, we expected to find various communities.

In **Chapter 9**, findings of the various studies presented in this dissertation are summarized and integrated in the broader knowledge base. Furthermore, a critical evaluation of our study, clinical implications and suggestions for future research are provided. The **Valorization Addendum** describes the societal, economic and scientific relevance of our trial, target groups and innovative aspects of this dissertation.

Footnotes

[1] Case example is an adapted version of a patient description provided by PsyWeb–Depression and Mental Health Resource.

References

Ablon, J. S., & Jones, E. E. (1999). Psychotherapy process in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 67*(1), 64–75.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision*. Washington D.C.: American Psychiatric Association.

Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., . . . Cuijpers, P. (2013). Comparative Efficacy of Seven Psychotherapeutic Interventions for Patients with Depression: A Network Meta-Analysis. *PloS Medicine, 10*(5), e1001454.

Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York, NY: Hoeber.

Beck, A. T. (1987). Cognitive models of depression. *Journal of Cognitive Psychotherapy: An International Quarterly, 2*, 5–37.

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

Bijl, R. V., & Ravelli, A. (2000). Current and residual functional disability associated with psychopathology: findings from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Psychological Medicine, 30*, 657–668.

Bromet, E., Andrade, L., Hwang, I., Sampson, N., Alonso, J., de Girolamo, G., . . . Kessler, R. (2011). Cross-national epidemiology of DSM-IV major depressive episode. *BMC Medicine, 9*(1), 90.

Clark, D. A., Beck, A. T., & Alford, B. A. (1999). *Scientific Foundations of Cognitive Theory and Therapy of Depression*. New York, NY: John Wiley and Sons.

Coleman, D., Cole, D., & Wuest, L. (2010). Cognitive and Psychodynamic Mechanisms of Change in Treated and Untreated Depression. *Journal of Clinical Psychology, 66*(3), 215–228.

Collins, L. M., & Graham, J. W. (2002). The effect of the timing and temporal spacing of observations in longitudinal studies of tobacco and other drug use: Temporal design considerations. *Drug and Alcohol Dependence, 68*, 85–96.

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

Cuijpers, P., Geraedts, A. S., van Oppen, P., Andersson, G., Markowitz, J. C., & van Straten, A. (2011). Interpersonal Psychotherapy for Depression: A Meta-Analysis. *American Journal of Psychiatry, 168*, 581–592.

Cuijpers, P., Karyotaki, E., Weitz, E., Andersson, G., Hollon, S. D., & van Straten, A. (2014). The effects of psychotherapies for major depression in adults on remission, recovery and improvement: A meta-analysis. *Journal of Affective Disorders, 159*(0), 118–126.

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology, 76*(6), 909–922.

DeRubeis, R. J., Evans, M. D., Hollon, S. D., Garvey, M. J., Grove, W. M., & Tuason, V. B. (1990). How Does Cognitive Therapy Work? Cognitive Change and Symptom Change in Cogntive Therapy and Pharmacotherapy for Depression. *Journal of Consulting and Clinical Psychology, 58*(6), 862–869.

Eaton, W. W., Shao, H., Nestadt, G., Lee, H. B., Bienvenu, O. J., & Zandi, P. (2008). Population-based study of first onset and chronicity in major depressive disorder. *Archives of General Psychiatry, 65*(5), 513–520.

Elkin, I., Shea, M. T., Watkins, J. T., & Imber, S. D. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry, 46*(11), 971–982.

Frank, E., Kupfer, D. J., Perel, J. M., Cornes, C., Jarret, D. B., Mallinger, A. G., . . . Grochocinski, V. (1990). Three-year outcomes for maintenance therapies in recurrent depression. *Archives of General Psychiatry, 47*(12), 1093–1099.

Garrat, G., Ingram, R. E., Rand, K. L., & Sawalani, G. (2007). Cognitive Processes in Cognitive Therapy: Evaluation of the Mechanisms of Change in the Treatment of Depression. *Clinical Psychology, 14*, 224–239.

Graaf de, R., ten Have, M., & van Dorsselaer, C. (2010). *De psychische gezondheid van de Nederlandse bevolking–NEMESIS-2: Opzet en eerste resultaten*. Utrecht: Trimbos-Instituut.

Graaf de, R., ten Have, M., van Gool, C., & van Dorsselaer, C. (2012). Prevalentie van psychische aandoeningen en trends van 1996 tot 2009; resultaten van NEMESIS-2. *Tijdschrift voor Psychiatrie, 1*, 27–38.

Graaf de, R., Tuithof, M., van Dorsselaer, S., & ten Have, M. (2011). *Verzuim door psychische en somatische aandoeningen bij werkenden: Resultaten van de 'Netherlands Mental Health Survey and Incidence Study-2' (NEMESIS-2)*. Utrecht: Trimbos-Instituut.

Haaga, D. A. F., & Stiles, W. B. (2000). Randomized clinical trials in psychotherapy research; Methodology, design and evaluation. In C. R. Snyder & R. E. Ingram (Eds.), *Handbook of psychological change; psychotherapy processes and practices for the 21st Century* (pp. 14–39). New York, NY: Wiley.

Haubert, L. C., & Dobson, K. S. (2007). Treatment of Depression and Mechanisms of Change: Strengthening the Links among Theory, Research and Practice. *Clinical Psychology: Science and Practice, 14*, 247–251.

Hollon, S., D., Stewart, M. O., & Strunk, D. (2006). Enduring effects for cognitive behaviour therapy in the treatment of depression and anxiety. *Annual Review of Clinical Psychology, 57*, 285–315.

Hollon, S. D., DeRubeis, R. J., Evans, M. D., Wiemer, M. J., Garvey, M. J., Grove, W. M., & Tuason, V. B. (1992). Cognitive therapy and pharmacotherapy for depression: Singly and in combination. *Archives of General Psychiatry, 49*(10), 774–781.

Hollon, S. D., & Ponniah, K. (2010). A review of empirically supported psychological therapies for mood disorders in adults. *Depression and Anxiety, 27*(10), 891–932.

Hollon, S. D., Thase, M. E., & Markowitz, J. C. (2002). Treatment and prevention of depression. *Psychological Science and Public Interest, 3*, 39–77.

Jakobsen, J. C., Hansen, J. L., Simonsen, S., Simonsen, E., & Gluud, C. (2012). Effects of cognitive therapy versus interpersonal psychotherapy in patients with major depressive disorder: a systematic review of randomized clinical trials with meta-analyses and trial sequential analyses. *Psychological Medicine, 42*, 1343–1357.

Johansson, P., & Høglend, P. (2007). Identifying mechanisms of change in psychotherapy: Mediators of treatment outcome. *Clinical Psychology & Psychotherapy, 14*(1), 1–9.

Kazdin, A. E. (2007). Mediators and Mechanisms of Change in Psychotherapy Research. *Annual Review of Clinical Psychology, 3*, 1–27.

Kazdin, A. E. (2009). Understanding how and why psychotherapy leads to change. *Psychotherapy Research, 19*(4–5), 418–428.

Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, 44*(8), 1116–1129.

Keller, M. B., & Boland, R. J. (1998). Implications of failing to achieve successful long-term maintenance treatment of recurrent unipolar major depression. *Biological Psychiatry, 44*(5), 348–360.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., . . . Wang, P. S. (2003). The epidemiology of major depressive disorder: Results from the national comorbidity survey replication (ncs-r). *JAMA, 289*(23), 3095–3105.

Klein, D. F. (1990). NIMH collaborative research on treatment of depression. *Archives of General Psychiatry, 47*, 682–684.

Klerman, G. L., Dimascio, A., Weissman, M., Prusoff, B., & Paykel, E. S. (1974). Treatment of Depression by Drugs and Psychotherapy. *American Journal of Psychiatry, 131*, 186–191.

Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. S. (1984). *Interpersonal psychotherapy for depression*. New York, NY: Basis Books.

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*, 877–883.

Kruijshaar, M. E., Hoeymans, N., Bijl, R. V., Spijker, J., & Essink-Bot, M. L. (2003). Levels of disability in Major Depression. Findings from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Journal of Affective Disorders, 77*(1), 53–64.

Kupfer, D. J., Frank, E., Perel, J. M., Cornes, C., Mallinger, A. G., Thase, M. E., . . . Grochocinski, V. J. (1992). 5-year outcome for maintenance therapies in recurrent depression. *Archives of General Psychiatry, 49*(769–773).

Laurenceau, J.-P., Hayes, A. M., & Feldman, G. C. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review, 27*(6), 682–695.

Lipsitz, J. D., & Markowitz, J. C. (2013). Mechanisms of change in interpersonal therapy (IPT). *Clinical Psychology Review, 33*(8), 1134–1147.

Luppa, M., Heinrich, S., Angermeyer, M. C., König, H.-H., & Riedel-Heller, S. G. (2007). Cost-of-illness studies of depression: A systematic review. *Journal of Affective Disorders, 98*(1–2), 29–43.

Luty, S. E., Carter, J. D., McKenzie, J. M., Rae, A. M., Frampton, C. M. A., Mulder, R. T., & Joyce, P. R. (2007). Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *British Journal of Psychiatry, 190*, 496–502.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation Analysis. *Annual Review of Clinical Psychology, 58*, 593–614.

Markowitz, J. C., & Weissman, M. M. (2004). Interpersonal psychotherapy: principles and applications. *World Psychiatry, 3*, 136–139.

Mathers, C., Ma Fat, D., & Boerma, J. T. (2008). *The global burden of disease: 2004 update*. Geneva: World Health Organization.

National Institute for Health and Care Excellence (NICE). (2009). *Depression in Adults: The treatment and management of depression in adults (CG90)*. London: National Institute for Health and Care Excellence.

Oei, T. P. S., & Shuttlewood, G. J. (1996). Specific and Nonspecific factors in psychotherapy; a case of cognitive therapy for depression. *Clinical Psychology Review, 16*(2), 83–103.

Paykel, E. S. (2008). Partial remission, residual symptoms, and relapse in depression. *Dialogues in Clinical Neuroscience, 14*, 431–437.

Prien, R. F., & Kupfer, D. J. (1986). Continuation drug therapy for major depressive episodes: How long should it be maintained? *American Journal of Psychiatry, 143*, 18–23.

Quilty, L. C., McBride, C., & Bagby, R. M. (2008). Evidence for the cognitive mediational model of cognitive behavioural therapy for depression. *Psychological Medicine, 38*, 1531–1541.

Reynolds, C. F., Frank, E., Perel, J. M., Imber, S. D., Cornes, C., Miller, M. D., . . . Kupfer, D. J. (1999). Nortriptyline and interpersonal psychotherapy as maintenance therapies for recurrent major depression: A randomized controlled trial in patients older than 59 years. *JAMA, 281*(1), 39–45.

RIVM. (2013). *Volksgezondheid Toekomst Verkenning: Nationaal Kompas Volksgezondheid, versie 4.16*. Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu (RIVM).

Simon, G. E., Goldberg, D. P., von Korff, M., & Üstün, T. B. (2002). Understanding cross-national differences in depression prevalence. *Psychological Medicine, 32*(4), 585–594.

Slobbe, L. C. J., Smit, J. M., Groen, J., Poos, M. J. J. C., & Kommer, G. J. (2011). *Kosten van Ziekten in Nederland 2007: Trends in de Nederlandse zorguitgaven 1999–2010*. Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu (RIVM).

Spijker, J., Bockting, C. L. H., Meeuwissen, J. A. C., van Vliet, I. M., Emmelkamp, P. M. G., Hermens, M. L. M., & van Balkom, A. L. J. M. (2013). *Multidisciplinaire Richtlijn Depressie (Derde revisie): Richtlijn voor de diagnostiek, behandeling en begeleiding van volwassen patiënten met een depressieve stoornis*. Utrecht: Trimbos-instituut.

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 67*(6), 894–904.

Velde van de, S., Bracke, P., & Levecque, K. (2010). Gender differences in depression in 23 European countries. Cross-national variation in the gender gap in depression. *Social Science and Medicine, 71*(2), 305–313.

Warmerdam, L., van Straten, A., Jongsma, J., Twisk, J., & Cuijpers, P. (2010). Online cognitive behavioral therapy and problem-solving therapy for depressive symptoms: Exploring mechanisms of change. *Journal of Behavior Therapy and Experimental Psychiatry, 41*, 64–70.

Weissman, M. M., Markowitz, J. C., & Klerman, G. L. (2000). *Comprehensive guide to interpersonal psychotherapy*. New York, NY: Basic Books.

Weissman, M. M., Markowitz, J. C., & Klerman, G. L. (2007). *Clinician's Quick Guide to Interpersonal Psychotherapy*. New York, NY: Oxford University Press.

Whisman, M. A. (1993). Mediators and moderators of change in cognitive therapy of depression. *Psychological Bulletin, 114*, 248–265.

Willemse, Y., & Trijsburg, R. W. (2005). Cognitieve gedragstherapie en interpersoonlijke psychotherapie. Een analyse van kritische succesfactoren. *Tijdschrift voor Psychiatrie, 47*, 593–602.

World Health Organization (WHO). (2012). Fact Sheet Depression (N°369). from http://www.who.int/mediacentre/factsheets/fs369/en/

2

Effectiveness, Relapse prevention and
Mechanisms of Change of Cognitive Therapy vs.
Interpersonal Psychotherapy for Depression:
Study protocol for a Randomized Controlled Trial

2

Abstract

Major depression is a common mental disorder that substantially impairs quality of life and has high societal costs. Although psychotherapies have proven to be effective antidepressant treatments, initial response rates are insufficient and the risk of relapse and recurrence is high. Improvement of treatments is badly needed. Studying the mechanisms of change in treatment might be a good investment for improving everyday mental health care. However, the mechanisms underlying therapeutic change are still largely unknown. The objective of the current study is to assess both the effectiveness of two commonly used psychotherapies for depression in terms of reduction of symptoms and prevention of relapse on short and long term, as well as identifying their underlying mechanisms of change. In a randomized trial we will compare (a) Cognitive Therapy (CT) with (b) Interpersonal Psychotherapy (IPT), and (c) an 8-week waiting-list control (WLC) condition followed by treatment of choice. One hundred eighty depressed patients (aged 18–65) will be recruited in a mental health care centre in Maastricht (the Netherlands). Eligible patients will be randomly allocated to one of the three intervention groups. The primary outcome measure of the clinical evaluation is depression severity measured by the Beck Depression Inventory-II (BDI-II). Other outcomes include process variables such as dysfunctional beliefs, negative attributions, and interpersonal problems. All self-report outcome assessments will take place on the internet at baseline, three, seven, eight, nine, ten, eleven, twelve and twenty-four months. At 24 months a retrospective telephone interview will be administered. Furthermore, a rudimentary analysis of the cost-effectiveness will be embedded. The study has been ethically approved and registered. By comparing CT and IPT head-to-head and by investigating multiple potential mediators and outcomes at multiple time points during and after therapy, we hope to provide new insights in the effectiveness and mechanisms of change of CT and IPT for depression, and contribute to the improvement of mental health care for adults suffering from depression. The study has been registered at the Netherlands Trial Register, part of the Dutch Cochrane Centre (ISRCTN67561918).

Introduction

With a lifetime prevalence of 17%, depression is a major health problem with serious clinical and social consequences. It is expected that depression will be the leading global cause of years of health lost due to disease in 2030 (Mathers, Ma Fat, & Boerma, 2008). With initial response rates up to 60%, and the majority of patients regaining their normal level of functioning within three years, several psychotherapies and antidepressant medication have proven to be effective in treating acute Major Depressive Disorder (MDD; Hollon & Ponniah, 2010; Roelofs & Muris, 2005). This might sound promising, but also shows that at least 40% of depressed patients do not respond (sufficiently) to initial treatment. Furthermore, depression has an unfavourable prognosis; even when treated effectively in the acute phase, recovery is often incomplete, which increases the chance of relapse and recurrence up to 87% over 15 years (Beers et al., 2005; Keller & Boland, 1998; Ormel, Oldehinkel, Brilman, & Brink, 2993; van Londen, Molenaar, Goekoop, Zwinderman, & Rooijmans, 1998).

With this in mind, the challenge in contemporary depression research is to improve treatments to increase acute response rates and prevent relapse and recurrence in the long term. Many researchers agree that knowledge of the underlying mechanisms that can explain therapeutic change is a key to improving treatment (e.g. Kazdin & Nock, 2003; Kraemer, Wilson, Fairburn, & Agras, 2002). Knowing how a therapy works allows honing its components to make it more efficient and (cost-) effective (Warmerdam, van Straten, Jongsma, Twisk, & Cuijpers, 2010). The current study will focus on the effectiveness, prevention of relapse and recurrence and mechanisms of change of two commonly used types of psychotherapy for depression: Cognitive Therapy (CT; Beck, Rush, Shaw, & Emery, 1979) and Interpersonal Psychotherapy (IPT; Klerman, Weissman, Rounsaville, & Chevron, 1984). A rudimentary analysis of the cost-effectiveness from a societal perspective will be embedded.

Effectiveness

Of the psychotherapeutic interventions for depression, CT and IPT might be the two best studied and empirically validated (Cuijpers et al., 2011; Cuijpers, van Straten, Andersson, & van Oppen, 2008; Hollon, Thase, & Markowitz, 2002). They come from different theoretical backgrounds: CT derives from Beck's (1979) cognitive theory and explains depression as a result of maladaptive information-processing, whereas IPT links depressive episodes to distressing life events and insufficient social support (Beck et al., 1979; Klerman et al., 1984). Both therapies have proven to be well-standardized, efficacious treatments for acute treatment of MDD (Butler, Chapman, Forman, & Beck, 2006; DeRubeis & Crits-Christoph, 1998; Emmelkamp, 1992; Hollon & Ponniah, 2010; Klerman & Weissman, 1993; Klerman et al., 1984; Strunk & DeRubeis, 2001; Weissman, 1979; Weissman & Markowitz, 1994). However, there is no consensus yet about whether the effect of one therapy outper-

forms the other. Many studies have investigated the effects of CT and IPT separately, but only 3 large studies have compared them head-to-head (Elkin, Shea, Watkins, & Imber, 1989; Luty et al., 2007; Quilty, McBride, & Bagby, 2008). Furthermore, doubts have arisen about the validity of one of these studies (Elkin et al., 1989) because analysis of treatment adherence showed no contrast between the two intervention groups (Ablon & Jones, 1999). Thus, the current view is based on only two methodologically well-designed studies. Therefore, there is a need for additional head-to-head comparisons of both therapies.

Relapse Prevention

In addition to the fact that CT and IPT have shown to be efficacious acute treatments of MDD, they may also reduce the risk of relapse (episode of MDD after remission) and recurrence (episode of MDD after recovery) in the long term. The effects and evidence differ for the two therapies. Research has shown that CT has an enduring effect that extends beyond the end of therapy (Bockting et al., 2005; Fava et al., 2004; Frank, Kupfer, Perel, & Cornes, 1990; Hollon, 2003; Hollon et al., 2005a; Klerman et al., 1984; Paykel et al., 2005), thereby reducing the chances of relapse and recurrence. The evidence for this is strong and consistent (Hollon, 2003; Jarret et al., 2001; Teasdale et al., 2000). However, the long term effect of IPT has not been tested extensively yet. Up until now it has only been tested as a maintenance treatment (Frank et al., 2007; Frank, Kupfer, Wagner, & McEachran, 1991), and the question remains whether IPT also has an enduring effect that remains after therapy is finished. This question should be further explored. Insight in the long term effects of IPT furthermore creates the opportunity to compare CT and IPT to assess whether one therapy is superior to the other in preventing relapse and recurrence in the long term (Hollon et al., 2002).

Mechanisms of Change

As noted above, insight into mechanisms of change might contribute to the process of therapy improvement. However, the mechanisms that cause therapeutic change are still largely unknown. Despite several research attempts to identify the mechanisms of change in psychotherapy, no study has identified a model that explains change in CT or IPT completely (Garrat, Ingram, Rand, & Sawalani, 2007; Longmore & Worrell, 2007). Mechanism research is complicated and has to cope with several methodological and theoretical difficulties (Kazdin, 2007, 2009). Theoretical difficulties arise because there are conflicting hypotheses about which mechanisms need to be studied and there is no consensus about the most important causes of change (DeRubeis et al., 1990; Garrat et al., 2007). For example, it is unclear whether therapeutic change can be better explained by change in treatment specific factors (DeRubeis et al., 1990; Furlong & Oei, 2002; Kwon & Oei, 2003; Quilty, McBride, & Bagby, 2008) or non-specific (common) factors (Coleman,

Cole, & Wuest, 2010; Kolko, Brent, Baugher, Bridge, & Birmaher, 2000; Oei, Bullbeck, & Campbell, 2006). Specific factors are elements marked as the active causes of change in the theory of the therapy, such as change in cognitions in CT and optimization of interpersonal functioning in IPT. Non-specific factors refer to elements in a therapy that contribute to improvement, but that are common to all psychotherapies, such as expectancy and therapeutic alliance (Oei & Shuttlewood, 1996). Furthermore, it is not exactly known whether changes achieved in therapy are best reflected by explicit or by implicit measures of psychopathology. Explicit measures depend on introspection, and a disadvantage is that it is known that people do not have access to all of the mechanisms that underlie their behaviour (Nisbett & Wilson, 1977). An implicit measure is defined by De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009) as "a measurement outcome that is causally produced by the to-be-measured attribute in the absence of certain goals, awareness, substantial cognitive resources, or substantial time" (p. 350). To the extent that implicit measures reflect uncontrollable, unaware, and fast mechanisms, they could provide information that augments that from explicit measures (Roefs et al., 2011).

Furthermore, methodological difficulties arise because many study designs do not meet the criteria for reputable mechanisms research (Kazdin, 2007, 2009). Theories often explain change in terms of causal processes. However, in many studies it is difficult to identify temporal relationships in order to investigate these causal pathways because of the absence of an appropriate time line and assessment on multiple time points (Kazdin, 2007, 2009). It is clear that there is a need for renewed, methodologically well-considered mechanism research.

The question that remains is what is necessary for proper mechanism research. According to Kazdin (2007), a first step into investigating mechanisms of change is studying mediating variables (Kazdin, 2007). A mediator explains why and in what way a treatment has an effect on the outcome, and plays a crucial role in the development of causal pathways. In identifying mediators, Kazdin has built upon the MacArthur guidelines of Kraemer, Stice, Kazdin, Offord, and Kupfer (2001) which are based on the more traditional guidelines for statistical mediation formulated by Baron and Kenny (1986). In addition to statistical mediation, Kazdin emphasizes the importance of the temporal relationship, and consistency and specificity of the mediator. The importance of the aspect of temporality is also emphasized by Murphy, Cooper, Hollon, and Fairburn (2009). Taking this into account, the current study will investigate potential mediators of CT for depression and test their specificity in comparison to IPT, and vice versa, by measuring multiple potential mediators and outcomes at multiple time points during and after therapy. This method enables us to investigate temporal relationships between changes in potential mediators and symptom reduction and to assess whether change in a mediator precedes, follows from, or goes together with changes in depression. In addition, this method can show us whether change in one mediator precedes change in another mediator.

Main research questions and Hypotheses

The following main research questions were formulated: 1) Are CT and IPT effective interventions in treating the acute phase of MDD, and is one therapy superior to the other? 2) What are the underlying psychological mechanisms of change in CT and IPT and are these mechanisms therapy-specific? 3) Are CT and IPT effective in preventing relapse or recurrence of MDD in the long-term? Is one therapy superior to the other, and if so, how can these preventive effects be explained?

In line with previous research, it is hypothesized that the amount of change in depressive symptoms after therapy will be similar in both the CT and the IPT group, indicating that both treatments are equally effective in treating depression in the acute phase (Cuijpers et al., 2008; Dorrepaal, van Nieuwenhuizen, Schene, & de Haan, 1998; Elkin et al., 1989; Willemse & Trijsburg, 2005). With regard to the mechanisms of change, many hypotheses are possible, especially when it comes to the order of change and causal pathways that lead to recovery. Following the theoretical models of CT and IPT one would expect that both treatments target depression through different key processes. Consequently, changes in cognitive schemas, attitudes, and cognitions are hypothesized to be the most significant contributors to symptom change in CT, whereas in IPT it is assumed that improvement of interpersonal functioning will lead to a reduction of symptoms (Beck et al., 1979; Klerman et al., 1984). Furthermore, we expect that it is the direct comparison of explicit versus implicit measures that will elicit new insights into the underlying mechanisms of change. To the extent that implicit measures reflect uncontrollable, unaware and fast mechanisms, they can provide useful additional information as compared to explicit measures. As to the prevention of relapse, we hypothesize that both CT and IPT may reduce the risk of relapse and recurrence in the long term. However, we expect CT to prevent relapse to a greater extent, because it has shown to have an enduring effect that extends beyond the end of therapy, while IPT so far only seems to be effective in treating depression as long as the treatment is continued (Hollon et al., 2005b).

Methods

Design of the study

We will conduct a randomized controlled trial (RCT) in which participants will be allocated to one of three conditions: (a) CT ($N=75$), (b) IPT ($N=75$), (c) or a waiting list control (WLC) condition followed by treatment of choice ($N=30$). Participants allocated to the waiting list condition start their treatment after an 8-week waiting period. To compensate for the waiting, they may choose their preferred treatment (CT or IPT). The anticipated flow of subject enrolment is graphically shown in Figure 1. The Medical Ethics Committee of Maastricht University approved the study protocol. The study is registered at the Netherlands Trial Register, part of the Dutch Cochrane Centre (ISRCTN67561918).

Participants

The population we aim to investigate consists of 180 depressed adults. Patients will be eligible to participate if they meet the following criteria: age between 18 – 65 years old; the presence of a depressive episode as indicated by a diagnosis on the Structured Clinical Interview for DSM-IV Axis I disorders (SCID-I; First, Spitzer, Gibbon, & Williams, 1997); internet access and an e-mail address; no bipolar or chronic (current episode > 5 years) depression; no current use of antidepressant medication, or concomitant, psychological treatment; no drugs- and alcohol abuse/dependence and/or mental retardation (IQ < 80).

Sample size

To calculate the sample size of the CT and IPT group, we combined rudimentary findings from previous randomized trials, which resulted in a mean relapse/recurrence rate of 33% after two years of follow-up, compared to 67% following antidepressant medication and/
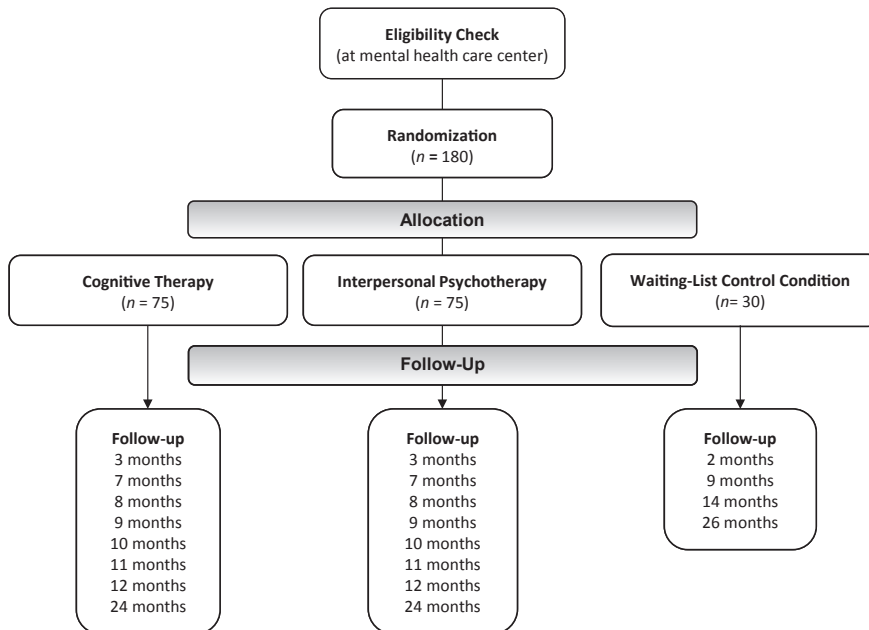
Figure 1. Study design. *Note: a)* the 2-month follow-up assessment in the waiting-list control condition is identical to the baseline assessment. Furthermore, the 9, 14, and 26-month follow-up assessments parallel respectively the 7, 12, and 24 month assessments of the CT and IPT group; *b)* in all three conditions, depression severity was assessed before each therapy session; *c)* as can be seen in the figure, there was no direct assessment at 2-months for the active groups CT and IPT. In order to obtain a measure of depression severity comparable to the 2-month assessment of the WLC condition, we used data that was obtained in the session 2 months after the start of treatment.

or clinical management (Bockting et al., 2005; Fava et al., 2004; Hollon, 2003; Hollon et al., 2005a; Paykel et al., 2005). Based on these results, we assume a difference of 20% in relapse between CT and IPT. Using survival analysis (two-sided, $\alpha = 0.05$, power $= 80\%$), we calculated that 75 patients per treatment condition would suffice to detect this statistically significant difference rate in relapse/recurrence between the two conditions in the long term.

A waiting list condition is included to investigate whether the effects of CT and IPT exceed the outcome in patients who do not receive treatment, a finding that has been demonstrated in numerous previous studies (Dobson, 1989; Gloaguen, Cottraux, Cucherat, & Blackburn, 1998; Parker, Parker, Brotchie, & Stuart, 2006; Robinson, Berman, & Neimeyer, 1990). A power calculation showed that 30 participants in the waiting list condition will be sufficient to detect a statistically significant difference between the two active conditions combined (CT and IPT) and the waiting list after eight weeks.

Recruitment

Participants will be recruited during regular intakes at our clinical site, the Academic Community Mental Health Centre (RIAGG) Maastricht. The SCID-I will be administered to assess current and lifetime episodes of Axis I disorders (First et al., 1997b). Patients meeting the inclusion criteria will be approached for participation. If they agree to participate, the SCID for Axis II disorders (SCID-II) will be administered as well (First, Gibbon, Spitzer, Williams, & Benjamin, 1997a). After inclusion, the participants' GP will be informed about his or her participation in the study. Participants who do not meet the inclusion criteria will be offered suitable alternative treatment options.

Randomization and procedure

Eligible participants will be contacted by telephone for a baseline visit at the university. Preceding this initial visit to the research centre, participants will receive an appointment letter, including study information. This will allow them to consider their study participation well before they agree to participate (approximately one week). Furthermore, to decrease the burden of the first assessment, participants will be asked to already fill out a substantial part of the baseline questionnaires at home. Randomization will take place after informed consent is obtained and the second part of the baseline questionnaires are filled out. The researcher will press the 'assign' button on the computer screen, after which the database randomly allocates the participant to one of the three conditions (CT, IPT, WLC) using block randomization. We pre-stratified groups by first (1 episode) and recurrent (> 1 episodes) depression. The waiting list condition is included because it has been frequently argued in psychotherapy research that the effect in treatment studies cannot be attributed to the intervention(s) that are studied, unless a placebo or no-treatment group is included (Klein, 1990). With regard to the nature of the treatments, blinding of the participants and researchers is not possible.

In reporting the follow-up period of the CT and IPT condition in this study, we distinguish three phases; (a) the 'Treatment Phase' (baseline to 7 months); the period when therapy will be delivered; (b) the 'Trial Follow-Up Phase' (month 8–12) in which depression severity will be measured monthly; and (c) the 'Long-term Follow-Up Phase' (12 to 24 months); the period covered by retrospective assessment. All questionnaires will be administered on a computer. Pre- and post-treatment assessment (baseline and 7 months) will take place at the University. All other follow-up self-report assessments will take place at home via the internet at three, eight, nine, ten, eleven, twelve and twenty-four months after inclusion.

Although the baseline assessment is the same, the follow-up period of the WLC condition is somewhat different as compared to the CT and IPT condition (see Figure 1). Participants in this condition will be measured less often, and at different points in time. This is because their treatment does not start immediately after baseline, but after an 8-week waiting list period. To be able to investigate the effect of this waiting period, and to mark the starting point of the treatment phase, a second identical baseline assessment will take place two months after the initial baseline assessment. As a consequence, the post-treatment assessment in this condition will take place after nine months (instead of after seven months in the CT and IPT condition). Furthermore, patients in this condition will only have two more assessment after their post-treatment measure. These assessments will take place at 14 and 26 months, and are identical to the 12 and 24 months assessment in the CT and IPT condition. To summarize, patients in the WLC condition will have the same baseline assessment as compared to participants in the CT and IPT condition. However, they will have a second identical baseline assessment at two months, and their 9, 14, and 26 month follow-up assessments will parallel respectively the 7, 12, and 24 months assessments of the CT and IPT group (see Figure 1).

In all conditions the BDI-II will be filled out before each session during the treatment phase. After the two-year follow-up period, a modified version of the semi-structured Longitudinal Interval Follow-Up Evaluation (LIFE; Keller et al., 1987) will be administered to retrospectively map out the course of depressive symptoms after treatment termination. To optimize the adherence of self-report assessments, participants will receive an email foregoing each assessment point. If one does not complete the questionnaires within one week, an email-reminder will be sent. If the participant still fails to fill out the questionnaires, a phone call will be made.

Interventions

The interventions will be offered at the Academic RIAGG Maastricht. At this site, CT and IPT are already delivered as treatment as usual. Interventions are written out in a treatment manual, and both treatments will contain 16 to 20 sessions of 45 minutes, depending on the individual progress of patients. Interventions will be delivered by qualified therapists, who were trained by Steven Hollon (CT) and John Markowitz (IPT), experts in the field

of depression treatment. To prevent contamination of treatment conditions, participating therapists will deliver either CT or IPT. Therapists will consult each other on current cases in biweekly consultation meetings. The CT protocol is based on the manual by Beck et al. (1979) and is modified for the current study with a special focus on relapse prevention. IPT is based on the manual of Klerman et al. (1984). CT and IPT differ in terms of target (cognitions and behaviour vs. interpersonal functioning), approach (directive vs. empathic-reflective) and method (homework assignment vs. no assignments; Ablon & Jones, 1999).

All sessions will be videotaped, and a random selection of tapes will be analysed by independent assessors on treatment adherence (content and quality) using the Collaborative Study Psychotherapy Rating Scale – version 6 (CSPRS-6) (Hill, O'Grady, & Elkin, 1992), and on quality of therapeutic alliance using the Observer version of the Working Alliance Inventory-Short (WAI-O-S) (Hatcher & Barends, 2006; Horvath & Greenberg, 1989; Osbourne, 2010; Tichenor & Hill, 1989; Tracey & Kokotovic, 1989).

Instruments

Several instruments will be used to assess the effects of CT and IPT on depressive symptoms and other secondary outcomes. Furthermore, instruments will be used to assess the mechanisms of change and the cost-effectiveness of CT and IPT. An overview of all measurements per assessment is given in Table 1.

Clinical outcome measures

Beck Depression Inventory-II

The Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996; van der Does, 2002a) will be used as the primary outcome measure of depressive severity. The total score is the sum of the 21 items ranging from 0 to 63. Higher scores reflect more severe depressive symptoms. Several studies have shown that the BDI-II is a strong screening measure for depression (Beck et al., 1996; van der Does, 2002a; Whisman, Perez, & Ramel, 2000).

Quick Inventory of Depressive Symptoms

The Quick Inventory of Depressive Symptoms (QIDS; Rush et al., 2003) is a treatment sensitive measure of symptom severity in depression. The 16 items that measure 9 criterion domains of MDD are derived from the 30-item Inventory of Depressive Symptomatology (IDS; Rush, Carmody, & Reimitz, 2000). The QIDS has been shown to have highly acceptable psychometric properties (Trivedi et al., 2004).

Table 1. Overview of instruments per time point in the CT and IPT group.*

Instrument	0**	3	7	8–11	12	24
Clinical Outcome measures						
Beck Depression Inventory II	•	•	•	•	•	•
Quick Inventory of Depressive symptoms	•				•	
Brief Symptom Inventory	•		•		•	•
Diagnostic Interview for Depression		•	•	•	•	•
Beck Hopelessness Scale	•		•			
Longitudinal Interval Follow-up Evaluation						•
Process variables						
Attributional Style questionnaire	•	•	•		•	
Dysfunctional Attitude Scale	•	•	•		•	
Inventory of Interpersonal Problems	•	•	•		•	•
Leiden Index of Depression Severity	•	•	•			
Ruminative Response Scale NL	•	•	•			
Self-Liking and Self-Competence Scale	•	•	•			
Single Category Implicit Association Task	•		•			
Economic Evaluation measures						
Health care use Questionnaire	•	•	•		•	•
Productivity and Disease Questionnaire	•		•		•	•
Work and Social Adjustment Scale	•		•		•	•
Euroqol-6D	•		•		•	•
Rand 36	•		•		•	•
Other measures	•					
Genetic information						
Working Alliance Inventory		•				
Treatment Adherence	•	•	•***			
Observer Rated Working Alliance	•	•	•***			

Note: * Patients in the waiting list condition will receive the same baseline assessment, they will have a second baseline assessment at 2 months follow-up, and their 9, 14, and 26 month follow-up assessments parallel respectively the 7, 12, and 24 month assessments of the CT and IPT group. ** The following general predictors will be administered at baseline as well: demographic variables, family anamnesis, childhood experiences, life-events, reliability and expectations. *** During the treatment phase, all sessions will be videotaped. Only a random selection of tapes will be rated.

Brief Symptom Inventory

The Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983) is a 53-item self-report instrument that will be used to measure general psychological distress. It is derived from the SCL-90-R and several studies have demonstrated it to be an acceptable short alternative of its longer version (de Beurs & Zitman, 2006; Galdón et al., 2008).

Diagnostic Inventory for Depression

The Diagnostic Inventory for Depression (DID) will be used to assess the psychosocial impairment due to depression, and to evaluate subjective quality of life (de Graaf & Huibers, 2009; Zimmerman, Sheeran, & Young, 2004). Psychometric analysis shows good reliability and validity (de Graaf & Huibers, 2009).

Longitudinal Interval Follow-up Evaluation

A modified version of the semi-structured Longitudinal Interval Follow-up Evaluation (LIFE; Keller et al., 1987) will be used to map out the course of depressive symptoms in the long-term Follow-Up Phase (12–24 months). This will be administered at the end of the follow-up (24 months). The interview contains the MDD section of the original interview, and several questions on general functioning, working status, relations to others, and treatment- and study experiences. The LIFE has shown to be a reliable and valid instrument for characterizing the course of several mental disorders examined retrospectively over the period of one year (Warshaw, Dyck, Allsworth, Stout, & Keller, 2001; Warshaw, Keller, & Stout, 1994).

Process (or mechanism) variables

Dysfunctional Attitudes

The Dysfunctional Attitude scale version A (DAS-A; de Graaf, Roelofs, & Huibers, 2009; Weissman & Beck, 1978) is a self-report scale designed to measure patterns of negative and absolute thinking in depressed patients. Respondents need to report whether or not they agree with a series of dysfunctional assumptions on a seven point Likert-Scale. With its good internal consistency and validity, it appears to be a valid measure of dysfunctional cognitions in depressed patients (Beck, Brown, Steer, & Weissmann, 1991; Nelson, Stern, & Cicchetti, 1992; Oliver & Baumgart, 1985; Power et al., 1994).

Attributional Style

Attributional style will be measured using the Attributional Style Questionnaire (ASQ; Cohen, van den Bout, Kramer, van Vliet, 1986; Peterson et al., 1982). In the ASQ, participants

2

receive 12 hypothetical events and are asked to imagine themselves in that scenario. Subsequently they have to decide what the major cause of the situation would be if it happened to themselves, and rate the cause along a 7-point Likert-scale on three dimensions: Personal, Permanent and Pervasive. Depressive symptoms are associated with an attributional style in which negative events are attributed to internal, stable, and global causes, whereas positive events are explained as being external, unstable and specific (van der Molen, Perreijn, & van den Hout, 1997). The ASQ has shown to have good psychometric properties (Cohen et al., 1986).

Implicit measure of self-esteem

A variant of the Single Category Implicit Association Test (Sc-IAT; Karpinski & Steinman, 2006) will be administered to obtain an implicit measure of associations with the self (self-esteem). For the current study, we adapted the original Sc-IAT paradigm to specifically measure feelings of helplessness, unlovability and worthlessness, which are considered the main themes of core beliefs in depression. The test consists of 3 blocks of trials and starts with a practice block. In this block, positive and negative words are presented (all words are related to the specific core beliefs), and the task for participants is to categorize these stimuli as such by pressing the appropriate key (left or right) without making too many errors. After the practice block, there are two test blocks. In one critical block, stimuli representing the self ('me-trials') share a response key with positive stimuli, while the other key needs to be pressed for negative stimuli. In the other critical block, the target key assignment for me-trials is switched. Stimuli representing the self now share a response key with the negative stimuli. Each block is preceded by a set of instructions concerning the dimension(s) of the categorisation task and the appropriate key response. The rationale behind the Sc-IAT is that the response time will be faster when two associated concepts are mapped together on the same key, compared to when two less compatible concepts are mapped together. For example, in people with a positive sense of self-esteem, when 'valuable' and 'me' meaning words are mapped to the same key, the response time will be faster than for the non-compatible combination ('worthless' and 'me'). The Sc-IAT effect will be calculated using the improved D-Algorithm (Greenwald, Nosek, & Banaji, 2003). In comparison with implicit measures obtained via other measurement procedures, the Sc-IAT shows good psychometric qualities (Bosson, Swann, & Pennebaker, 2000; Glashouwer & de Jong, 2008; Nosek, Greenwald, & Banaji, 2005; Olson & Fazio, 2003).

Hopelessness

The Beck Hopelessness Scale (BHS; Beck, Weissman, Lester, & Trexler, 1974) is a self-report instrument that determines the extent of positive and negative beliefs about the future over the previous 7 days using 20 true-false statements. All statements are scored 0 or 1

with the total being calculated by summing the pessimistic responses for the items. The BHS has been shown to have good psychometric properties for application in clinical samples (Beck & Steer, 1988; Dyce, 1996; Young, Halper, Clark, & Scheftner, 1992).

Interpersonal Problems

The Inventory of Interpersonal Problems (IIP; Horowitz, Rosenberg, Baer, Ureño, & Villaseñor, 1988) is a 127-item self-report measure describing the types of problems that people experience in their relationships with others, and the level of distress associated with them. Items are divided in two sections: participants are asked to rate problematic behaviour that is hard for them to do (78 items) and behaviour that they are doing too much (49 items). The IIP provides a total score and scores on six or eight subscales. Higher scores indicate more interpersonal problems. Psychometric properties of the IIP were found to be good (Horowitz et al., 1988).

Cognitive Reactivity

During a depressive episode, an association is formed between depressed mood and dysfunctional thinking. This association may remain intact even when the depressive episode remits. The strength of this association is called cognitive reactivity (de Beurs, 2009). The Leiden Index of Depression Sensitivity (LEIDS; van der Does, 2002b) is used to measure this cognitive reactivity on six subscales. Psychometric qualities of this instrument are found to be good (van der Does, 2002b).

Rumination

The Ruminative Response Scale (RRS; Nolen-Hoeksema & Morrow, 1991; Raes, Hermans, & Eelen, 2003) is a 22-item questionnaire that will be used to detect the ruminative responses to depressed mood that are focused on self, symptoms, or possible consequences and causes of mood. People are asked to indicate how often they think about certain things on a four point Likert-scale. The RRS shows good reliability and satisfactory validity (Nolen-Hoeksema & Morrow, 1991; Raes et al., 2003).

Self-Esteem

The Self-Liking and Self-Competence Scale Revised (SLCS-R; Tafarodi & Swann, 2001; Vandromme, Hermans, Spruyt, & Eelen, 2007) is a self-report questionnaire constructed to measure Self-Competence and Self-Liking, two dimensions of self-esteem. The SLSC-R contains eight items for each of the two dimensions. Items are rated on a 5-point Likert scale. The subscales scores can range from 8 to 40, with higher scores indicating higher self-competence or higher self-liking. Psychometric properties were found to be good (Tafarodi & Swann, 2001).

2

Economic Evaluation

Costs

A rudimentary analysis of the cost-effectiveness of CT and IPT will be conducted from a societal perspective. We will distinguish three cost categories: health care sector costs, costs for the patient and family, and productivity cost (Drummond, Schulpher, Torrance, O'Brien, & Stoddart, 2005). Healthcare costs and patient and family costs will be measured using a periodic retrospect health care consumption questionnaire designed by de Graaf et al. (2008) for use in the field of depression research. This questionnaire is based on an existing cost diary (Goossens, Rutten-van Mölken, Vlaeyen, & van der Linden, 2000) and retrospective cost questionnaires (Hakkart–van Roijen, 2002; van Asselt et al., 2008). Containing 52 items, it measures the number and content of contacts with GPs, specialists, paramedics, alternative care, psychological care, hospital visits, medication, and self-help over a period of three months. To measure production losses the patient modules (A-E) of the productivity and disease questionnaire (PRODISQ; Koopmanschap et al., 2004) will be used.

Quality of Life measures

To measure the experiential impact of the disorder, the Work and Social Adjustment Scale (W&SAS; Mundt, Marks, Greist, & Shear, 2002) will be used. This 5-item self-report scale examines to what extent the disorder impairs a person's everyday functioning. Items are rated on a 9 point Likert-scale, with higher scores indicating more severe impairment. Psychometric properties of the W&SAS are found to be good (Mataix-Cols et al., 2005; Mundt et al., 2002). Furthermore, to measure generic quality of life of the patients on the basis of societal appreciation of health condition, the EuroQol (EuroQolGroup, 1990) will be used. To measure quality of life in terms of impairments due to physical and mental health status, we will use the RAND-36 (van der Zee & Sanderman, 1993). This 36 item questionnaire measures physical and social functioning, role restriction due to physical or emotional problems, mental health, energy, pain, and general health perception. The RAND-36 has shown to be a reliable, valid and sensitive measure (van der Zee & Sanderman, 1993).

Other measures

Working Alliance

The past decades, it has become clear that therapeutic alliance is an essential element of the therapeutic process (Martin, Garske, & Davis, 2000). The stronger the collaborative and affective bond between patient and therapist, the larger the therapeutic change (Horvath & Bedi, 2002). To measure this bond, the Working Alliance Inventory (WAI; Horvath & Greenberg, 1989; Vervaeke & Vertommen, 1996) will be used. The question-

naire consists of 36 items rated on a 5-point Likert scale and will be filled out by both patient and therapist. Internal consistency of this instrument was found to be high. Furthermore, since all therapy sessions will be videotaped, the quality of the therapeutic alliance will also be assessed by independent observers using the Observer rated version of the Working Alliance Inventory-Short (WAI-O-S) (Tichenor & Hill, 1989; Tracey & Kokotovic, 1989).

Genetics

There is evidence that the same genetic factors that appear to moderate the efficacy of different classes of antidepressants are also involved in aspects of cognitive functioning (Goldberg & Weinberger, 2004). The same polymorphisms are also associated with reactions on aversive stimuli (Domschke et al., 2006; Pezawas et al., 2005; Smolka et al., 2005). Because changes in regulation of emotions and cognitive processes are core elements in therapeutic approaches such as CT and IPT, we expect that the efficacy of these therapies is influenced by genetic variations in polymorphisms. Therefore, we will obtain genetic material at baseline (buccal cotton swabs).

Analyses

Data-analysis will include intention-to-treat analyses, per-protocol analyses, change analyses and prognostic analyses. Analyses will include elementary head-to-head comparisons of the intervention groups as well as more complicated multivariate analyses (e.g., individual time series analysis, logistic regression analysis, multilevel analyses). Furthermore, in ancillary analysis, subgroups of the study sample according to symptom severity and DSM-IV classification will be examined for differential effects. In addition, we will determine the proportion of patients that show reliable and clinically significant improvement on the outcome measures. Our calculations will be based on the method of Jacobson and Truax (1991) which prescribes that Clinical Improvement (CI) is based on both Reliable Change (RC), the extent to which the pre-to-post-difference score is reliable; and on Clinical Significant change (CSC), the extent to which post-treatment scores are clinically meaningful (Evans, Margison, & Barkham, 1998). We will use chi-square tests to test the frequency differences in the RC, CSC, and CI between the three intervention groups.

Relapse (episode of MDD after remission) and recurrence (episode of MDD after recovery) in the course of follow-up (12 and 24 months) will be assessed using survival analysis (Cox proportional hazards regression). Furthermore, using multilevel mediation models, we will investigate various potential treatment mediators (psychological processes responsible for therapeutic change) to identify mechanisms of change. We will use the revised and expanded McArthur guidelines with a central focus on the temporal

aspect (Kazdin, 2007, 2009). Furthermore, we will use structural equation modelling to test the relative strength of the factors involved (path analysis), an approach already successfully applied by Oei and colleagues (Oei et al., 2006).

Discussion

We presented the protocol of a study assessing the effectiveness, mechanisms of change and prevention of relapse/recurrence of CT and IPT for depression. While it is well known that both CT and IPT are effective treatments for depression, it is not entirely clear yet whether one therapy is superior to the other, especially in the long term. Furthermore, since response rates of both therapies are insufficient and the risk of relapse and recurrence is high, the challenge in contemporary depression research is the improvement of treatments to increase response rates and prevent relapse and recurrence. Studying the mechanisms of change in treatment might be a first step in improving every day clinical practice. Although in the past decades attention for mechanisms of change in psychotherapy has grown, this type of research is still in its infancy. It may be clear that there is a need for renewed methodologically well-considered research. With the current study, we hope to provide new insights in the issues stated above.

Methodological Considerations

The current study has several strengths including randomization of patients to three conditions (CT, IPT and a waiting list condition). The used design (RCT) is not only the standard for the evaluation of effectiveness of psychiatric treatments, but is also very valuable in studying mechanisms of therapeutic change (Haaga & Stiles, 2000; Nock, 2007). Furthermore, we follow the recommendation of Kazdin to measure multiple potential mediators simultaneously and to include multiple follow-up measures at various time-points throughout a 2-year follow-up period. We assess both mediators and outcomes before and after therapy. This repeated measures design will provide a unique opportunity to evaluate whether change in cognitions occurs in advance of, and is related to, reduction in symptoms. In addition, to our knowledge, we are the first to compare CT with IPT head-to-head for relapse prevention and assess the underlying mechanisms of change using both explicit and implicit measures in a treatment study of depression to examine the level of therapeutic changes. To conclude, we will use state-of-the art statistical techniques to analyse temporality, causality and mechanisms of change.

However, the current study also has a number of limitations. In spite of the fact that we consider many potential mechanisms, there probably will be important (latent) processes that are not assessed in the current study. Even though we will use analys-

ing techniques to prevent these biases as much as possible, there still is a chance that results will be contributed to the measured variables, whereas they are actually caused by other latent variables.

Conclusion

The current study will compare CT and IPT head-to-head in terms of effectiveness and the prevention of relapse. Furthermore, in order to investigate mechanisms of change, we will investigate potential mediators of CT for depression and test their specificity in comparison to IPT and vice versa. By including assessments on multiple time points in both the treatment- and Follow-Up Phase, we try to investigate temporal relationships between change in potential mediators and outcome measures that are needed to identify causal pathways of therapeutic change. We will assess both specific and common treatment factors using both explicit and implicit measures. With this we hope to provide new insights in the mechanisms of change of CT and IPT for depression and hereby contribute to the improvement of mental health care for adults suffering from depression.

Abbreviations

ASQ, Attributional Style Questionnaire; BDI-II, Beck Depression Inventory, second edition; BHS, Beck Hopelessness Scale; BSI, Brief Symptom Inventory; CI, Clinical Improvement; CSC, Clinical Significant Change; CSPRS-6, Collaborative Study Psychotherapy Rating Scale – version 6; CT, Cognitive Therapy; DAS-A, Dysfunctional Attitude scale version a; DID, Diagnostic Inventory for Depression; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, fourth edition; IDS, Inventory of Depressive Symptomatology; IIP, Inventory of Interpersonal Problems; IPT, Interpersonal Psychotherapy; LEIDS, Leiden Index of Depression Sensitivity; LIFE, Longitudinal Interval Follow-Up Evaluation; MDD, Major

Depressive Disorder; PRODISQ, Productivity and Disease questionnaire; QIDS, Quick Inventory of Depressive Symptoms; RC, Reliable Change; RCT, Randomized Controlled Trial; RIAGG, Regional Academic Community Mental Health Centre; RRS, Ruminative Response Scale; Sc-IAT, Single Category Implicit Association Test; SCID-I, Structural Clinical Interview for DSM-IV Axis I disorders; SCID-II, Structural Clinical Interview for DSM-IV Axis II disorders; SLSC-R, Self Liking and Self Competence Scale Revised; W&SAS, Work and Social Adjustment Scale; WAI, Working Alliance Inventory; WAI-O-S, Observer rated version of the Working Alliance Inventory Short.

2

References

Ablon, J. S., & Jones, E. E. (1999). Psychotherapy process in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 67*(1), 64–75.

Asselt van, A. D. I., Dirksen, C. D., Arntz, A., Giesen-Bloo, J. H., van Dyck, R., Spinhoven, P., . . . Severens, J. L. (2008). Out-patient psychotherapy for borderline personality disorder: Cost-effectiveness of schema-focused therapy v. transference-focused psychotherapy. *British Journal of Psychiatry, 192*(6), 450–457.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.

Beck, A. T., Brown, G. K., Steer, R. A., & Weissmann, A. N. (1991). Factor Analysis of the Dysfunctional Attitude Scale in a Clinical Population. *Journal of Consulting and Clinical Psychology, 3*(3), 478–483.

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

Beck, A. T., Steer, R., & Brown, G. K. (1996). *Beck Depression Inventory II: Manual*. Boston, MA: Hartcourt Brace.

Beck, A. T., & Steer, R. A. (1988). *Manual for the Beck Hopelessness Scale*. San Antonio, TX: Psychological Corporation.

Beck, A. T., Weissman, A., Lester, D., & Trexler, L. (1974). The measurement of pessimism: The Hopelessness Scale. *Journal of Consulting and Clinical Psychology, 42*(6), 861–865.

Beers, M. H., Fletscher, A. J., Jones, T. V., Porter, R., Berkwits, M., & Kaplan, J. L. (2005). *Merck Manual*. Houten: Bohn Stafleu van Loghum.

Beurs de, D. P. (2009). Further validation of the Leiden Index of Depression Sensitivity *Tijdschrift voor Psychiatrie: samenvattingen 37ste voorjaarscongres, 51*(3), 14.

Beurs de, E., & Zitman, F. (2006). De Brief Symptom Inventory (BSI). De betrouwbaarheid en validiteit van een handzaam alternatief voor de SCL-90. *Maandblad Geestelijke Volksgezondheid, 61*, 120–141.

Bockting, C. L. H., Schene, A. H., Spinhoven, P., Koeter, M. W. J., Wouters, L. F., Huyser, J., & Kamphuis, J. H. (2005). Preventing Relapse/Recurrence in Recurrent Depression With Cognitive Therapy: A Randomized Controlled Trial. *Journal of Consulting and Clinical Psychology, 73*(4), 647–657.

Bosson, J. K., Swann, W. B., Jr., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79*(4), 631–643.

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.

Cohen, L., van den Bout, J., Kramer, W., & van Vliet, T. (1986). A Dutch Attributional Style Questionnaire: Psychometric Properties and Findings of Some Dutch-American Differences. *Cognitive therapy and research, 10*(6), 665–669.

Coleman, D., Cole, D., & Wuest, L. (2010). Cognitive and Psychodynamic Mechanisms of Change in Treated and Untreated Depression. *Journal of Clinical Psychology, 66*(3), 215–228.

2

Cuijpers, P., Geraedts, A. S., van Oppen, P., Andersson, G., Markowitz, J. C., & van Straten, A. (2011). Interpersonal Psychotherapy for Depression: A Meta-Analysis. *American Journal of Psychiatry, 168*, 581–592.

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology, 76*(6), 909–922.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*(3), 347–368.

Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine, 13*(3), 595–605.

DeRubeis, R. J., & Crits-Christoph, P. (1998). Empirically supported individual and group psychological treatments for adult mental disorders. *Journal of Consulting and Clinical Psychology, 66*(1), 37–52.

DeRubeis, R. J., Evans, M. D., Hollon, S. D., Garvey, M. J., Grove, W. M., & Tuason, V. B. (1990). How Does Cognitive Therapy Work? Cognitive Change and Symptom Change in Cogntive Therapy and Pharmacotherapy for Depression. *Journal of Consulting and Clinical Psychology, 58*(6), 862–869.

Dobson, K. S. (1989). A meta-analysis of the efficacy of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 57*(3), 414–419.

Does van der, W. (2002a). *BDI-II-NL: Handleiding; De Nederlandse versie van de Beck Depressie Inventory. Second edition*. Enschede: Ipskamp.

Does van der, W. (2002b). Cognitive reactivity to sad mood: structure and validity of a new measure. *Behaviour Research and Therapy, 40*, 105–120.

Domschke, K., Braun, M., Ohrmann, P., Suslow, T., Kugel, H., Bauer, J., . . . Deckert, J. (2006). Association of the functional –1019C/G 5-HT1A polymorphism with prefrontal cortex and amygdala activation measured with 3 T fMRI in panic disorder. *International Journal of Neuropsychopharmacology, 9*(3), 349–355.

Dorrepaal, E., van Nieuwenhuizen, C., Schene, A., & de Haan, R. (1998). De effectiviteit van cognitieve en interpersoonlijke therapie bij depressiebehandeling: een meta-analyse. *Tijdschrift voor Psychiatrie, 40*(1), 27–39.

Drummond, M. F., Schulpher, M. J., Torrance, G. W., O'Brien, B., & Stoddart, G. L. (2005). *Methods for the Economic Evaluation of Health Care Programmes, 3rd edition*. New York, NY: Oxford University Press.

Dyce, J. A. (1996). Factor structure of the Beck Hopelessness Scale. *Journal of Clinical Psychology, 52*(5), 555–558.

Elkin, I., Shea, M. T., Watkins, J. T., & Imber, S. D. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry, 46*(11), 971–982.

Emmelkamp, P. M. G. (1992). Cognitieve gedragstherapeutische interventies bij depressie. In K. v. d. Velden (Ed.), *Directieve therapie 4*. Houten: Bohn Stafleu van Loghum.

EuroQolGroup. (1990). EuroQol–a new facility for the measurement of health-related quality of life. *Health Policy, 16*, 199-208.

Evans, C., Margison, R., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental health, 1*, 70–72.

Fava, G. A., Ruini, C., Rafanelli, C., Finos, L., Conti, S., & Grandi, S. (2004). Six-Year Outcome of Cognitive Behavior Therapy for Prevention of Recurrent Depression. *The American Journal of Psychiatry, 161*(10), 1872–1876.

First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997a). *Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)*. Washington, D.C.: American Psychiatric Press.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997b). *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I)*. New York, NY: Biometrics Research Department New York State Psychiatric Institute.

Frank, E., Kupfer, D. J., Buysse, D. J., Swartz, H. A., Pilkonis, P. A., Houck, P. R., . . . Stapf, D. M. (2007). Randomized trial of weekly, twice-monthly, and monthly interpersonal psychotherapy as maintenance treatment for women with recurrent depression. *The American Journal of Psychiatry, 164*(5), 761–767.

Frank, E., Kupfer, D. J., Perel, J. M., & Cornes, C. (1990). Three-year outcomes for maintenance therapies in recurrent depression. *Archives of General Psychiatry, 47*(12), 1093–1099.

Frank, E., Kupfer, D. J., Wagner, E. F., & McEachran, A. B. (1991). Efficacy of interpersonal psychotherapy as a maintenance treatment of recurrent depression: Contributing factors. *Archives of General Psychiatry, 48*(12), 1053–1059.

Furlong, M., & Oei, T. P. S. (2002). Changes to automatic thoughts and dysfuntional attitudes in group CBT for depression. *Behavioural and Cognitive Psychotherapy, 30*(3), 351–360.

Galdón, M., Durá, E., Ferrando, M., Murgui, S., Pérez, S., & Ibañez, E. (2008). Psychometric properties of the Brief Symptom Inventory-18 in a Spanish breast cancer sample. *Journal of Psychometric Research, 65*, 533–539.

Garrat, G., Ingram, R. E., Rand, K. L., & Sawalani, G. (2007). Cognitive Processes in Cognitive Therapy: Evaluation of the Mechanisms of Change in the Treatment of Depression. *Clinical Psychology, 14*, 224–239.

Glashouwer, K. A., & de Jong, P. J. (2008). Impliciete persoonlijkheidstrekken en psychopathologie: Achtergrond, huidige inzichten en perspectieven. *Gedragstherapie, 41*(2), 121–134.

Gloaguen, V., Cottraux, J., Cucherat, M., & Blackburn, I. (1998). A meta-analysis of the effects of cognitive therapy in depressed patients. *Journal of Affective Disorders, 49*(1), 59–72.

Goldberg, T. E., & Weinberger, D. R. (2004). Genes and the parsing of cognitive processes. *Trends in Cognitive Sciences, 8*(7), 325–335.

Goossens, M. E. J. B., Rutten-van Mölken, M. P. M. H., Vlaeyen, J. W. S., & van der Linden, S. M. J. P. (2000). The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. *Journal of Clinical Epidemiology, 53*, 688–695.

Graaf de, L. E., Gerhards, S., Evers, S., Arntz, A., Riper, H., Severens, J., . . . Huibers, M. (2008). Clinical and cost-effectiveness of computerised cognitive behavioural therapy for depression in primary care: Design of a randomised trial. *BMC Public Health, 8*(1), 224.

Graaf de, L. E., & Huibers, M. J. H. (2009). Introductie van de Diagnostic Inventory for Depression in Nederland. *Tijdschrift voor Psychiatrie, 51*, 675–686.

Graaf de, L. E., Roelofs, J., & Huibers, M. J. H. (2009). Measuring dysfunctional attitudes in the general population: the DAS-A revised. *Cognitive therapy and research, 33*, 345–355.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). 'Understanding and using the Implicit Association Test: I. An improved scoring algorithm': Correction to Greenwald et al. (2003). *Journal of Personality and Social Psychology, 85*(3).

Haaga, D. A. F., & Stiles, W. B. (2000). Randomized clinical trials in psychotherapy research:Methodology, design, and evaluation. In C. R. Snyder & R. E. Ingram (Eds.), *Handbook of psychological change: Psychotherapy processes and practices for the 21st Century* (pp. 14–39). New York: Wiley.

Hakkaart-van Roijen, L. (2002). *Trimbos/iMTA questionnaire for Costs associated with Psychiatric Illness (TiC-P)*. Rotterdam: Institute for medical Technology Assessment, Erasmus University.

Hatcher, R. L., & Barends, A. W. (2006). How a return to theory could help alliance research. *Psychotherapy: Theory/Research/Practice/Training, 43*(3), 292–299.

Hill, C. E., O'Grady, K. E., & Elkin, I. (1992). Applying the Collaborative Study Psychotherapy Rating Scale to rate therapist adherence in cognitive-behavior therapy, interpersonal therapy, and clinical management. *Journal of Consulting and Clinical Psychology, 60*(1), 73–79.

Hollon, S. D. (2003). Does cognitive therapy have an enduring effect? *Cognitive therapy and research, 27*(1), 71–75.

Hollon, S. D., DeRubeis, R. J., Shelton, R. C., Amsterdam, J. D., Salomon, R. M., O'Reardon, J. P., . . . Gallop, R. (2005a). Prevention of Relapse Following Cognitive Therapy vs Medications in Moderate to Severe Depression. *Archives of General Psychiatry, 62*(4), 417–422.

Hollon, S. D., Jarrett, R. B., Nierenberg, A. A., Thase, M. E., Trivedi, M., & Rush, A. J. (2005b). Psychotherapy and Medication in the Treatment of Adult and Geriatric Depression: Which Monotherapy or Combined Treatment? *Journal of Clinical Psychiatry, 66*(4), 455–468.

Hollon, S. D., & Ponniah, K. (2010). A review of empirically supported psychological therapies for mood disorders in adults. *Depression and Anxiety, 27*(10), 891–932.

Hollon, S. D., Thase, M. E., & Markowitz, J. C. (2002). Treatment and prevention of depression. *Psychological Science and Public Interest, 3*, 39–77.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G., & Villaseñor, V. S. (1988). Inventory of Interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*, 885–892.

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

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*(2), 223–233.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to define meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12–19.

Jarret, R. B., Kraft, D., Doyle, J., Foster, B. M., Eaves, C. G., & Silver, P. C. (2001). Preventing recurrent depression using cognitive therapy with and without a continuation phase; a randomised clinical trial. *Archives of General Psychiatry, 58*, 381–388.

Karpinski, A., & Steinman, R. B. (2006). The Single Category Implicit Association Test as a measure of Implicit Social Cognition. *Journal of Personality and Social Psychology, 91*, 16–32.

Kazdin, A. E. (2007). Mediators and Mechanisms of Change in Psychotherapy Research. *Annual Review of Clinical Psychology, 3*, 1–27.

Kazdin, A. E. (2009). Understanding how and why psychotherapy leads to change. *Psychotherapy Research, 19*(4–5), 418-428.

Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, 44*(8), 1116–1129.

Keller, M. B., & Boland, R. J. (1998). Implications of failing to achieve successful long-term maintenance treatment of recurrent unipolar major depression. *Biological Psychiatry, 44*(5), 348–360.

Keller, M. B., Lavori, P. W., Friedman, B., Nielsen, E., Endicott, J., McDonald-Scott, P., & Andreasen, N. C. (1987). The Longitudinal Interval Follow-up Evaluation: A Comprehensive Method for Assessing Outcome in Prospective Longitudinal Studies. *Archives of General Psychiatry, 44*, 540–548.

Klein, D. F. (1990). NIMH collaborative research on treatment of depression. *Archives of General Psychiatry, 47*, 682–684.

Klerman, G. L., & Weissman, M. M. (1993). *New applications of interpersonal psychotherapy*. Washington, D.C.: American Psychiatric Association.

Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. S. (1984). *Interpersonal psychotherapy for depression*. New York, NY: Basis Books.

Kolko, D. J., Brent, D. A., Baugher, M., Bridge, J., & Birmaher, B. (2000). Cognitive and family therapies for adolescent depression: Treatment specificity, mediation, and moderation. *Journal of Consulting and Clinical Psychology, 68*(4), 603–614.

Koopmanschap, M., Meerding, W., Evers, S., Severens, J., Burdorf, A., & Brouwer, W. (2004). *Handleiding voor het gebruik van PRODISQ. Een modulaire vragenlijst over de relatie tussen ziekte en productiviteitskosten. Toepasbaar bij economische evaluaties van gezondheidszorgprogramma's voor patiënten en werknemers*. Rotterdam/Maastricht: Erasmus Universiteit Rotterdam/Erasmus Medisch Centrum/Universiteit Maastricht.

Kraemer, H. C., Stice, E., Kazdin, A. E., Offord, D., & Kupfer, D. J. (2001). How Do Risk Factors Work Together? Mediators, Moderators, and Independent, Overlapping and Proxy Risk Factors. *American Journal of Psychiatry, 158*(6), 848–883.

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*, 877–883.

Kwon, S., & Oei, T. P. S. (2003). Cognitive change processes in a group cognitive behavior therapy of depression. *Journal of Behavior Therapy and Experimental Psychiatry, 34*, 73–85.

Londen van, L., Molenaar, R. P. G., Goekoop, J. G., Zwinderman, A. H., & Rooijmans, H. G. M. (1998). Three- to 5-year prospective follow-up of outcome in major depression. *Psychological Medicine, 28*(03), 731–735.

Longmore, R. J., & Worrell, M. (2007). Do we need to challenge thoughts in cognitive behavior therapy? *Clinical Psychology Review, 27*(2), 173–187.

Luty, S. E., Carter, J. D., McKenzie, J. M., Rae, A. M., Frampton, C. M. A., Mulder, R. T., & Joyce, P. R. (2007). Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *British Journal of Psychiatry, 190*, 496–502.

2

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *Journal of Consulting and Clinical Psychology, 68*, 438–450.

Mataix-Cols, D., Cowley, A. J., Hankins, M., Schneider, A., Bachofen, M., Kenwright, M., . . . Marks, I. M. (2005). Reliability and validity of the Work and Social Adjustment Scale in phobic disorders. *Comprehensive Psychiatry, 46*, 223- 228.

Mathers, C., Ma Fat, D., & Boerma, J. T. (2008). *The global burden of disease: 2004 update*. Geneva: World Health Organization.

Molen van der, H. T., Perreijn, S., & van den Hout, M. A. (1997). *Klinische psychologie: theorieën en psychopathologie*. Groningen: Wolters-Noordhoff.

Mundt, J. C., Marks, I. M., Greist, J. H., & Shear, K. (2002). The Work and Social Adjustment Scale: A simple accurate measure of impairment in functioning. *British Journal of Psychiatry, 180*, 461–464.

Murphy, R., Cooper, Z., Hollon, S. D., & Fairburn, C. G. (2009). How do psychological treatments work? Investigating mediators of change. *Behavioural Research Therapy, 47*(1), 1–5.

Nelson, L. D., Stern, S. L., & Cicchetti, D. V. (1992). The Dysfunctional attitude scale, how well can it measure depressive thinking? *journal of Psychopathology and behavioural assessment, 4*, 217–223.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259.

Nock, M. K. (2007). Conceptual and design essentials for evaluating mechanisms of change. *Alcoholism: Clinical and Experimental Research, 31*(Suppl 3), 4S-12S.

Nolen-Hoeksema, S., & Morrow, J. (1991). A prospective study of depression and posttraumatic stress symptoms after a natural disaster: The 1989 Loma Prieta earthquake. *Journal of Personality and Social Psychology, 61*(1), 115–121.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and Using the Implicit Association Test: II. Method Variables and Construct Validity. *Personality and Social Psychology Bulletin, 31*(2), 166–180.

Oei, T. P. S., Bullbeck, K., & Campbell, J. M. (2006). Cognitive change process during group cognitive behaviour therapy for depression. *Journal of Affective Disorders, 92*, 231–241.

Oei, T. P. S., & Shuttlewood, G. J. (1996). Specific and Nonspecific factors in psychotherapy; a case of cognitive therapy for depression. *Clinical Psychology Review, 16*(2), 83–103.

Oliver, M., & Baumgart, E. P. (1985). The dysfunctional attitude scale: psychometric properties and relation to depression in an unselected adult population. *Cognitive therapy and research, 9*, 161–167.

Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science, 14*(6), 636–639.

Ormel, J., Oldehinkel, T., Brilman, E., & van den Brink, W. (1993). Outcome of Depression and Anxiety in Primary Care: A Three-Wave 31/2-Year Study of Psychopathology and Disability. *Archives of General Psychiatry, 50*(10), 759–766.

Osbourne, C. A. (2010). *The Role of Alliance and Symptomatic Change Within Cognitive Behaviour Therapy for Depression (doctoral thesis)*. Massey University, Albany.

Parker, G., Parker, I., Brotchie, H., & Stuart, S. (2006). Interpersonal psychotherapy for depression? The need to define its ecological niche. *Journal of Affective Disorders, 95*(1–3), 1–11.

Paykel, E. S., Scott, J., Cornwall, P. L., Abbott, R., Crane, C., Pope, M., & Johnson, A. L. (2005). Duration of relapse prevention after cognitive therapy in residual depression: Follow-up of controlled trial. *Psychological Medicine: A Journal of Research in Psychiatry and the Allied Sciences, 35*(1), 59–68.

Peterson, C., Semmel, A., von Baeyer, C., Abramson, L. Y., Metalsky, G. I., & Seligman, M. E. P. (1982). The attributional style questionnaire. *Cognitive therapy and research, 6*, 287–300.

Pezawas, L., Meyer-Lindenberg, A., Drabant, E. M., Verchinski, B. A., Munoz, K. E., Kolachana, B. S., . . . Weinberger, D. R. (2005). 5-HTTLPR polymorphism impacts human cingulate-amygdala interactions: A genetic susceptibility mechanism for depression. *Nature Neuroscience, 8*(6), 828–834.

Power, M. J., Katz, R., McGuffin, P., Duggan, C. F., Lam, D., & Beck, A. T. (1994). The Dysfunctional Attitude Scale (DAS): A Comparison of Forms A and B and Proposals for a New Subscaled Version. *Journal of Research in Personality, 28*(3), 263–276.

Quilty, L. C., McBride, C., & Bagby, R. M. (2008). Evidence for the cognitive mediational model of cognitive behavioural therapy for depression. *Psychological Medicine, 38*, 1531–1541.

Raes, F., Hermans, D., & Eelen, P. (2003). Kort instrumenteel De Nederlandstalige versie van de Ruminative Response Scale (RRS-NL) en de Rumination on Sadness Scale (RSS-NL). *Gedragstherapie, 36*(2), 97–104.

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.

Roefs, A., Huijding, J., Smulders, F. T. Y., MacLeod, C. M., de Jong, P. J., Wiers, R. W., & Jansen, A. T. M. (2011). Implicit measures of association in psychopathology research. *Psychological Bulletin, 137*(1), 149–193.

Roelofs, J., & Muris, P. (2005). Psychological treatments of depression. In E. J. L. Griez, D. Faravelli, D. J. Nutt & J. Zohar (Eds.), *Mood disorders: clinical management and research issues.* Chichester: John Wiley & Sons.

Rush, A. J., Carmody, T., & Reimitz, P. E. (2000). The Inventory of Depressive Symptomatology (IDS) : clinician (IDS-C) and Self-Report (IDS-SR) ratings of depressive symptoms. *International Journal of Methods in Psychiatric Research, 9*, 45–59.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., . . . Keller, M. B. (2003). The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*(5), 573–583.

Smolka, M. N., Schumann, G., Wrase, J., Grusser, S. M., Flor, H., & Mann, K. (2005). Catechol-O-Methyltransferase val158met Genotype Affects Processing of Emotional Stimuli in the Amygdala and Prefrontal Cortex. *Journal of Neuroscience, 25*, 145–159.

Strunk, D. R., & DeRubeis, R. J. (2001). Cognitive Therapy for Depression: A Review of Its Efficacy. *Journal of Cognitive Psychotherapy, 15*, 289–297.

Tafarodi, R. W., & Swann, W. B. (2001). Two-dimensional self-esteem: Theory and measurement. *Personality and Individual Differences, 31*(5), 653–673.

Teasdale, J. D., Segal, Z. V., Williams, J. M., Ridgeway, V. A., Soulsby, J. M., & Lau, M. A. (2000). Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. *Journal of Consulting and Clinical Psychology, 68*, 615–623.
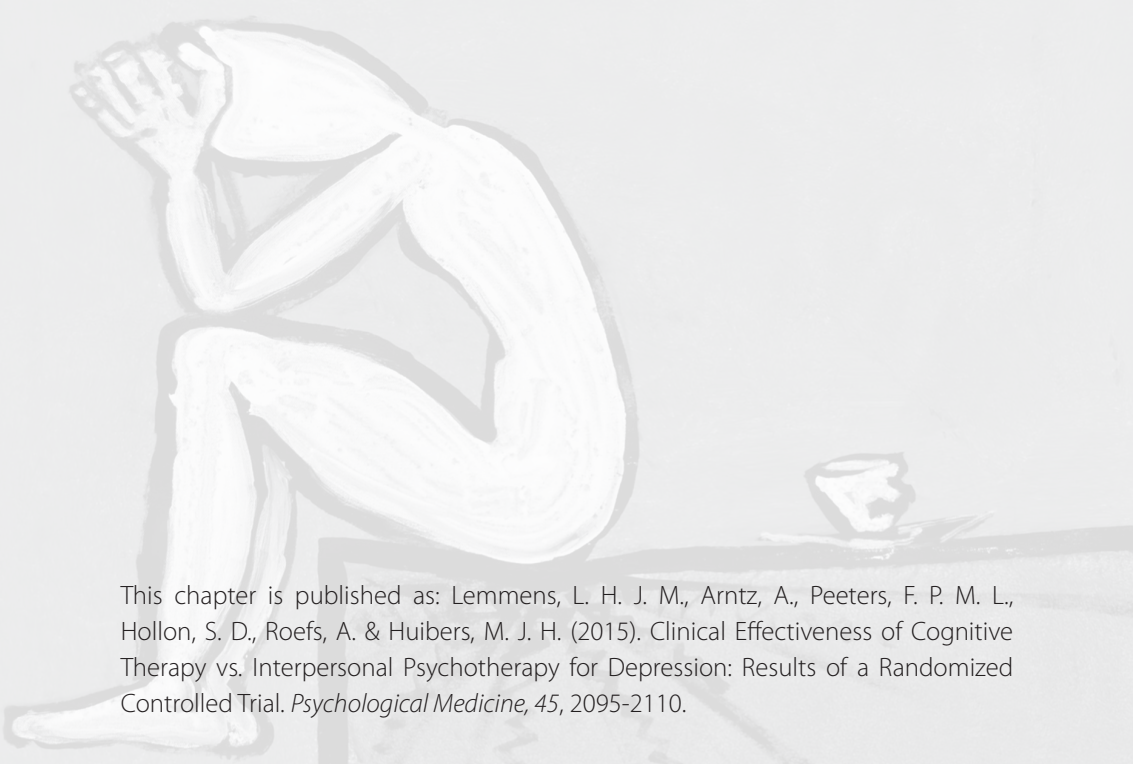
Tichenor, V., & Hill, C. E. (1989). A comparison of six measures of working alliance. *Psychotherapy: Theory, Research, Practice, Training, 26*(2), 195–199.

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*(3), 207–210.

Trivedi, M. H., Rush, A. J., Ibrahim, H. M., Carmody, T. J., Biggs, M. M., Suppes, T., . . . Kashner, T. M. (2004). The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: A psychometric evaluation. *Psychological Medicine: A Journal of Research in Psychiatry and the Allied Sciences, 34*(1), 73–82.

Vandromme, H., Hermans, D., Spruyt, A., & Eelen, P. (2007). Dutch translation of the Self-Liking/Self-Competence Scale – Revised: A confirmatory factor analysis of the two-factor structure. *Personality and Individual Differences, 42*(1), 157–167.

Vervaeke, G. A. C., & Vertommen, H. (1996). De Werkalliantievragenlijst (wav). *Gedragstherapie, 2*, 139–144.

Warmerdam, L., van Straten, A., Jongsma, J., Twisk, J., & Cuijpers, P. (2010). Online cognitive behavioral therapy and problem-solving therapy for depressive symptoms: Exploring mechanisms of change. *Journal of Behavior Therapy and Experimental Psychiatry, 41*, 64–70.

Warshaw, M. G., Dyck, I., Allsworth, J., Stout, R. L., & Keller, M. B. (2001). Maintaining reliability in a long-term psychiatric study: an ongoing inter-rater reliability monitoring program using the longitudinal interval follow-up evaluation. *Journal of Psychiatric Research, 35*, 297–305.

Warshaw, M. G., Keller, M. B., & Stout, R. L. (1994). Reliability and Validity of the Longitudinal interval follow-up evaluation for assessing outcome of anxiety disorders. *Journal of Psychiatric Research, 28*(6), 531–545.

Weissman, M. M. (1979). The efficacy of drugs and psychotherapy in the treatment of acute depressive episodes. *The American Journal of Psychiatry, 136*(4-B), 555–558.

Weissman, M. M., & Markowitz, J. C. (1994). Interpersonal psychotherapy: Current status. *Archives of General Psychiatry, 51*(8), 599–606.

Weissmann, A. N., & Beck, A. T. (1978). *Development and validation of the Dysfunctional Attitude Scale; a prelimary investigation*. Paper presented at the Annual Meeting of the American Educational research association, Toronto.

Whisman, M., Perez, J., & Ramel, W. (2000). Factor structure of the Beck Depression Inventory-Second Edition (BDI-II) in a student sample. *Journal of Clinical Psychology, 56*, 545–551.

Willemse, Y., & Trijsburg, R. W. (2005). Cognitieve gedragstherapie en interpersoonlijke psychotherapie. Een analyse van kritische succesfactoren. *Tijdschrift voor Psychiatrie, 47*, 593–602.

Young, M. A., Halper, I. S., Clark, D. C., & Scheftner, W. A. (1992). An item-response theory evaluation of the Beck Hopelessness Scale. *Cognitive therapy and research, 16*(5), 579–587.

Zee van der, K. I., & Sanderman, R. (1993). *Het meten van de algemene gezondheidstoestand met de RAND-36: een handleiding*. Groningen: Noordelijk Centrum voor Gezondheidsvraagstukken.

Zimmerman, M., Sheeran, T., & Young, D. (2004). The Diagnostic Inventory for Depression: a self-report scale to diagnose DSM-IV major depressive disorder. *Clinical Psychology, 60*, 87–110.

3

Clinical Effectiveness of Cognitive Therapy vs. Interpersonal Psychotherapy for Depression: Results of a Randomized Controlled Trial

Abstract

Although both Cognitive Therapy (CT) and Interpersonal Psychotherapy (IPT) have shown to be effective treatments for Major Depressive Disorder (MDD), it is not clear yet whether one therapy outperforms the other with regard to severity and course of the disorder. The current study examined the clinical effectiveness of CT vs. IPT in a large sample of depressed patients seeking treatment in a Dutch outpatient mental health clinic. We tested whether one of the treatments was superior to the other at post-treatment and at 5-months follow-up. Furthermore, we tested whether active treatment was superior to no treatment. We also assessed whether initial depression severity moderated the effect of time and condition and tested for therapist differences. 182 depressed adults were randomized to either CT ($n = 76$), IPT ($n = 75$) or a 2-month Waiting-List Control (WLC) condition ($n = 31$). Main outcome was depression severity, measured with the Beck Depression Inventory-II (BDI-II), assessed at baseline, 2, 3, and 7 months (treatment-phase) and monthly up to 5 months follow-up (8–12 months). No differential effects between CT and IPT were found. Both treatments exceeded response in the WLC condition, and led to considerable improvement in depression severity that was sustained up to one year. Baseline depression severity did not moderate the effect of time and condition. Within our power and time ranges, CT and IPT appeared not to differ in the treatment of depression in the acute phase and beyond.

3

Introduction

Both Cognitive Therapy (CT) and Interpersonal Psychotherapy (IPT) have proven to be well-standardised, effective treatments for Major Depressive Disorder (MDD; Cuijpers, van Straten, Andersson, & van Oppen, 2008). However, it is not clear yet whether one therapy outperforms the other with regard to severity and course of the disorder. Recently, Jakobsen, Hansen, Simonsen, Simonsen, and Gluud (2012) meta-analysed results of five randomized controlled trials (RCT) comparing post-treatment effects of CT and IPT on the Beck Depression Inventory (BDI). Three trials examined individual CT and IPT (Elkin, Shea, Watkins, & Imber, 1989; Luty et al., 2007; Quilty, McBride, & Bagby, 2008), one study examined the effects of group therapy (Bellino, Zizza, Rinaldi, & Bogetto, 2007), and one study included psychodynamic-IPT instead of traditional IPT (Hardy et al., 1995). Meta-analysis of the three studies that examined individual CT and IPT showed no significant differences in post-treatment depression severity scores between the two interventions (mean difference in favour of CT = 0.62 points on the BDI-II, p = 0.59, 95% CI [2.86, 1.61])[1]. Even though existing research consistently suggests that both treatments are equally effective, additional trials are necessary. As Jakobsen et al. (2012) pointed out, so far the data collected in randomized comparisons of CT and IPT are insufficient to reliably decide whether the two therapies have differential effects on the BDI. More specifically, trial sequential analysis on the three existing studies examining the effects of individual CT and IPT for depression ($n = 388$), indicated that data of at least 108 more patients is necessary to detect or reject an intervention effect with a minimal relevant difference of 4 BDI points (80% power). For an effect of 3 BDI points, this is much higher (791 additional patients)[2]. Furthermore, all studies included in Jakobsen's meta-analysis were classified as having a high risk of bias (especially Elkin et al., 1989), and none of them included data after treatment termination, leaving long-term effects largely unknown. Another unresolved issue in the comparison of CT and IPT is whether one of the two treatments is indicated for severe depression, as two studies found contradictory evidence: Elkin et al. (1989) favoured IPT, whereas Luty et al. (2007) found that CT was more effective for severe depression. In addition, treatment moderators and therapist effects are less well studied, but may be at least as important as treatment protocol. Lastly, it has been frequently argued in psychotherapy research that the effects of treatment studies cannot be attributed to the intervention(s) studied unless a placebo or no-treatment group is included (Klein, 1990). Even though several studies have individually compared CT or IPT to a non-active Waiting-List Control (WLC) condition, to date there are no randomized comparisons of CT, IPT and a WLC condition.

Study aims

To address these issues, the current study examined the clinical effectiveness of CT and IPT in a large sample of depressed patients seeking treatment in an outpatient mental health clinic in the Netherlands. The aim of the study was threefold: first, we tested

whether one of the treatments was superior to the other in the reduction of depressive symptoms as assessed with the BDI-II and on a set of secondary outcomes at post-treatment (7 months) and up to 5-months follow-up. Second, we tested for therapist differences and examined several potential moderators including baseline depression severity and total number of sessions. Third and final, we examined whether short-term effects of active treatment were superior to waiting-list control after 2 months. With this we add to previous studies conducted in the US and New Zealand with a large European sample. In order to determine the relative contribution of our study to the field, we added our data to the existing evidence base of trials that examined individual CT and IPT (Elkin et al., 1989; Luty et al., 2007; Quilty et al., 2008), and meta-analysed the findings. Furthermore, using the same set of four studies, trial sequential analysis was performed to determine whether our study added the information necessary to conclude equivalence between CT and IPT. Finally, the risk of bias in our study was assessed.

Methods

Trial Design

Details about the study design have been fully described elsewhere (Lemmens et al., 2011). The data came from a large single-centre RCT (parallel group design) examining the effectiveness and mechanisms of change of individual CT and IPT for depression. In this study, depressed outpatients were randomly allocated to one of three conditions: (a) CT, (b) IPT, (c) or a 2-month WLC condition followed by treatment of choice. In reporting the follow-up period of the study, three phases can be distinguished; (1) the 'Treatment Phase' (baseline to 7 months); the period when therapy was delivered; (2) the 'Trial Follow-Up Phase' (month 8–12) in which depression severity was measured monthly; and (3) the 'Long-term Follow-Up Phase' (month 12–24); the period covered by retrospective assessment at 24 months. The data presented here concern the acute outcome of therapy at the end of the 7-month Treatment Phase and data collected in the Trial Follow-Up Phase. The Medical Ethics Committee of Maastricht University approved the study protocol, and all participants provided written informed consent. The study is registered at the Netherlands Trial Register, part of the Dutch Cochrane Centre (ISRCTN67561918).

Participants and recruitment

Patients were adult outpatients (18 – 65 years[3]) referred to the mood disorder unit of the Maastricht Community Mental Health Centre with a primary diagnosis of MDD as confirmed by the Structural Clinical Interview for DSM-IV Axis I disorders (SCID-I; First, Spitzer, Gibbon, & Williams, 1997) conducted by a trained evaluator. Further inclusion criteria were: internet access, an e-mail address, and sufficient knowledge of the Dutch language. Exclu-

sion criteria were: bipolar or chronic (current episode > 5 years) depression, elevated acute suicide risk, concomitant pharmacological or psychological treatment[4], drugs- and alcohol abuse/dependence, and mental retardation (IQ < 80).

Procedure

Participants were recruited during regular intakes at our clinical site. After informed consent was obtained and a baseline assessment was completed, the researcher pressed the 'assign' button on the computer screen, after which the database randomly allocated the participant to one of three conditions using computer-generated block randomization (10:10:4). The random allocation sequence was generated by an independent computer scientist and was concealed from the researchers who were involved in the randomization procedure to prevent prediction of future assignment. Randomization was pre-stratified according to presence or absence of previous episodes. Blinding of patients and therapists for treatment condition was not possible. Sample size calculations of the active conditions CT and IPT were based on long-term expectations of CT vs. IPT (Lemmens et al., 2011). By combining rudimentary findings from previous randomized trials, a 20% difference in relapse/recurrence rate between the conditions was expected at the end of the two-year follow-up period. An a-priori power analysis indicated that with 75 patients per arm in the active conditions and taking 15% attrition into account, the study was powered at 80% (two-tailed α = .05) to detect a 20% difference in relapse rate between CT and IPT. With this sample size, taken 15% attrition into account, the study is powered at 80% to detect a medium effect size difference (Cohen's d = .50) in dimensional outcomes at p = .05, like in the present study. A second power calculation (powered at 80%; α = .05) showed that 28 patients in the WLC condition would suffice to detect a statistically significant difference of six BDI-II points between the two active conditions combined and the waiting list at 2 months (Cohen's d = .58)[5]. The main assessment points for the active conditions (CT and IPT) were baseline, 3, 7, 9 and 12 months. The primary outcome was also assessed at 8, 10 and 11 months. All assessments were administered on a computer. Pre- and post-treatment assessments (baseline and 7 months) were administered at the research centre. Mid-treatment (3 months) and follow-up assessments (8–12 months) took place via the internet. In addition, in order to examine the effect of the waiting list, depression severity of the total sample was assessed at 2 months[6].

Interventions

Treatment consisted of 16 to 20 individual sessions of 45 minutes, depending on the progress of the individual patient. The protocol allowed flexibility in scheduling appointments less often than weekly. Patients were considered to have had an adequate dose of treatment if they attended at least 12 sessions[7]. Both interventions were written out in a treatment manual. The CT protocol was based on the manual by Beck,

Rush, Shaw, and Emery (1979) which states that depression results from maladaptive information-processing strategies and is maintained by dysfunctional behavioural responses. In order to decrease depressive symptoms, CT focuses on identifying and altering the function, content and structure of cognitions, schemas and attitudes associated with negative affect. (Beck et al., 1979). The IPT protocol followed the guidelines laid out by Klerman, Weissman, Rounsaville, and Chevron (1984). IPT tries to understand the social and interpersonal context in which the depressive symptoms arose and investigates how they relate to the current social and personal context. The theorized mechanism is that if the patient can solve the interpersonal problem or is able to change the relation to this problem, the depressive symptoms should resolve as well (Markowitz & Weissman, 2004). In spite of their different theoretical backgrounds, CT and IPT also share several common features: both are time-limited, symptom-targeted and present-focused, and encourage the patient to regain control of mood and functioning. Furthermore, both try to increase the patient's activity level and pay special attention to the identification of expectations and assumptions using interventions such as exploration and clarification (Ablon & Jones, 1999; Willemse & Trijsburg, 2005). In addition, both therapies emphasize the importance of other (non-specific) factors such as structure, motivation and alliance.

Treatment Integrity

Treatment integrity and boundaries of both therapies were carefully monitored during the therapist training phase and the study itself. To prevent contamination, therapists were uniquely assigned to one of the treatment conditions. Furthermore, prior to the study, they received additional training by Steven Hollon (CT) and John Markowitz (IPT), experts in the respective interventions. The training (2x8 hours) addressed the theoretical framework and skills for each treatment using tailored case examples and role-play. During the study, therapists and the investigators who also participated as therapists in the study (MH for CT and FP for IPT) met biweekly in separate consultation sessions for each treatment condition to discuss their caseload. In addition, all sessions were videotaped. Four independent raters evaluated a random selection of 106 tapes on treatment adherence (content and quality) and competence. The raters (3 psychotherapists and 1 psychologist) were highly experienced in providing, teaching and supervising CT and IPT. All raters were trained prior to the study and were masked to treatment outcome. Nine tapes were double coded in order to assess interrater reliability.

　　To examine competence in both therapies, the overall quality scores of the Cognitive Therapy Scale (Dobson, Shaw, & Vallis, 1985) and a short version of the IPT Adherence and Quality Scale (Stuart, 2011) were used. In order to obtain a measure of adherence, the Collaborative Study Psychotherapy Rating Scale – version 6 (CSPRS-6; Hollon, Waskow, Evans, & Lowery, 1984; Hollon et al., 1988) was used. This 96-item questionnaire is able to distinguish between CT, IPT and clinical management (Hill, O'Grady, & Elkin,

1992). Items are rated on a 7-point Likert scale ("not at all" to "extensively") and higher scores indicate better adherence. Following Luty et al. (2007), we modified the original version to distinguish only between CT and IPT by omitting the 20 items pertinent to clinical management and medication, reducing the scale to 76 items.

Outcomes

Primary outcome was depression severity as measured with the Beck Depression Inventory, Second Edition (BDI-II; Beck, Steer, & Brown, 1996). Scores range from 0 to 63 with higher scores indicating more severe symptoms of depression. Secondary outcome measures included the following: General psychological distress was assessed with the Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983). A high score is indicative of severe distress (range 0–212). To obtain a measure of impairment in social functioning due to depression, the Work and Social Adjustment Scale (WSAS; Mundt, Marks, Greist, & Shear, 2002) was used. Scores range from 0 (no impairment) to 40 (severe impairment). Besides clinical outcome measures, several quality of life measures were included because they have shown to provide an important additional dimension to the evaluation of treatments for depression (Jakobsen et al., 2012; Kennedy, Eisfeld, & Cooke, 2001). The RAND-36 (van der Zee & Sanderman, 1993) was used to measure quality of life in terms of impairments due to physical and mental health status. Scores were transformed to a 100-point scale with higher scores indicating more positive experiential quality of life. The EuroQol-6D (EQ6D; EuroQolGroup 1990) was used to obtain an overall utility score for population-based quality of life (EQ5D)[8]. Utilities refer to the preference for any particular set of health outcomes and are indicated by a number between 0 (worst imaginable condition) and 1 (perfect health; Dolan, 1997; Drummond, Schulpher, Torrance, O'Brien, & Stoddart, 2005). For the current study, the Dutch value set was used (Lamers, McDonnell, Stalmeier, Krabbe, & Busschbach, 2006). For a more detailed description of these instruments and their psychometric properties we refer to our protocol paper (Lemmens et al., 2011).

Data Analysis

A detailed description of statistical procedures is presented in data supplement I. First, we mapped out patient flow from screening to randomization. After that we explored pre-treatment demographic and clinical variables of the three groups using descriptive statistics and checked for baseline differences between conditions in terms of size and clinical importance. Subsequently, therapist characteristics, treatment- and study compliance, and treatment integrity were determined, followed by examination of descriptive statistics on all clinical outcome measures at each time-point.

To examine whether CT and IPT differed in the reduction of depressive symptoms mixed (multilevel) regression analysis was used. We applied intention-to-treat analysis, with the BDI-II as the primary outcome. Our basic model was a 2-level (patients

and measurements) repeated measures design with depression severity as dependent variable, condition as a between subjects factor and time of measurement in weeks as a within-subjects factor. The difference between CT and IPT was represented by the time*condition interaction in the model. Furthermore, in all analyses, we controlled for baseline EQ5D and BDI-II scores because they showed considerable differences between the groups. Because visual inspection showed separate linear time slopes for the acute (0–7 months) and the follow-up (7–12 months) phase, change over time was assessed in two separate analyses (one for the acute phase, and one for the follow-up). Furthermore, change on secondary outcomes was assessed by testing the 2-level basic model on dependent variables BSI, WSAS, RAND36 and EQ5D. For BDI-II and EQ5-D mixed Gamma regression was used in the follow-up because of skewed distributions. Effect sizes Cohen's d and r (Cohen, 1988) for the continuous primary and secondary outcomes were computed from the multilevel estimates.

After that, we tested whether initial depression severity moderated the effect of time and condition by adding the two- and three-way interaction(s) of baseline depression severity (BDI-II) with time and condition to the basic model of the primary outcome BDI-II. Subsequently, we checked for influence of therapist and number of sessions by univariately adding them as fixed factors (main effect, and interaction with time) to the final model. Several other baseline characteristics that displayed potentially relevant differences between the treatment groups (gender, work- and marital status) were added to the model as covariates to see whether they would affect the results. All effects were tested at the $p < 0.05$ level (two-tailed). Then, we examined whether therapy outperformed the waiting list by comparing change in BDI-II scores of patients in the active groups after 2 months of therapy with those of patients in the WLC condition after 2 months of no-treatment. All analyses were carried out in SPSS version 21.0 and results are reported according to the CONSORT guidelines for reporting trials (Moher et al., 2010). Furthermore, the methodology of Jacobson & Truax (1991) was used to determine the proportion of patients that showed clinically meaningful change on the BDI-II. Response (the minimum amount of decrease in symptoms that has to be accomplished during therapy) was defined as a decrease of at least 9 BDI-II points during the Treatment Phase. Remission (the cut-off point between healthy and 'ill') was defined as an absolute value of 9 or less on the BDI-II. Frequency differences in response and remission rates between the groups were examined using mixed binary logistic regression.

To determine the relative contribution of our study to the field, we added our BDI-II data to the existing evidence base and meta-analysed findings from all four randomized trials that examined individual CT and IPT (Elkin et al., 1989; Luty et al., 2007; Quilty et al., 2008; current study) using the statistical program Open Meta Analysis (Wallace et al., 2012). Since existing trials did not adjust their outcome variable according to baseline values, non-covariate corrected means were included as estimates of the effects of the current study. As a sensitivity analysis, we repeated the analysis with (baseline BDI-II & EQ5D) corrected post-treatment BDI-II score. The same set of four

studies was used to perform trial sequential analysis on the BDI-II. Following Jakobsen et al. (2012), we conducted two analyses; one with a minimal relevant difference of 4 BDI points and 80% power, and one with more strict presumptions (BDI difference of 3 points and 90% power). Both analyses were based on a type I error of 5% and on the variance of all trials. Similar to the meta-analysis, effects of both non-covariate corrected as well as covariate corrected means as estimates of the effects of the current study were explored. In addition, resembling Jakobsen et al. (2012), bias risk was assessed with regard to sequence generation, allocation concealment, intention-to-treat analysis, blinding, drop-out, outcome measure reporting, presence of economic- and academic bias (see data supplement V for a full description of the criteria). An independent rater[9] checked the generated table entries for accuracy.

3
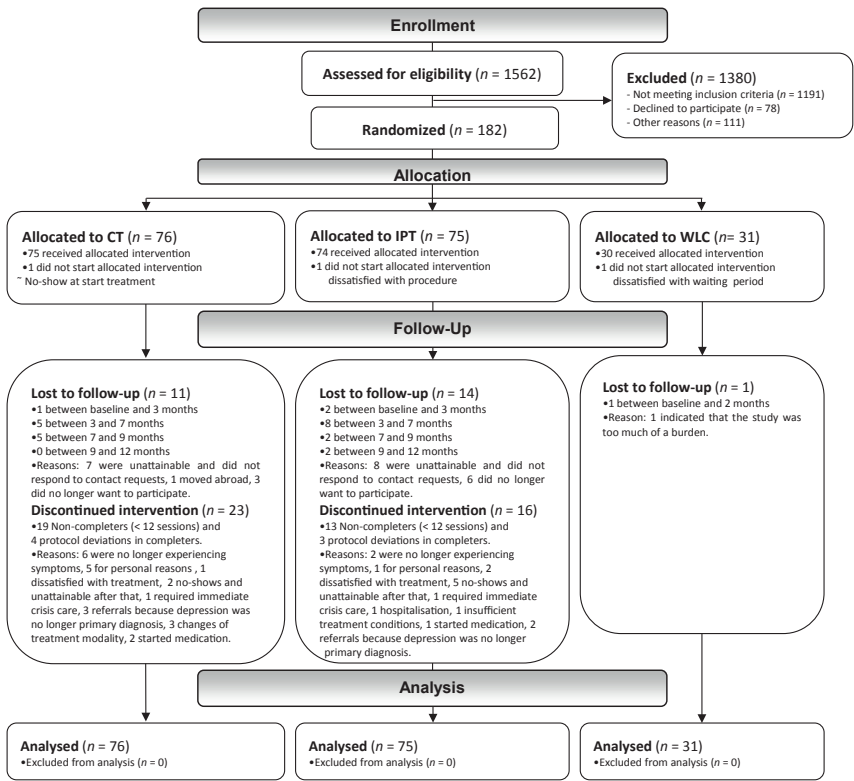
Figure 1. Flow of the participants. *Note:* CT = Cognitive Therapy; IPT = Interpersonal Psychotherapy; WLC = Waiting-List Control Condition.

Results

Description of the sample

Patient flow is shown in Figure 1. Of the 1562 patients who were initially screened for eligibility, 1191 did not meet inclusion criteria (mainly because of the use of antidepressant medication ($n=362$) or MDD not being the primary diagnosis ($n=434$)), 78 patients met inclusion criteria but declined to participate, and 111 were excluded for other reasons. A total of 182 patients were randomized ($n=76$, $n=75$, $n=31$ for CT, IPT and WLC respectively).

Pre-treatment characteristics of the total sample stratified according to condition are displayed in Table 1. There were no relevant differences between the patients in the two treatment conditions combined and the WLC condition for any of the sociodemographic variables or depression specifiers. However CT and IPT showed considerable differences on the BDI-II and EQ5D. Therefore, we controlled for this in all analyses. Mean age in the total group was 41.2 (SD = 12.1) and 100 participants (66.2%) were female. Mean baseline BDI-II score was 29.8 (SD = 9.0) and 49% of the patients were diagnosed with recurrent depression. More than half of the patients (55.6%) suffered from severe depression (BDI-II score ≥ 29) and the majority (59.6%) was still actively employed.

Table 1. Pre-treatment characteristics of the total sample ($n=182$) stratified according to condition.

	Cognitive Therapy	Interpersonal Psychotherapy	Waiting-List Control
	($n=76$)	($n=75$)	($n=31$)
Sociodemographic variables			
- Female sex, n (%)	54 (71.1)	46 (61.3)	16 (51.6)
- Age in years, mean (SD)	41.2 (12.4)	41.3 (11.8)	37.0 (12.5)
- Education, n (%)			
- Low	16 (21.1)	13 (17.3)	8 (25.8)
- Medium	48 (63.2)	41 (54.7)	20 (64.5)
- High	12 (15.8)	21 (28.0)	3 (9.7)
- Partner, yes, n (%)	43 (56.6)	51 (68.0)	18 (58.1)
- Active Employment*, yes, n (%)	43 (56.6)	47 (62.7)	24 (77.4)
Depression specifiers			
- BDI-II score, mean (SD)	28.4 (9.0)	31.2 (8.9)	29.2 (10.6)
- Recurrent depression, n (%)	38 (50.0)	36 (48.0)	15 (48.4)
- Severe depression**, n (%)	37 (48.7)	47 (62.7)	16 (51.6)

Note: BDI-II = Beck Depression Inventory Second Edition; * = data unavailable for CT ($n=1$); ** = BDI-II score ≥ 29.

Therapists

The 10 therapists (five for each intervention) who participated in the study were licensed psychologists, psychotherapists and psychiatrists. They ranged in age from 24 to 47 years and 80% were female. At study onset, therapists had an average of 9.1 (SD = 5.4) years of clinical experience (range 4 to 21 years). Each therapist treated between 7 and 30 patients, except for one therapist who left the clinical site after treating only two patients. No differences were found between the CT and IPT therapists with regard to years of experience (t (8) = 1.76, p = .12) or the number of patients treated (t (7) = 0.07, p = .95)[10].

Treatment and study compliance, and treatment integrity

Recruitment took place from February 2007 to April 2012. Treatment was delivered until December 2012. Follow-up data was complete in April 2013. At treatment termination, 117 patients (78%) in the active groups completed at least 12 sessions of therapy, with the mean number of sessions being 17 (SD = 2.9) in both conditions. There were no significant differences in session scheduling between CT and IPT[11]. Of the 34 patients that attended fewer than 12 sessions, 8 were early completers (< 12 sessions because they were no longer experiencing symptoms), 2 (one per condition) never started treatment (zero sessions), and the remaining 24 withdrew or were lost to follow-up. Non-completion rates were similar for the two intervention groups. There were no baseline differences between (early) treatment completers and non-completers.

Quality of therapy was rated as being (very) good to excellent in 83.3% of the CT tapes. For IPT this was 90.4%. Analysis of CSPRS-6 scores revealed significant differences in therapy specific behaviour between conditions, indicating that therapists adhered to protocol. CT specific behaviour was significantly more evident in CT than in IPT (mean score of 80.80 vs. 52.42, t (79.21) = 7.23, $p < .001$). Scores on IPT subscales in turn, were higher in IPT than in CT (85.75 vs. 44.57, t (86.96) = 10.79, $p < .001$). No significant differences between therapists were found. Mean ICC across subscales of double-rated tapes was 0.63 (range: 0.50 – 0.75).

The percentages of participants in the active conditions completing all assessments at 3 and 7 months were 98% and 89% respectively. Follow-up assessments at 9 and 12 months were completed by 85% of the patients. Patients lost to follow-up were significantly younger (36 vs. 42 years; t (149) = 2.13, $p = .03$) and more severely depressed at baseline (33.5 vs. 29.0; t (149) = 2.29, $p = .02$) compared to patients who completed all assessments. No significant differences in attrition rates emerged across conditions (see Figure 1).

Results on Primary and Secondary Outcome measures

Table 2 presents the observed mean [95% CI] scores and corrected mixed models estimated means [95% CI] on all outcome measures at each time-point in the treatment groups CT and IPT ($n = 151$). Estimated means from mixed regression are more valid than observed values because they correct for baseline BDI-II and EQ5D differences and take missing values into consideration.

Table 3 shows the final models of the mixed regression analysis on the primary and secondary outcomes in the Treatment Phase[12] and in the follow-up. All time*condition interactions were non-significant, indicating that there were no differential effects in symptom reduction between the two intervention groups. In line with this, no group differences were found on the mean scores at each time point (see Table 2). The analysis of the basic model (time, condition, time*condition controlling for baseline BDI-II and EQ5D) on the primary outcome BDI-II further showed a significant main effect of time in the Treatment Phase, indicating that depressive symptoms significantly decreased during treatment. The non-significant effect of time in the Follow-Up Phase shows that BDI-II scores remained stable up until 5 months after finishing treatment. A graphic representation of change in depressive symptoms over time as measured with the BDI-II is shown in Figure 2 (panel a). Analyses on secondary outcomes yielded a similar pattern of results, except for the BSI which also showed a significant effect of time in the Follow-Up Phase (see Table 3).

Table 2. Observed Mean scores [95% CI], corrected Mixed Regression based Estimated Means [95% CI] and Mean Differences [95 CI, *p*] between CT and IPT for all outcome measures in the intention-to-treat sample (*n* = 151).

| | Observed Means | | Estimated Means* | | | |
	CT (n = 76)	IPT (n = 75)	CT (n = 76)	IPT (n = 75)	CT-IPT	p
Beck Depression Inventory II						
- Baseline	28.4 [26.4, 30.4]	31.2 [29.1, 33.2]	29.6 [28.2, 31.0]	29.6 [28.2, 31.0]		
- 3 months	22.4 [19.8, 24.9]	24.1 [21.0, 27.1]	22.6 [20.4, 25.1]	20.9 [18.8, 23.2]	1.73 [-1.50, 4.96]	.29
- 7 months	13.7 [11.1, 16.3]	16.0 [12.7, 19.3]	14.1 [11.8, 16.8]	13.9 [11.6, 16.7]	0.14 [-3.44, 3.72]	.94
- 9 months	13.3 [10.5, 16.2]	15.8 [12.2, 19.4]	14.3 [11.8, 17.3]	13.2 [10.8, 16.1]	1.11 [-2.72, 4.93]	.57
- 12 months	12.6 [9.7, 15.6]	17.2 [13.3, 21.0]	13.4 [11.0, 16.3]	14.9 [12.2, 18.1]	-1.46 [-5.41, 2.49]	.47
Brief Symptom Inventory						
- Baseline	65.9 [58.9, 73.0]	71.1 [64.4, 77.7]	66.2 [61.6, 71.0]	62.7 [58.4, 67.3]	3.49 [-3.06, 10.04]	.30
- 7 months	38.6 [33.4, 47.4]	45.5 [13.0, 17.5]	39.8 [33.4, 47.4]	40.3 [33.8, 48.2]	-0.54 [-10.54, 9.46]	.92
- 12 months	29.5 [21.8, 37.2]	42.5 [32.5, 52.5]	31.0 [25.0, 8.3]	37.6 [30.2, 46.7]	-6.60 [-17.15, 3.94]	.22
Work and Social Adjustment Scale						
- Baseline	23.2 [21.4, 25.0]	22.4 [20.7, 24.0]	23.5 [21.9, 25.2]	21.7 [20.2, 23.2]	1.83 [-0.42, 4.08]	.11
- 7 months	15.4 [12.9, 17.9]	15.2 [13.0, 17.5]	15.6 [13.5, 18.1]	14.3 [12.3, 16.6]	1.34 [-1.77, 4.46]	.40
- 12 months	12.6 [10.1, 15.1]	14.0 [11.5, 16.4]	13.3 [11.1, 15.8]	13.2 [11.0, 15.8]	0.07 [-3.31, 3.45]	.97

3

Table 2. (continued)

	Observed Means		Estimated Means*			
	CT (n=76)	IPT (n=75)	CT (n=76)	IPT (n=75)	CT-IPT	p
RAND-36 Quality of Life						
- Baseline	46.7 [42.9, 50.5]	43.8 [40.6, 47.0]	45.7 [42.5, 49.0]	45.2 [42.0, 48.5]	0.49 [-4.18, 5.15]	.84
- 7 months	56.9 [51.6, 62.1]	54.5 [49.3, 59.6]	56.4 [51.7, 61.0]	55.7 [50.9, 60.4]	0.72 [-5.95, 7.39]	.83
- 12 months	58.3 [52.9, 63.7]	56.2 [51.0, 61.5]	58.1 [53.2, 63.0]	56.7 [51.7, 61.7]	1.37 [-5.63, 8.37]	.70
EuroQol-5D Utility score						
- Baseline	0.63 [0.58, 0.69]	0.55 [0.49, 0.62]	0.63 [0.57, 0.67]	0.58 [0.52, 0.63]	0.04 [-0.03, 0.12]	.26
- 7 months	0.77 [0.72, 0.83]	0.77 [0.70, 0.82]	0.78 [0.72, 0.83]	0.78 [0.72, 0.83]	0.00 [-0.08, 0.08]	.99
- 12 months	0.79 [0.73, 0.85]	0.79 [0.75, 0.86]	0.78 [0.70, 0.83]	0.81 [0.74, 0.85]	-0.03 [-0.12, 0.05]	.48

Note: CT = Cognitive Therapy; IPT = Interpersonal Psychotherapy; 95% CI = 95% Confidence Interval; * = Mixed Regression based Estimated Means corrected for BDI-II and EQ5D baseline scores (except for models in which BDI-II or EQ5D were dependent variables); Mean difference scores [95%CI, p] are based on the corrected mixed model estimates because they are more valid than observed values because they correct for baseline BDI-II and EQ5D differences and take missing values into consideration; Data unavailable for 3, 17, 23, and 25 patients at 3, 7, 9, and 12 months respectively.

Table 3. Results of analyses on primary and secondary outcome measures in the Treatment Phase (0–7 months) & Trial Follow-Up Phase (7–12 months).

	Treatment Phase (0–7)						Trial FU Phase (7–12)					
	B	95% CI	F	d.f.	r*	p	B	95% CI	F	d.f.	r*	p
Primary Outcome analysis												
Beck Depression Inventory–II[1]												
Intercept	2.71	[2.59, 2.83]	37.59	274	0.35	<.001	2.72	[2.60 – 2.85]	10.72	758	0.12	<.001
Baseline Severity	0.35	[0.28, 0.43]	88.74	274	0.49	<.001	0.34	[0.21 – 0.46]	27.38	758	0.19	<.001
Baseline Q.o.L.	-0.10	[-0.17, -0.02]	6.30	274	0.15	.01	-0.18	[-0.31 – -0.06]	8.27	758	0.10	<.001
Time	-0.02	[-0.03, -0.02]	59.81	274	0.42	<.001	0.00	[-0.00, 0.00]	0.12	758	0.01	.73
Condition	-0.01	[-0.25, 0.23]	0.01	274	0.01	.94	0.05	[-0.20, 0.30]	0.14	758	0.01	.71
Time x Condition	0.00	[-0.01, 0.02]	0.38	274	0.04	**.54**	0.01	[-0.00, 0.01]	2.06	758	0.05	**.15**
Analyses on Secondary Outcome Measures												
Brief Symptom Inventory[2]												
Time	-0.016	[-0.019, -0.012]	59.66	277	0.42	<.001	-0.007	[-0.012, 0.002]	6.51	252	0.16	.01
Condition	0.013	[-0.230, 0.257]	0.011	277	0.01	.92	0.188	[-0.109, 0.485]	1.55	252	0.08	.22
Time x Condition	0.002	[-0.006, 0.010]	0.31	277	0.03	**.58**	0.009	[-0.002, 0.019]	2.82	252	0.11	**.09**
Work and Social Adjustment Scale[2]												
Time	-0.013	[-0.016, -0.010]	58.51	279	0.42	<.001	-0.005	[-0.010, 0.000]	3.81	254	0.12	.05
Condition	-0.084	[-0.279, 0.111]	0.72	279	0.05	.40	-0.005	[-0.242, 0.232]	0.002	254	0.00	.97
Time x Condition	-0.000	[-0.007, 0.006]	0.004	279	0.00	**.95**	-0.005	[-0.005, 0.016]	1.09	254	0.07	**.30**

3

Table 3. (continued)

	Treatment Phase (0–7)						Trial FU Phase (7–12)					
	B	95% CI	F	d.f.	r*	p	B	95% CI	F	d.f.	r*	p
RAND-36 Quality of Life²												
Time	0.351	[0.240, 0.462]	39.13	137	0.47	<.001	0.063	[-0.077, 0.203]	0.79	126	0.08	.38
Condition	-0.621	[-7.320, 6.077]	0.034	137	0.02	.86	-0.634	[-7.804, 6.535]	0.031	130	0.02	.86
Time x Condition	-0.008	[-0.230, 0.214]	0.005	137	0.01	**.94**	-0.029	[-0.309, 0.252]	0.041	126	0.02	**.84**
EuroQol-5d Utility score²												
Time	-0.019	[-0.024, -0.014]	56.62	281	0.41	<.001	-0.003	[-0.011, 0.005]	0.504	257	0.04	.48
Condition	-0.003	[-0.332, 0.327]	0.000	281	0.00	.99	-0.139	[-0.525, 0.248]	0.500	257	0.04	.48
Time x Condition	-0.004	[-0.014, 0.006]	0.564	281	0.04	**.45**	-0.004	[-0.019, 0.012]	0.202	257	0.03	**.65**

Note: [1] =Mixed Models Gamma Regression with log-link on Beck Depression Inventory-II (BDI-II) + 1; Baseline Severity is standardised BDI-II score at baseline; Baseline Quality of Life (Q.o.L.) is standardised EQ5D utility score at baseline; Condition is CT vs. IPT, centered at -.5 and .5 respectively; Time is the linear trend in weeks, with week=0 at 7 months for the Treatment Phase and week=0 at 12 months for the trial FU Phase; Time effects in the Treatment Phase represent change from baseline to 7 months; Time effects in the Trial FU phase represent change from 7 to 12 months; Condition main effects represent the difference between conditions at 7 months in the Treatment Phase and 12 months for the Trial FU phase; [2] =Mixed Model Regression on Rand-36 Quality of Life; Mixed Models Gamma Regression with log-link on Brief Symptom Inventory (BSI) + 1, Work and Social Adjustment Scale (WSAS) + 1; (-1 * EQ5D Utility score + 1.01) * 100. Condition is CT vs. IPT, centered at -.5 and .5 respectively; Time is the linear trend in weeks, with week=0 at 7 (respectively 12) months; Time effects represent change from baseline to 7 months (respectively 7 – 12 months), Condition main effects represent the difference between conditions at 7 (respectively 12) months. All models are controlled for standardised baseline BDI-II and EQ5D scores (except for model in which EuroQol was the dependent variable); * =Effect size r=√(F/(F+d.f.)); Data unavailable for 3, 17, 23, 23, 25, and 25 patients at 3, 7, 8, 9, 10, 11 and 12 months respectively; 95% CI=95% Confidence Interval.
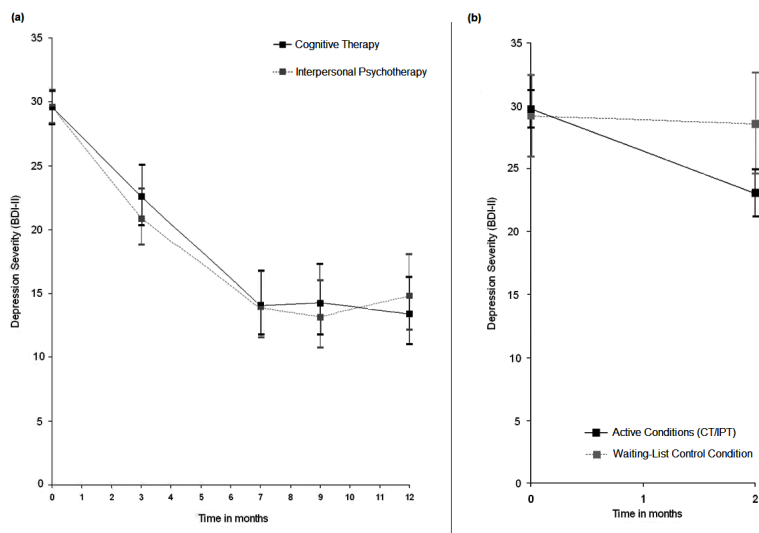
Figure 2. Course of Depression over time. Panel 2a: Mixed regression based estimated means and 95% CI (corrected for baseline severity and quality of life utility score) for the Beck Depression Inventory-II at baseline, 3, 7, 9, and 12 months for Cognitive Therapy (CT) and Interpersonal Psychotherapy (IPT). Panel 2b: Mixed Regression based estimated means and 95% CI for the Beck Depression Inventory-II at baseline and 2 months for the Active Conditions (CT/IPT) combined and the Waiting-List Control Condition.

Effect sizes

Corrected mixed model estimates and effect sizes (Cohen's *d* with baseline SD) of within- and between-condition changes of the final models are presented in Table 4. Within-group effect sizes at the end of the Treatment Phase were large in both groups (> 0.80; Cohen, 1988), and remained stable after that. Between-group effect sizes were small.

Table 4. Corrected Mixed Model Estimates of within and between condition changes and Cohen's *d* based on the BDI-II.

| | Improvement Effect Size[a] | | | | Between-Group Effect Size[b] | |
| | CT | | IPT | | CT vs. IPT | |
	Mean Change	*d*	Mean Change	*d*	Mean change difference	*d*
Baseline – 7 months	-15,52	1,71	-15,66	1,73	-0,14	0,02
Baseline – 12 months	-16,17	1,78	-14,71	1,62	1,46	-0,16

Note: BDI-II = Beck Depression Inventory-II; CT = Cognitive Therapy; IPT = Interpersonal Psychotherapy; [a] = (M_{t0}-M_{ti})/SD_{t0}; [b] = difference in improvement effect sizes between two groups; All models corrected for baseline severity and baseline quality of life utility score.

Moderation analyses

The moderation analysis of baseline severity (BDI-II) in the acute phase (results not shown in Table 3) did not show a time*condition*baseline severity interaction (F (1,755) = 0.05, p = .83). After removing the three-way interaction from the model, the two-way interaction between treatment condition and baseline severity was still not significant (F (1,757) = 1.10, p = .29), indicating that baseline severity did not moderate treatment outcome in the acute phase. The effect of therapist was not significant either (F (1,267) = 0.35, p = .93; see data supplement II), indicating that individual differences between therapists did not influence treatment outcome. In addition, adjustment for number of sessions or potential confounds (gender, work- and marital status) did not affect the results (see data supplement I). In the Follow-Up Phase, adjustment for baseline severity, as well as for the other potential confounds (total number of sessions, therapist, gender, work- and marital status) did not affect the results of the effectiveness analysis as well (see data supplement III).

Active treatment groups versus Waiting list

Response to therapy exceeded response in the WLC condition. Patients in the WLC condition showed minimal changes in depression severity across the 2-month waiting list period, suggesting that there was no spontaneous recovery (see Figure 2, panel b). Improvement in the active treatment conditions after 2 months of therapy (mean number of sessions = 6.5, SD = 1.7) was significantly larger (mean BDI-II change difference between active conditions combined and WLC condition = 6.16, t (173) = 2.47, p < .02, 95% CI [1.21, 10.05], r = .18).

Response and Remission

A complete overview of the observed and mixed regression based estimated response and remission rates of participants in each group for each time point, can be found in data supplement IV. Overall, observed clinical improvement after 3 months was modest, yielding remission in 13% of the patients. After 7 months, remission in the total sample was 34%. Remission rates of 37% at 9 and 12-months show that treatment effects remained stable after treatment termination. Even though mixed regression analysis yielded lower remission rates at three months, results at subsequent assessment points resembled those of the observed values. There were no significant differences between the two intervention groups (all p's > .12). The trend in favour of CT on the observed remission rates at 12 months (42.1% vs. 32.0%) disappeared after controlling for baseline differences in the mixed model analysis.

Relative contribution to the field

Meta-analysis with the random-effects model on non-covariate corrected post-treatment BDI data from the four trials that compared individual CT and IPT (Elkin et al., 1989; Luty et al., 2007; Quilty et al., 2008; current study) showed no significant differences between the two interventions (Mean difference in favour of CT = 1.1 BDI (-II) points, 95% CI [-3.07, 0.84]; see Figure 3, panel a). A second analysis using covariate corrected estimated means for the current study showed smaller differences between the two groups (see Figure 3, panel b). Trial sequential analysis on these four trials–using the non-covariate corrected means as estimates of the effects of the current study–based on a minimal relevant difference of 4 BDI points and 80% power, showed that the futility boundary was crossed (minimal required information size of $n = 526$, total $n = 539$). This indicates that our study provided the final information necessary to conclude equivalence of CT and IPT, within a 4-point limit (see Figure 4, panel a). Analysis with more strict presumptions (3-point BDI difference and 90% power) indicated that more data (total $n = 1251$) is needed to definitively settle the question of a differential effect. Analyses using covariate corrected means yielded similar results (see Figure 4, panel b; required information sizes of $n = 494$ and $n = 1174$ for the lenient and strict procedure respectively). Assessment of bias risk indicates that our study meets the majority of criteria for low risk trials (see supplementary data file V).

	Cognitive Therapy			Interpersonal Psychotherapy			Weight	Mean Difference IV, fixed [95% CI]	Mean Difference IV, fixed, 95% CI
	n	Mean	SD	n	Mean	SD			
3a. Non-covariate corrected									
Elkin et al. (1989)	59	13.4	10.6	61	12.0	10.6	26.5%	1.40 [-2.39 to 5.19]	
Luty et al. (2007)	86	14.8	12.4	91	17.1	12.9	27.4%	-2.30 [-6.03 to 1.43]	
Quilty et al. (2008)	45	11.0	11.1	46	12.0	8.9	22.4%	-0.96 [-5.09 to 3.17]	
Lemmens et al. (2015)*	76	14.0	11.4	75	16.7	13.7	23.7%	-2.69 [-6.71 to 1.32]	
Subtotal	266			273			100%	-1.11 [-3.07 to .84]	
Heterogeneity: χ^2 = 2.68, df = 3 (p = 0.44), I^2 = 0%									
3b. Sensitivity analysis									
Elkin et al. (1989)	59	13.4	10.6	61	12.0	10.6	24.9%	1.40 [-2.39 to 5.19]	
Luty et al. (2007)	86	14.8	12.4	91	17.1	12.9	25.8%	-2.30 [-6.03 to 1.43]	
Quilty et al. (2008)	45	11.0	11.1	46	12.0	8.9	21.0%	-0.96 [-5.09 to 3.17]	
Lemmens et al. (2015)**	76	14.1	11.1	75	13.9	11.2	28.4%	0.20 [-3.35 to 3.75]	
Subtotal	266			273			100%	-0.39 [2.28 to 1.50]	
Heterogeneity: χ^2 = 2.04, df = 3 (p = 0.56), I^2 = 0%									

Figure 3. Post-treatment effect of CT vs. IPT on the Beck Depression Inventory (random-effect model using the unstandardized mean difference score with a 95% CI). *Note:* Since existing trials did not adjust their outcome variable according to baseline values, non-covariate corrected means were included as estimates of the effects of the current study (3a). As a sensitivity analysis, we repeated the analysis with (baseline BDI-II & EQ5D) corrected post-treatment BDI-II scores (3b). *Note:* SD = Standard Deviation; IV = Inverse Variance; CI = Confidence interval; df = Degrees of Freedom; * = Non-covariate corrected estimated means. ** = Covariate corrected estimated means.
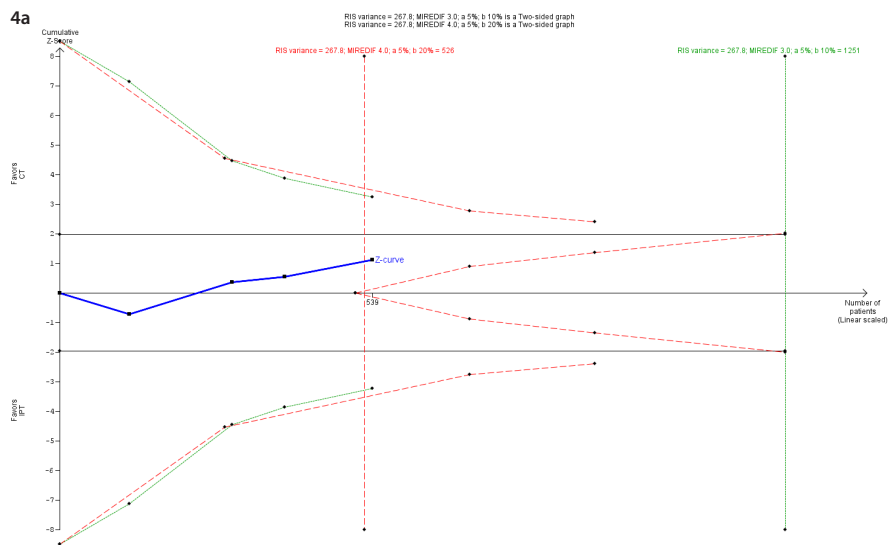
Figure 4, panel A. Trial Sequential Analysis (TSA) of the cumulative meta-analysis of the effect of Cognitive Therapy (CT) versus Interpersonal Psychotherapy (IPT) for Depression on the Beck Depression Inventory (BDI). TSA using non-covariate corrected means as estimates of the effects of the current study; the required information size of 526 participants (dotted lines) is calculated based on an intervention effect compare with IPT of 4 BDI points, and a power of 80%. With these presumptions, the cumulated Z curve (bold curve) crosses the futility boundary, implying that there are no significant differences in effect between the two interventions and no more trials are needed. The required information size of 1251 (etched lines) is calculated based on an intervention effect compared with IPT of 3 BDI points, and a power of 90%. With these presumptions, the cumulated Z curve (bold curve) does not cross the trial sequential monitoring boundaries, implying that there is no firm evidence for a beneficial effect of CT compared with IPT. Both analyses based on a variance of 267.8 and a risk of type I error of 5%.

4b
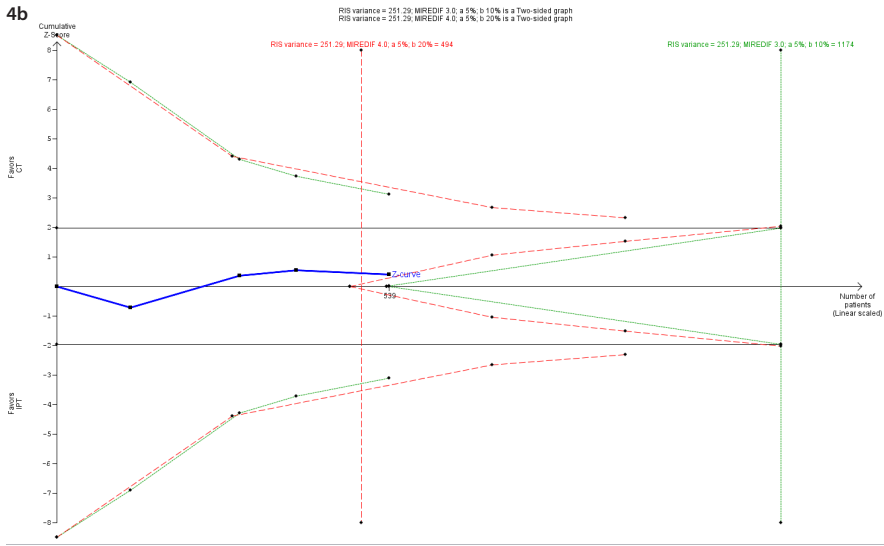
Figure 4, panel B. Trial Sequential Analysis (TSA) of the cumulative meta-analysis of the effect of Cognitive Therapy (CT) versus Interpersonal Psychotherapy (IPT) for Depression on the Beck Depression Inventory (BDI). Panel 4b: TSA using covariate corrected means (Baseline BDI-II and EQ5D) as estimates of the effects of the current study; the required information size of 494 participants (dotted lines) is calculated based on an intervention effect compare with IPT of 4 BDI points, and a power of 80%. With these presumptions, the cumulated Z curve (bold curve) crosses the futility boundary, implying that there are no significant differences in effect between the two interventions and no more trials are needed. The required information size of 1174 (etched lines) is calculated based on an intervention effect compared with IPT of 3 BDI points, and a power of 90%. With these presumptions, the cumulated Z curve (bold curve) does not cross the trial sequential monitoring boundaries, implying that there is no firm evidence for a beneficial effect of CT compared with IPT. Both analyses based on a variance of 251.3 and a risk of type I error of 5%.

Discussion

Main results

In this study it was found that CT and IPT do not differ in the treatment of depression in the acute phase and beyond. By demonstrating that both treatments exceeded response in the WLC condition and led to considerable improvement in depressive symptoms that was sustained 5 months beyond the end of treatment, our study does not only replicate, but also extends findings of other RCTs that examined individual CT and IPT head-to-head (Elkin et al., 1989; Luty et al., 2007; Quilty et al., 2008). A substantial number of patients (67%) responded to therapy, and the overall improvement in depressive symptoms as assessed with the BDI-II was about 50% in each group. At the end of the 7-month Treatment Phase, 34% of patients were in remission. Even though this is lower than the rates of response, it is within the range of reported effects in RCT efficacy contexts (Keitner, Ryan, & Solomon, 2006). However, it does indicate that the majority of patients experienced residual symptoms at the end of treatment, making them vulnerable for relapse. Pre- to post-treatment effect sizes in our trial were within range of those reported by Elkin et al. (1989), Luty et al. (2007), and Quilty et al. (2008). It has to be noted that our treatment phase was somewhat longer than in other trials, which gave patients more time to improve. Therapist effects were found to be non-significant, indicating that experienced therapists in a research-oriented routine clinical setting who receive additional training and are carefully monitored throughout, show comparable results. This is in line with previous research by Crits-Christoph et al. (1991) that concludes that therapist effects are more likely to be found in outcome studies with minimal training and supervision.

Change in our trial occurred later in time than in other (often US based) trials. A direct comparison at 2 months shows an average drop of only 6 BDI-II points in the current study as compared to 10–15 points in other trials (Dimidjian et al., 2006; Jarrett et al., 1999). An explanation might be the difference in session frequency. In the Netherlands, patients receive treatment mostly once a week, whereas in US (efficacy) trials, this is mostly twice a week, especially in the first phase of treatment. It is expected that an increase from one to two sessions per week in the initial phase of treatment, while keeping the total number of treatment sessions constant, explains the differences in early response rates, whereas post-treatment results remain similar. Furthermore, in contrast to previous findings, baseline severity did not moderate the effect of time and condition. However, this effect has been mainly established in comparisons between active treatments vs. nonspecific controls (Driessen, Cuijpers, Hollon, & Dekker, 2010; Fournier et al., 2010). We would not expect such differences between the two active treatments, unless they differed in quality of treatment delivered, which was not the case in the current study.

Methodological considerations

This study has both strengths and limitations. First of all, our study is one of the largest clinical trials in the field. With a sample size of $n = 151$ in the active groups, we provided the final information necessary to conclude equivalence of individual CT and IPT for adult depression within a 4 BDI points limit. Furthermore, it is the first study that replicates research results obtained in the US and New-Zealand in a European sample and it thereby extends knowledge about the relative effectiveness of these treatments around the world. In addition, our study was the first to add a WLC condition to a randomized comparison of CT and IPT. The inclusion of an untreated control group diminished uncertainty about whether or not the observed effect was a result of natural course. Given that the WLC condition showed minimal spontaneous remission, the within-condition estimates of the active groups are good estimators of the true effects of CT and IPT. With 85% of patients providing data at the 12-month assessment, attrition rates were low. We therefore are confident that only minimal biases occurred as a result of missing data. By training therapists in a research-oriented routine clinical setting we took care of both the trial's quality as well as generalisability of the trial findings. Other strengths include a broad range of outcomes, the use of modern statistical techniques, intention-to-treat analyses, assessment of bias risk, and an extensive integrity check. With regard to the integrity check, it has to be noted however, that even though the overall ICC was acceptable, the range was somewhat less satisfactory, which is a limitation of the present study.

Some other limitations should be mentioned as well. First, as outlined by Cuijpers, Li, Hofmann, and Andersson (2010) it is important to include both clinician-rated and self-report measures of depression to assess improvement in depressive symptoms over the course of psychotherapy. Although structured diagnostic interviews ensured accurate classification of patient's diagnoses at baseline, all follow-up measures were based on self-report. We therefore have no information on actual diagnoses at follow-up. Furthermore, interrater reliability data on the SCID-I diagnoses is lacking. Third, the duration of the WLC condition was significantly shorter than the treatment time. Therefore a comparison of the full effects of treatment vs. no treatment was not possible. However, given the study population (patients with major psychopathology who already applied for treatment on their own accord), and the distress and risks related to depression, we considered it unethical to include a full WLC condition. We therefore decided to include a waiting list that was as short as possible. Fourth, because we used a fixed post-treatment assessment and allowed flexibility in scheduling treatment sessions, not all patients finished treatment within the 7-month treatment phase. As a result, the score they provided at the 7-month assessment is a reflection of their progress up until that point, and cannot be considered an absolute post-treatment score. However, because these patients were in the final stage of therapy and only received one or two additional sessions on average, we do not think that this has led to drastic changes in outcomes. This was further underlined by the results at

9 months, the point by which all patients had finished treatment, which yielded the same pattern of results. Furthermore, the exclusion of patients receiving concomitant treatment, and those using antidepressant medications in particular, may have reduced generalisability to the whole population of treatment-seeking depressed patients. However, because medication is not a preferred option for initial treatment of episodic depression in the Netherlands (see Spijker et al., 2013), only a small group was excluded from the trial for this reason. We therefore think that the negative effects on generalizability for the Dutch situation are relatively small. However, it should be noted that this is probably different for countries in which the use of medication plays a more prominent role in the treatment of MDD. Lastly, we did not control for multiple testing on the secondary outcomes.

Clinical implications and future research

The present study is the fourth to conclude that no significant differences could be detected between individual CT and IPT in treating depression in the acute phase across the full range of severity. What does this mean for clinical practice? Although it may seem attractive to shift the focus to cost-effectiveness, we think that this would be too early because several crucial aspects for treatment evaluation are still unknown. First of all, existing studies are powered to demonstrate a medium effect between conditions at a modest power level (e.g. 80%). This leaves room for smaller differences between the two approaches that could not be detected. True equivalence trials are necessary to test whether CT and IPT are really equivalent. However, it has to be noted that these studies usually require very large sample sizes (depending on what difference is viewed as clinically relevant), and therefore are a challenge in itself. Furthermore, in spite of general equality of treatments, it might be the case that one treatment is superior to the other with regard to certain subgroups of depressed patients. It is therefore important to identify (patient) characteristics that predict differential treatment response. Unfortunately these moderators and predictors of treatment success are still largely unknown. In addition, the fact that therapies that show a distinguishable theory and use specific therapeutic techniques do not necessarily lead to specific outcomes, calls for specification of their underlying mechanisms. One could argue that different specific pathways lead to similar results (specificity hypothesis). However, it is also possible that change is driven by more common factors such as motivation and therapeutic alliance (nonspecificity hypothesis). Process research is needed to settle this question (Kazdin, 2009). Finally, existing studies, including the present, only focus on the short-term effectiveness. It might be the case however, that even though both treatments are equally effective in the acute phase of treatment, differences will come forward in the long-term with regard to relapse prevention. It is therefore also important to examine long-term effects more closely, as we plan to do in a subsequent paper.

Footnotes

[1] Data derived from data reported in Jakobsen et al. (2012).

[2] Data and method derived from Jakobsen et al. (2012).

[3] These cut-off points were chosen because in the Dutch health care system, patients under 18 and over 65 are treated in different units.

[4] This was done because the study was also designed to examine mechanisms of change. It was therefore important that patients only received one type of treatment at a time.

[5] Further details about power calculations can be found in our design paper (Lemmens et al. 2011).

[6] Effects of treatment given after wait in the WLC condition were assessed as well, but are beyond the scope of the current paper and will be presented elsewhere.

[7] 12 sessions is a commonly used cut-off for the definition of completion in clinical practice and research (see e.g., Beck, 1995; Weissman et al., 2000).

[8] The EuroQol-6D assesses six health state dimensions (mobility, self-care, usual activity, pain/discomfort, anxiety/depression and cognitive impairment). The resulting profiles of these dimensions were converted in utilities. However, the norms for calculating utility scores are only available for a five dimension version of the EuroQol (EuroQol-5D). As a result, the dimension cognition is not taken in consideration in the utility score.

[9] Given his role as editor of the Cochrane Hepato-Biliary Group (CHBG), and expertise in assessment of bias risk, we asked dr. J. Jakobsen (PhD, MD) for his opinion.

[10] Calculation based on 9 therapists. The therapist that left after treating only two patients was not included in this analysis.

[11] Even though patients in the IPT condition were a bit faster in scheduling session 1 to 6 (CT=59 days vs. IPT=53 days) and session 1 to 12 (CT=133 days vs. IPT=121 days), these differences were not statistically significant (t (140)=1.87, p=.06, and t (118)=1.61, p=.11 for session 6 and 12 respectively). The difference in overall duration of treatment (CT=178 days vs. IPT=197 days) was also not significant (t (147)=-1.51. p=.13).

[12] Due to the possibility of scheduling sessions less often than weekly, 35% of patients were still in treatment at the end of the acute phase (7 months). The mean number of sessions at the 7-month assessment was 14.5 (SD 3.1) and on average patients had an additional 1.7 (SD 2.6) sessions after that. Since we consider these patients treatment completers in the final stage of therapy at the time of 7-month assessment we included them in the analyses.

3

References

Ablon, J. S., & Jones, E. E. (1999). Psychotherapy process in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 67*(1), 64–75.

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

Beck, A. T., Steer, R., & Brown, G. K. (1996). *Beck Depression Inventory II: Manual*. Boston, MA: Hartcourt Brace.

Beck, J. S. (1995). *Cognitive Therapy: Basics and Beyond*. New York, NY: Guilford Press.

Bellino, S., Zizza, M., Rinaldi, C., & Bogetto, F. (2007). Combined therapy of major depression with concomitant borderline personality disorder: Comparison of interpersonal and cognitive psychotherapy. *The Canadian Journal of Psychiatry / La Revue canadienne de psychiatrie, 52*(11), 718–725.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Crits-Christoph, P., Baranackie, K., Kurcias, J., Beck, A., Carroll, K., Perry, K., . . . Zitrin, C. (1991). Meta-Analysis of Therapist Effects in Psychotherapy Outcome Studies. *Psychotherapy Research, 1*(2), 81–91.

Cuijpers, P., Li, J., Hofmann, S. G., & Andersson, G. (2010). Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: A meta-analysis. . *Clinical Psychology Review, 30*, 768–778.

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology, 76*(6), 909–922.

Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine, 13*(3), 595–605.

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., . . . Jacobson, N. S. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology, 74*(4), 658–670.

Dobson, K. S., Shaw, B. F., & Vallis, T. M. (1985). Reliability of a measure of the quality of cognitive therapy. *British Journal of Clinical Psychology, 24*(4), 295–300.

Dolan, P. (1997). Modeling Valuations for EuroQol Health States. *Medical Care, 35*(11), 1095–1108.

Driessen, E., Cuijpers, P., Hollon, S. D., & Dekker, J. J. M. (2010). Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. *Journal of Consulting and Clinical Psychology, 78*(5), 668–680.

Drummond, M. F., Schulpher, M. J., Torrance, G. W., O'Brien, B., & Stoddart, G. L. (2005). *Methods for the Economic Evaluation of Health Care Programmes* (3rd ed.). New York, NY: Oxford University Press.

Elkin, I., Shea, M. T., Watkins, J. T., & Imber, S. D. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry, 46*(11), 971–982.

EuroQolGroup. (1990). EuroQol–a new facility for the measurement of health-related quality of life. *Health Policy, 16*, 199–208.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I)*. New York, NY: Biometrics Research Department New York State Psychiatric Institute.

Fournier, J. C., DeRubeis, R. J., Hollon, S. D., Dimidjian, S., Amsterdam, J. D., Shelton, M. C., & Fawcett, J. (2010). Antidepressant drug effects and depression severity: A patient-level meta-analysis. *JAMA, 303*(1), 47–53.

Hardy, G. E., Barkham, M., Shapiro, D. A., Stiles, W. B., Rees, A., & Reynolds, S. (1995). Impact of Cluster C personality disorders on outcomes of contrasting brief psychotherapies for depression. *Journal of Consulting and Clinical Psychology, 63*(6), 997–1004.

Hill, C. E., O'Grady, K. E., & Elkin, I. (1992). Applying the Collaborative Study Psychotherapy Rating Scale to rate therapist adherence in cognitive-behavior therapy, interpersonal therapy, and clinical management. *Journal of Consulting and Clinical Psychology, 60*(1), 73–79.

Hollon, S. D., Evans, M. D., Auerbach, A., DeRubeis, R. J., Elkin, I., Lowery, A., . . . Piasecki, J. (1988). *Development of a system for rating therapies for depression: Differentiating cognitive therapy, interpersonal psychotherapy, and clinical management pharmacotherapy*. Unpublished Manuscript.

Hollon, S. D., Waskow, I. E., Evans, M., & Lowery, H. A. (1984). *Systems for rating therapies for depression*. Paper presented at the Annual Convention of the American Psychiatric Association, Los Angeles, CA.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to define meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12–19.

Jakobsen, J. C., Hansen, J. L., Simonsen, S., Simonsen, E., & Gluud, C. (2012). Effects of cognitive therapy versus interpersonal psychotherapy in patients with major depressive disorder: a systematic review of randomized clinical trials with meta-analyses and trial sequential analyses. *Psychological Medicine, 42*, 1343–1357.

Jarrett, R. B., Schaffer, M., McIntire, D., Witt-Browder, A., Kraft, D., & Risser, R. C. (1999). Treatment of atypical depression with cognitive therapy or phenelzine: A double-blind, placebo-controlled trial. *Archives of General Psychiatry, 56*(5), 431–437.

Kazdin, A. E. (2009). Understanding how and why psychotherapy leads to change. *Psychotherapy Research, 19*(4–5), 418–428.

Keitner, G. I., Ryan, C. E., & Solomon, D. A. (2006). Realistic expectations and a disease management model for depressed patients with persistent symptoms. *Journal of Clinical Psychiatry, 67*(9), 1412–1421.

Kennedy, S. H., Eisfeld, B. S., & Cooke, R. G. (2001). Quality of life: an important dimension in assessing the treatment of depression? *Journal of Psychiatry and Neuroscience, 26*, 23–28.

Klein, D. F. (1990). NIMH collaborative research on treatment of depression. *Archives of General Psychiatry, 47*, 682–684.

Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. S. (1984). *Interpersonal psychotherapy for depression*. New York, NY: Basis Books.

Lamers, L. M., McDonnell, J., Stalmeier, P. F. M., Krabbe, P. F. M., & Busschbach, J. J. V. (2006). The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Economics, 15*(10), 1121–1132.

Lemmens, L. H. J. M., Arntz, A., Peeters, F. P. M. L., Hollon, S. D., Roefs, A., & Huibers, M. J. H. (2011). Effectiveness, relapse prevention and mechanisms of change of cognitive therapy vs. interpersonal therapy for depression: study protocol for a randomized controlled trial. *Trials, 12*(1), 150–162.

3

Luty, S. E., Carter, J. D., McKenzie, J. M., Rae, A. M., Frampton, C. M. A., Mulder, R. T., & Joyce, P. R. (2007). Randomized controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *British Journal of Psychiatry, 190*, 496–502.

Markowitz, J. C., & Weissman, M. M. (2004). Interpersonal psychotherapy: principles and applications. *World Psychiatry, 3*, 136–139.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gotzsche, P. C., Deveraux, P. J., . . . Altman, D. G. (2010). Research Methods & Reporting: CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomized trials. *BMJ (Clinical Research Ed.), 340*, c869.

Mundt, J. C., Marks, I. M., Greist, J. H., & Shear, K. (2002). The Work and Social Adjustment Scale: A simple accurate measure of impairment in functioning. *British Journal of Psychiatry, 180*, 461–464.

Quilty, L. C., McBride, C., & Bagby, R. M. (2008). Evidence for the cognitive mediational model of cognitive behavioural therapy for depression. *Psychological Medicine, 38*, 1531–1541.

Spijker, J., Bockting, C. L. H., Meeuwissen, J. A. C., van Vliet, I. M., Emmelkamp, P. M. G., Hermens, M. L. M., & van Balkom, A. L. J. M. (2013). *Multidisciplinaire Richtlijn Depressie (Derde revisie): Richtlijn voor de diagnostiek, behandeling en begeleiding van volwassen patiënten met een depressieve stoornis*. Utrecht: Trimbos-instituut.

Stuart, S. (2011). *IPT Adherence and Quality Scale.* (Unpublished Manuscript). Interpersonal Psychotherapy Institute. Iowa.

Wallace, B. C., Daharbreh, I. J., Trikalinos, T. A., Lau, J., Trow, P., & Schmid, C. H. (2012). Closing the Gap between Methodologists and End-Users: R as a Computational Back-End. *Journal of Statistical Software, 49*(5), 1–15.

Weissman, M. M., Markowitz, J. C., & Klerman, G. L. (2000). *Comprehensive guide to interpersonal psychotherapy*. New York, NY: Basic Books.

Willemse, Y., & Trijsburg, R. W. (2005). Cognitieve gedragstherapie en interpersoonlijke psychotherapie. Een analyse van kritische succesfactoren. *Tijdschrift voor Psychiatrie, 47*, 593–602.

Zee van der, K. I., & Sanderman, R. (1993). *Het meten van de algemene gezondheidstoestand met de RAND-36: een handleiding*. Groningen: Noordelijk Centrum voor Gezondheidsvraagstukken.

Supplementary data

Data Supplement I

Detailed description of Statistical Analyses

We started by mapping out patient flow from screening to randomization. After that, we explored pre-treatment demographic and clinical variables of the three groups using descriptive statistics and checked for baseline differences between conditions in terms of size and clinical importance. Subsequently, therapist characteristics, treatment- and study compliance, and treatment integrity was determined, followed by examination of descriptive statistics on all clinical outcome measures at each time-point.

In order to examine whether CT and IPT differed in the reduction of depressive symptoms mixed (multilevel) regression analysis using restricted maximum likelihood estimation was used. Visual inspection of change on the primary outcome BDI-II over time showed separate linear time slopes for the acute phase and the follow-up. Since it was not possible to fit them in one model, we separately assessed change over time for the Treatment Phase (0–7 months) and the Trial Follow-Up Phase (7–12 months). Further inspection of the data and residuals of ordinary mixed regression showed that the assumption of normality was violated by a right skewed distribution from month 7 onwards. Therefore, gamma regression with a log link was used with BDI-II + 1 scores (1 added as gamma regression cannot handle zero's and quite some participants had zero scores). All analyses were intention-to-treat, meaning that all patients that enrolled in the study were included in the analyses, irrespective of completing therapy or assessments (Hollis & Campbell, 1999). Acute effects were examined by modeling time effects from baseline to 7 months using an unstructured covariance structure for the repeated parts as simpler models had poorer fit. For the enduring effects (7 – 12 months), a Toeplitz covariance structure was used, being the simplest structure NS different from unstructured.

Our initial basic model was a 3-level repeated measures design (therapists, patients and measurements) with depression severity (measured with the BDI-II) as the dependent variable, condition (CT vs. IPT, centered at -0.5 and 0.5) as a between subject variable and time of measurement in weeks as a within-subject factor. We set weeks at zero at 7 months, respectively at 12 months, so that the main effect of condition represented the condition effect at these measurements. The difference between CT and IPT was represented by the time*condition interaction in the model. Because of relevant differences in baseline severity (BDI-II) and quality of life utilities (EQ5D) between CT and IPT, we controlled for this in all analyses by adding their standardised baseline scores as covariates to all models[1]. Despite trying various (simple) models and covariance structures, the estimations including therapists as random level failed to converge, probably because the number of patients nested within therapists was too

small. Therefore therapist was omitted from further analyses as a random effect. Nevertheless, the effect of therapist was examined later by adding it as a fixed effect to the final model (see further).

After that, change on secondary outcomes was assessed by testing the 2-level basic model (time, condition, and time*condition, controlling for baseline BDI-II and EQ5D scores) on dependent variables BSI, WSAS, RAND36 and EQ5D. For RAND36, mixed models regression was used. Given the right skewed distribution of BSI and WSAS scores, we used gamma regression with a log link for BSI +1 and WSAS+1 scores. Since EQ5D utilities showed a left skewed distribution, scores were transformed ((-1 * EQ5D utility score + 1.01)*100) to meet assumptions of gamma regression.

Effect sizes Cohen's d and r (Cohen, 1988) for the continuous primary and secondary outcomes were computed from the multilevel estimates. Within-condition change was defined as Cohen's d = (baseline mean − mean at time i)/($\sqrt{}$baseline variance). Between-group effect sizes were determined by calculating the difference between the within-condition effect sizes of CT and IPT at time i. r was defined as $\sqrt{(F/(F + d.f.))}$.

Then, we tested whether initial depression severity moderated the effect of time and condition by adding the two- and three-way interaction(s) of baseline depression severity (continuous standardised BDI-II score) with time and condition to the basic model of the primary outcome BDI-II. The moderator*condition and moderator*time*condition interactions were of primary interest in this analysis. Power analysis showed that our study was powered at 80% to detect medium effect size interactions (f = 0.25) at two tailed α = .05. Non-significant interactions were hierarchically excluded from the model until only statistically significant prognostic variables remained. Subsequently, we checked for influence of therapist and number of sessions by univariately adding them as fixed factors (main effect, and interaction with time) to the final model. To conclude, several other baseline characteristics that displayed potentially relevant differences between the treatment groups (gender, work- and marital status) were added to the model as covariates to see whether they would affect the results. All effects were tested at the $p < 0.05$ level (two-tailed).

Subsequently, we examined whether therapy outperformed the waiting list by comparing change in BDI-II scores of patients in the active groups after 2 months of therapy with those of patients in the WLC condition after 2 months of no-treatment. All analyses were carried out in SPSS version 21.0 and results are reported according to the CONSORT guidelines for reporting trials (Moher et al., 2010).

The methodology of Jacobson & Truax (1991) was used to determine the proportion of patients that showed clinically meaningful change on the BDI-II. Response (the minimum amount of decrease in symptoms that has to be accomplished during therapy) was defined as a decrease of at least 9 BDI-II points during the Treatment Phase. Remission (the cut-off point between healthy and 'ill') was defined as an absolute value

of 9 or less on the BDI-II. To examine frequency differences in response and remission rates between the groups, mixed binary logistic regression with an unstructured covariance structure was used. All models were controlled for standardised baseline BDI-II and EQ5D scores, Time*Baseline BDI-II, and Condition*Time. All analyses were carried out in SPSS version 21.0 and results are reported according to the CONSORT guidelines for reporting trials (Moher, Schulz, & Altman, 2001).

In order to determine the relative contribution of our study to the field, we meta-analysed findings from all four randomized trials that examined individual CT and IPT (Elkin, Shea, Watkins, & Imber, 1989; Luty et al., 2007; Quilty, McBride, & Bagby, 2008; current study). Using the statistical program Open Meta Analysis (Wallace et al., 2012), we analysed the post-treatment BDI-II scores with a random-effects model using the unstandardized mean difference score with a 95% confidence interval (CI). Since other trials did not adjust their outcome variable according to baseline values, we included non-covariate corrected means as estimates of the effects of the current study. As a sensitivity analysis, we repeated the analysis with adjusted post-treatment BDI-II scores. The same set of four studies was used to perform trial sequential analysis on the BDI-II. Following Jakobsen, Hansen, Simonsen, Simonsen, & Gluud (2012), we conducted two analyses; one with a minimal relevant difference of 4 BDI points and 80% power, and one with more strict presumptions (BDI difference of 3 points and 90% power). Both analyses were based on a type I error of 5% and on the variance of all trials. Similar to the meta-analysis, effects of both non-covariate corrected as well as covariate corrected means as estimates of the effects of the current study were explored. In addition, resembling Jakobsen et al. (2012) bias risk was assessed with regard to sequence generation, allocation concealment, intention-to-treat analysis, blinding, drop-out, outcome measure reporting, presence of economic- and academic bias (see data supplement V for a full description of the criteria). An independent rater checked the generated table entries for accuracy.

Footnotes

[1] The baseline covariates BDI-II and EQ5D were significantly correlated to all outcome measures (BDI-II, EQ5D, BSI, WSAS and RAND36) at 7 and 12 months (Pearson's r ranging from -36 to .50. at $p = .01$).

Data Supplement II

Data Supplement II. Final BDI-II model in the Treatment Phase (0–7 months) controlled for potential confounders.

	B	95% CI	F	d.f.	r*	p
Final Model						
Intercept	2.71	[2.59– 2.83]	37.59	274	0.35	<.001
Baseline Severity	0.35	[0.28– 0.43]	88.74	274	0.49	<.001
Baseline Quality of Life	-0.10	[-0.17– -0.02]	6.30	274	0.15	.01
Time	-0.02	[-0.03– -0.02]	59.81	274	0.42	<.001
Condition	-0.01	[-0.25– 0.23]	0.01	274	0.01	.94
Time x Condition	0.00	[-0.01– 0.02]	0.38	274	0.04	.54
Sessions						
Intercept	2.71	[2.59– 2.83]	31.48	273	0.32	<.001
Baseline Severity	0.35	[0.28– 0.43]	86.81	273	0.49	<.001
Baseline Quality of Life	-0.10	[-0.17– -0.02]	6.19	273	0.15	.01
Time	-0.02	[-0.03– -0.02]	59.49	273	0.42	<.001
Condition	-0.01	[-0.25– 0.23]	0.01	273	0.01	.95
Time x Condition	0.00	[-0.01– 0.02]	0.36	273	0.04	.55
Number of sessions	0.05	[-0.03– 0.12]	1.55	273	0.08	.21
Therapist						
Intercept	16.05	[9.88–22.22]	15.64	267	0.24	<.001
Baseline Severity	6.34	[4.77– 7.91]	63.37	267	0.44	<.001
Baseline Quality of Life	-2.27	[-3.83– -0.71]	8.19	267	0.17	.01
Time	-0.40	[-0.56– -0.36]	81.13	267	0.48	<.001
Condition	0.58	[-9.24–10.40]	0.01	267	0.01	.91
Time x Condition	0.06	[-0.14– 0.26]	0.30	267	0.03	.58
Therapist	-	-	0.35	267	0.04	.93
Employment						
Intercept	2.71	[2.59– 2.83]	32.67	271	0.33	<.001
Baseline Severity	0.36	[0.29– 0.43]	92.81	271	0.51	<.001
Baseline Quality of Life	-0.09	[-0.16– -0.01]	5.51	271	0.14	.02
Time	-0.03	[-0.03– -0.02]	59.60	271	0.42	<.001
Condition	-0.00	[-0.24– 0.24]	0.00	271	0.00	.99
Time x Condition	0.00	[-0.01– 0.02]	0.31	271	0.03	.58
Employment	0.11	[-0.03– 0.25]	2.32	271	0.09	.13

Data Supplement II. (continued)

	B	95% CI	F	d.f.	r*	p
Gender						
Intercept	2.69	[2.57– 2.81]	31.25	273	0.32	<.001
Baseline Severity	0.35	[0.27– 0.42]	84.60	273	0.49	<.001
Baseline Quality of Life	-0.09	[-0.17– -0.02]	5.80	273	0.14	.02
Time	-0.02	[-0.03– -0.02]	60.83	273	0.43	<.001
Condition	0.00	[-0.24– 0.24]	0.00	273	0.00	.97
Time x Condition	0.00	[-0.01– 0.02]	0.39	273	0.04	.53
Gender	0.09	[-0.06– 0.25]	1.48	273	0.07	.23
Partner						
Intercept	2.71	[2.59– 2.83]	31.35	273	0.32	<.001
Baseline Severity	0.34	[0.28– 0.43]	88.58	273	0.49	<.001
Baseline Quality of Life	-0.10	[-0.17– -0.02]	6.27	273	0.15	.13
Time	-0.02	[-0.03– -0.02]	60.27	273	0.43	<.001
Condition	-0.01	[-0.25– 0.23]	0.01	273	0.01	.93
Time x Condition	0.00	[-0.01– 0.02]	0.37	273	0.04	.54
Partner	0.02	[-0.12– 0.17]	0.08	273	0.02	.78

Note: Mixed Models Gamma Regression with log-link on Beck Depression Inventory-II (BDI-II) + 1; Baseline Severity is standardised BDI-II score at baseline; Baseline Quality of Life is standardised EQ5D utility score at baseline; Condition is CT vs. IPT, centered at -.5 and .5 respectively; Time is the linear trend in weeks, with week = 0 at 7 months; Time effects represent change from baseline to 7 months, Condition main effects represent the difference between conditions at 7 months; * = Effect size $r = \sqrt{(F/(F+d.f.))}$.

Data Supplement III

Data Supplement III. Final BDI-II model in the Trial FU Phase (7–12 months) controlled for potential confounders.

	B	95% CI	F	d.f.	r*	p
Final Model						
Intercept	2.72	[2.60– 2.85]	10.72	758	0.12	<.001
Baseline Severity	0.34	[0.21– 0.46]	27.38	758	0.19	<.001
Baseline Quality of Life	-0.18	[-0.31– -0.06]	8.27	758	0.10	<.001
Time	0.00	[-0.00– 0.00]	0.12	758	0.01	.73
Condition	0.05	[-0.20– 0.30]	0.14	758	0.01	.71
Time x Condition	0.01	[-0.00– 0.01]	2.06	758	0.05	.15
Sessions						
Intercept	2.72	[2.60– 2.85]	8.86	757	0.11	<.001
Baseline Severity	0.34	[0.21– 0.46]	26.24	757	0.18	<.001
Baseline Quality of Life	-0.18	[-0.31– -0.06]	8.17	757	0.10	<.001
Time	0.00	[-0.03– 0.00]	0.11	757	0.01	.74
Condition	0.05	[-0.21– 0.30]	0.14	757	0.01	.71
Time x Condition	0.01	[-0.00– 0.01]	2.06	757	0.05	.15
Number of sessions	0.00	[-0.13– 0.13]	0.00	757	0.00	.98
Therapist						
Intercept	13.43	[5.28–21.58]	3.40	751	0.07	<.001
Baseline Severity	4.10	[2.05– 6.15]	15.45	751	0.14	<.001
Baseline Quality of Life	-3.05	[-5.07– -1.04]	8.83	751	0.11	<.001
Time	0.01	[-0.05– 0.06]	0.04	751	0.01	.85
Condition	-2.65	[-15.81–10.50]	0.16	751	0.01	.69
Time x Condition	0.07	[-0.04– 0.19]	1.54	751	0.05	.22
Therapist	-	-	0.20	751	0.02	.99
Employment						
Intercept	2.72	[2.60– 2.85]	9.93	751	0.11	<.001
Baseline Severity	0.34	[0.22– 0.47]	29.05	751	0.19	<.001
Baseline Quality of Life	-0.17	[-0.30– -0.04]	6.89	751	0.10	.01
Time	0.00	[-0.00– 0.00]	0.05	751	0.01	.82
Condition	0.08	[-0.18– 0.33]	0.36	751	0.02	.55
Time x Condition	0.01	[-0.00– 0.01]	2.72	751	0.06	.10
Employment	0.22	[-0.02– 0.46]	3.14	751	0.06	.08

Data Supplement III. (continued)

	B	95% CI	F	d.f.	r*	p
Gender						
Intercept	2.70	[2.57– 2.84]	9.02	757	0.11	<.001
Baseline Severity	0.33	[0.21– 0.46]	26.45	757	0.18	<.001
Baseline Quality of Life	-0.19	[-0.31– -0.06]	8.27	757	0.10	<.001
Time	0.00	[-0.00– 0.01]	0.19	757	0.02	.66
Condition	0.06	[-0.20– 0.32]	0.21	757	0.02	.65
Time x Condition	0.01	[-0.00– 0.01]	1.99	757	0.05	.16
Gender	0.11	[-0.14– 0.37]	0.75	757	0.03	.39
Partner						
Intercept	2.73	[2.60– 2.86]	8.89	757	0.11	<.001
Baseline Severity	0.34	[0.21– 0.46]	27.21	757	0.19	<.001
Baseline Quality of Life	-0.19	[-0.31– -0.06]	8.31	757	0.10	<.001
Time	0.00	[-0.00– 0.01]	0.12	757	0.01	.73
Condition	0.05	[-0.21– 0.30]	0.13	757	0.01	.72
Time x Condition	0.01	[-0.00– 0.01]	2.04	757	0.05	.15
Partner	0.03	[-0.22– 0.27]	0.04	757	0.01	.84

Note: Mixed Models Gamma Regression with log-link on Beck Depression Inventory-II (BDI-II) + 1; Baseline Severity is standardised BDI-II score at baseline; Baseline Quality of Life is standardised EQ5D utility score at baseline; Condition is CT vs. IPT, centered at -.5 and .5 respectively; Time is the linear trend in weeks, with week = 0 at 12 months; Time effects represent change from 7 to 12 months, Condition main effects represent the difference between conditions at 12 months; * = Effect size $r = \sqrt{(F/(F+\text{d.f.}))}$.

Data Supplement IV

Data supplement IV. Observed and Mixed Regression based estimated Response and Remission rates (%) in the active conditions based on the BDI-II ($n = 151$).

	Observed Values		Mixed Regression Based Estimates						
	CT	IPT	CT	IPT	B	95% CI	F	d.f.	p
	($n = 76$)	($n = 75$)	($n = 76$)	($n = 75$)					
Response									
3 months	40.8%	44.0%	40.5%	45.0%	-0.05	[-0.22–0.12]	0.28	575	.60
7 months	68.4%	66.7%	69.5%	69.1%	0.00	[-0.15–0.16]	0.00	575	.96
9 months	57.9%	61.3%	58.1%	63.7%	-0.06	[-0.22–0.11]	0.45	575	.50
12 months	60.5%	54.7%	60.1%	57.4%	0.03	[-0.14–0.19]	0.10	575	.76
Remission									
3 months	11.8%	13.3%	2.0%	5.1%	-0.03	[-0.07–0.01]	2.47	575	.12
7 months	34.2%	34.7%	26.9%	36.5%	-0.10	[-0.27–0.08]	1.21	575	.27
9 months	36.8%	37.3%	31.3%	40.7%	-0.09	[-0.26–0.08]	1.20	575	.28
12 months	42.1%	32.0%	36.3%	34.1%	0.02	[-0.15–0.19]	0.06	575	.81

Note: Binary Logistic Mixed Model Regression on Response/Remission=yes/no; All models are controlled for standardised baseline BDI-II and EQ5D scores, Time*baseline BDI-II, and Condition*Time; Condition is CT vs. IPT, centered at -.5 and.5 respectively; Time ranges from 2 – 5 and represents 3, 7, 9, 12 month assessment; CT=Cognitive Therapy; IPT=Interpersonal Psychotherapy; BDI-II=Beck Depression Inventory Second Edition; Response=A decrease of at least 9 BDI-II points from baseline to the specific assessment point; Remission=Absolute BDI-II score of 9 or less (Jacobson & Truax, 1991). Data unavailable for 3, 17, 23, and 25 patients at 3, 7, 9, and 12 months respectively.

Data Supplement V

Data supplement V. Assessment of bias risk (using the method of Jakobsen et al. 2012).

Criterion	Description*	Yes/No	Line of reasoning (source: Lemmens et al. 2011; current study)
Sequence Generation	Description of a random component in the sequence generation process. Such as: referring to a random number table.	Y	'Randomization took place at the research centre. The researcher pressed the 'assign' button on the computer screen, after which the database randomly allocated the participant to one of three conditions using computer-generated block randomization (10:10:4) The random allocation sequence was generated by an independent computer scientist and concealed from the researchers that were involved in the randomization procedure in order to prevent prediction of future assignment.'
Allocation Concealment	Participant and investigators enrolling participants could not foresee assignment because allocation was concealed by central allocation, or sequentially number drug containers or envelopes, or an equivalent method.	Y	'Randomization took place at the research centre. The researcher pressed the 'assign' button on the computer screen, after which the database randomly allocated the participant to one of three conditions using computer-generated block randomization (10:10:4) The random allocation sequence was generated by an independent computer scientist and concealed from the researchers that were involved in the randomization procedure in order to prevent prediction of future assignment.'
ITT analysis	Patients are analysed according to the randomization scheme. In other words, for the purposes of ITT analysis, everyone who begins the treatment is considered to be part of the trial, whether he or she finishes it or not.	Y	'All analyses were intention-to-treat, meaning that all patients that enrolled in the study were included in the analyses, irrespective of completing therapy or assessments'

Data supplement V. (continued)

Criterion	Description*	Yes/No	Line of reasoning (source: Lemmens et al. 2011; current study)
Blinding	Knowledge of the allocated interventions was adequately prevented during the study. As indicated by one of three criteria 1) no blinding, but no consequences by lack of blinding on outcome (measures); 2) blinding and unlikely that blinding could have been broken, 3) only partial blinding, but not likely to introduce bias.	Y	*'With regard to the nature of interventions, blinding of patients and therapists for treatment condition was not possible'.* However, we think it is unlikely that the lack of blinding has influenced outcome, mainly because all outcome measures were self-report measures, and patients were not aware of study aims. However, the fact that the researchers who conducted statistical analyses were not blind for the coding of CT and IPT is a limitation of the current study.
Comparability of drop-outs in intervention groups?	Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups	Y	*'No significant differences in attrition rates emerged across conditions (See Consort Flow Chart)'.* Missing outcome data is balanced in numbers across intervention groups. At 7 months (the end of the acute phase), 6 patients in CT and 10 in IPT were lost to follow-up. At 12 months this was 11 (14.5%) for CT ad 14 (18.7%) for IPT. Reasons for drop-out were similar across groups. Patients either were unattainable/ did not respond to contact requests (7 in CT vs. 8 in IPT), or no longer wanted to participate in the trial (3 in CT vs. 6 in IPT). 1 moved abroad (CT). Even though the 12-month attrition rates are within (the low) range of other clinical trials, they might have introduced bias, which could be a limitation of the study. However, we do not consider it likely that the drop-out rates have caused any large biases because missings were handled carefully. By using mixed regression (a method that takes the nested structure of the data into consideration and can deal with autocorrelation and missing values, see Singer & Willet, 2003, Oxford University Press).
Free of selective outcome measure reporting	The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way	Y	We pre-specified all of the study's outcomes in our protocol paper (Lemmens et al., 2011). As can be seen in the protocol paper, we included several categories of measurements: primary and secondary outcome measures (in terms of symptoms and quality of life), process measures, and economic evaluation measures. The present study examines the clinical effectiveness, and therefore included all clinical outcome and quality of life measures.

3

Data supplement V. (continued)

Criterion	Description*	Yes/No	Line of reasoning (source: Lemmens et al. 2011; current study)
Free of Economic bias	Economic bias may be present if a trial is financed by an individual or organization that might have an interest in a given result from the trial	Y	*This research was financed by the research institute of Experimental Psychopathology (EPP) and the Academic Community Mental Health Centre (RIAGG). Both organizations have no special interests in specific outcomes of the trial.' EPP is an independent non-profit research institute that aims to unravel underlying mechanisms of the aetiology, maintenance and treatment of select psychopathological conditions. EPP has created an excellent track record in a continuum from basic to applied clinical research. RIAGG is a research oriented routine clinical setting in Maastricht, the Netherlands, providing EPP's researchers with a stable clinical infrastructure where long-term studies can be executed. RIAGG has no interest in specific outcomes of the trial, since both CT and IPT are already delivered as routine psychological therapies for depression in the clinic. We therefore feel free to say that our trial free of industry sponsorship or other types of for-profit support that may manipulate the trial design, conductance or results of the trial.*
Free of Academic bias	Academic bias may be present if one or more of the trialists have an academic or personal interest in a given result from the trial	Y	*'The authors declare to have no conflict of interest, or personal gain'*

Note: *Definitions come from Higgins & Green (2011) 'The Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0'. Obtained via the Cochrane Hepato-Biliary group's (CHBG) website http://hbg.cochrane.org/information-authors

4

The value of an implicit self-associative
measure specific to core beliefs of depression

Abstract

The present study examined differences in explicit and implicit measures of self-esteem between depressed patients and healthy controls using an indirect measurement procedure especially adapted to measure self-esteem aspects of core beliefs of depression. Furthermore, we examined whether our implicit and explicit self-associative measures were associated with each other and with depressive symptoms, and investigated the effect of a discrepancy between the implicit and explicit measure on depression.Participants were 87 depressed patients and 30 healthy controls. The Self-Liking and Self-Competence Scale was administered as a measure of explicit self-esteem. A depression-specific variant of the Single Category Implicit Association Test served as a measure of implicit self-esteem. Patients showed significantly lower levels of explicit self-esteem as compared to healthy controls. In spite of our adaptations, no differences were found on the implicit measure. The implicit measure of self-esteem was neither related to the explicit measure nor to depressive symptoms. Furthermore, although both the explicit measure of self-esteem and the difference score of the explicit and implicit measure were related to symptoms of depression, the relation between the explicit measure and depression was found to be significantly stronger. This study suggests that only the explicit measure of self-esteem–and not the implicit–is related to depression. Results should be interpreted with caution because it is not clear yet to what extent implicit measures really reflect self-esteem. Future research using well-designed measurement procedures for obtaining implicit and explicit measures could contribute to a better insight in the nature of these constructs.

4

Introduction

According to cognitive theory, the self-concept is negatively biased in depressed patients (Beck, 1987; Greenwald & Farnham, 2000). Depressed patients tend to think more negatively about themselves, and report lower self-esteem than healthy controls do (Ingram, Miranda, & Segal, 1998). This reduces subjective well-being. Cognitive theory states that individuals at risk for depression developed latent negative schemas about the self, the world and the future. More specifically, depressogenic core beliefs about the self are categorized as beliefs related to unlovability, helplessness, and incompetence (Beck, Rush, Shaw, & Emery, 1979). These schemas come to the surface in periods of stress and influence the responses to stressful life circumstances (Beck, 1987; Clark, Beck, & Alford, 1999). However, the relation between negative core beliefs about the self and depressive symptoms has not been elucidated yet, and the current state of the art in this field is mainly based on research using explicit measures. Explicit measures of self-esteem reflect rational and conscious processing of self relevant stimuli. However, in the past decades, an increasing number of researchers have acknowledged that self-schemata may also be reflected by more automatic and intuitive processing of affective experiences (Clark et al., 1999; Haaga, Dyck, & Ernst, 1991; Steinberg, Karpinski, & Alloy, 2007). To the extent that implicit measures reflect uncontrollable, unaware and fast mechanisms, they could reveal insights beyond those of explicit measures. Furthermore, the use of explicit measures has methodological disadvantages such as reliance on introspection capacities and socially desirable answering tendencies, which might bias the outcomes. Implicit measures partly overcome these drawbacks because they do not rely on introspection and participants only have limited possibilities to respond in a socially desirable way (De Houwer, 2006; Glashouwer & de Jong, 2008; Steffens, 2004).

As a result, the interest for the use of implicit measures in research on vulnerability to depression has grown, and various measurement procedures have been developed to obtain implicit measures of self-esteem (e.g. the Word Completion Task (WCT; Bosson, Swann, & Pennebaker, 2000); the Name Letter Task (NLT; Nuttin, 1985); and the Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001)), with the Implicit Association Task (IAT: Greenwald & Farnham, 2000) being the most frequently used in this field. As recommended by De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009), it is useful to distinguish between indirect measurement procedures and measures. A measurement procedure is the specific methodology; the set of guidelines followed that lead to an outcome. A measure is defined as 'a measurement outcome that is causally produced by the to-be-measured attribute in the absence of certain goals, awareness, substantial cognitive resources, or substantial time' (p.350). In this article, we will use the term 'implicit measure' to refer to the outcome of the Single Category Implicit Association Test (Sc-IAT: Karpinski & Steinman, 2006) specific to measure core beliefs of depression.

The introduction of implicit measures in the field of self-esteem initially led to a debate about which measure reflected a person's 'true' attitude (e.g. Fazio & Olson,

2003). However, dual-process models (Epstein, 1994) showed that both implicit and explicit cognitive processes influence depressive reactions to stressful life events, but do so in different ways (e.g. Beevers, 2005; Haeffel et al., 2007). The current thinking therefore treats implicit and explicit measures as complementary rather than as competitors, and consensus has been reached about the fact that using both measures can contribute to a deeper understanding of self-esteem (De Houwer, 2006; De Houwer et al., 2009; De Raedt, Schacht, Franck, & De Houwer, 2006; Karpinski & Steinman, 2006; Roefs et al., 2011).

Even though interest has grown, research with implicit measures of self-esteem is relatively scarce. Remarkably, the results of the few studies that actually obtained implicit measures of self-esteem in the context of depression often contradict results obtained by explicit measures. Although it is consistently shown in research using explicit measures that depressed patients have a more negative self-image than healthy controls (e.g. Hollon, Kendall, & Lumry, 1986; Silverman, Silverman, & Eardley, 1984; Xi, Zhang, & Li, 2007), up until now only two studies found evidence for decreased self-esteem on implicit measures in currently depressed patients (Glashouwer & de Jong, 2010; Risch et al., 2010). The vast majority of studies that obtained implicit measures of self depressive associations in depressed patients found evidence for positive self-esteem in both healthy individuals and in depressed patients, regardless of the measurement procedure that was used (De Raedt et al., 2006; Franck, De Raedt, & De Houwer, 2007; Franck, De Raedt, & De Houwer, 2008; Franck, De Raedt, Dereu, & van den Abbeele, 2007; Gemar, Segal, Sagrati, & Kennedy, 2001; Valiente et al., 2011).

The observation that implicit and explicit measures consistently show diverging results is at least unexpected, and might suggest that they reflect different constructs. Several research groups explain these findings as an indication that not the absolute levels of implicit and explicit measures of self-esteem separately, but the discrepancy between them (either expressed as an interaction or difference score of the explicit and implicit measure) plays a crucial role in the cause and maintenance of depression, because it represents deficiencies in the integration of self-representations, which is related to internalizing problems (e.g. Bosson, Brown, Zeigler-Hill, & Swann, 2003; Creemers, Scholte, Engels, Prinstein, & Wiers, 2012; Franck, De Raedt, & De Houwer, 2007; Schröder-Abé, Rudolph, & Schütz, 2007). Moreover, because these findings are consistent but not in line with cognitive theory of depression, one could even argue that the theory needs to be adapted. However, because measurement procedures for obtaining implicit measures are relatively young and the best way to obtain an implicit measure of self-esteem is still unclear (Garety & Freeman, 1999; McKay, Langdon, & Coltheart, 2007), these conclusions might be premature and it might be too soon to question the validity of cognitive theory. Another, in our view more plausible, reason for these contradictory findings between results obtained by implicit and explicit measures might be the lack of specificity of the instruments that have been used. Further research would benefit from careful design of measurement procedures to obtain implicit measures, paying attention to factors such as the exact choice of paradigm, and the selection of stimuli.

With regard to the selection of stimuli, it would make sense to design a task that is more specifically designed to measure depressogenic core beliefs related to self-esteem: unlovability, helplessness, and incompetence (Beck et al., 1979). This might lead to larger contrasts between the groups because specific beliefs are only expected to be revealed in depressed patients and not in healthy controls. Several studies have already used target words related to depression (e.g., Franck, De Raedt, Dereu, et al., 2007; Glashouwer & de Jong, 2010; Risch et al., 2010), and found results that were contradictory to each other and (in some cases) to cognitive theory. It has to be noted that these studies chose a different focus in the selection of stimuli. Glashouwer and de Jong (2010) and Risch et al. (2010) used words that were related to the general concept of depression (for example 'fragile', 'pessimistic', 'negative' vs. 'active', 'cheerful', 'lively'), whereas Franck, De Raedt, Dereu, et al. (2007) specifically focused on 'worthlessness', one of the core concepts of depression (e.g., 'capable', 'valuable' vs. 'inferior', 'worthless').

Furthermore, these studies (Franck, De Raedt, Dereu, et al., 2007; Glashouwer & de Jong, 2010; Risch et al., 2010) used an original IAT paradigm. A drawback of the original IAT is that it can only provide information regarding the relative strength of various associations ('me' compared to 'other'). Because the IAT opposes the 'self' category with 'other', the strength of the association between self and attributes is partially biased by the strength of the association between the category 'other' and attributes. The IAT is therefore unable to indicate the strength of the evaluation of 'me' or 'other' separately. To avoid this potential contamination, and measure the associations with the self in isolation, a variant of the IAT: the single category IAT (Sc-IAT: Karpinski & Steinman, 2006) can be used. Because the Sc-IAT has no reference group 'other', it is able to measure the absolute strength of associations between the target category (i.e., 'self' in our study) and the attributes.

The central goal of the present study was therefore to optimize our implicit measure by not only adapting the measurement procedure to reflect the depressogenic core beliefs, but also by using a Sc-IAT paradigm. To our knowledge, this is the first study using a single category measurement procedure exclusively adapted to this target group. By comparing scores on implicit and explicit measures of self-esteem of depressed patients and healthy controls, the hypothesis was tested that depressed patients, who were about to start psychotherapy, would show more negative self-associations than would healthy control participants. Furthermore, we examined whether our implicit and explicit self-associative measures were associated with each other and with depressive symptoms. In addition, even though the discrepancy hypothesis arose as a post-hoc explanation, it has now been used as a theoretical explanation of depression by several research groups (e.g., Creemers et al., 2012; Franck, De Raedt, & De Houwer, 2007, Franck, De Raedt, Dereu, et al., 2007), and warrants further examination. We therefore investigated the effects of the discrepancy between scores on the implicit and explicit measure of self-esteem on depression, using two different methods to assess discrepancy that are used in current literature (i.e., the interaction and the difference score of the explicit and the implicit measure).

Method

Participants

The clinical sample consisted of 87 depressed patients (69% female). They were between 18 and 62 years old (M=42.30, SD=10.87) and 41.4% was educated on medium level (low: 25.3%, high: 26.4%). Their average score on the Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996) was 27.24 (SD=9.04). The healthy control sample consisted of 30 participants (67% female), who ranged in age from 18 to 65 years (M=44.53, SD=15.13), and had BDI-II scores ranging from 0 to 9 (M=3.17, SD=2.67). The majority of the controls (56.7%) were educated on medium level (low: 10.0%, high: 26.7%). No significant differences were found between groups on age (t (40)=0.75, p=.46), gender (χ^2 (1, n=117)=0.06, p=.82), and level of education (F (1, 109)=0.51, p=.48).

Recruitment and procedure

Patients were recruited as part of a large ongoing randomized controlled trial aimed at the effectiveness, mechanisms of change and relapse prevention of Cognitive Therapy (CT) and Interpersonal Psychotherapy (IPT) for depression. The study is registered at the Netherlands Trial Register, part of the Dutch Cochrane Centre (ISRCTN67561918). The main design of the study is fully described elsewhere (Lemmens et al., 2011). Patients were adult outpatients (18–65 years) referred to the mood disorder unit of the Maastricht Community Mental Health Centre with a primary diagnosis of Major Depressive Disorder (MDD) confirmed by the Structured Clinical Interview for DSM-IV Axis I disorders (SCID-I; First, Spitzer, Gibbon & Williams, 1997). Further inclusion criteria were: internet access, an e-mail address, and sufficient knowledge of the Dutch language. Exclusion criteria were: bipolar or chronic (> 5 years) depression, current use of antidepressant medication, drugs- and alcohol abuse/dependence, and mental retardation (IQ <80). Healthy participants were recruited in the general population. After ruling out the presence of depressive symptoms (BDI-II < 10) and other psychopathology in a telephone screening, participants were invited.

After informed consent was given, the assessment started with a brief verbal description of the procedure. All participants were told that they were going to perform a word-sorting task that required concentration, and that they subsequently would answer several questionnaires about aspects of psychological functioning. Healthy participants were debriefed afterwards. For depressed patients, debriefing took place after completing the total study.

Measures

Beck Depression Inventory-II

The Beck Depression Inventory-II (BDI-II; Beck et al., 1996; Dutch translation by van der Does, 2002) is a 21-item self-report questionnaire used to measure the severity of depression. Each item is rated on a 4-point Likert-scale ranging from 0 to 4. Higher scores indicate higher levels of depression. The questionnaire has strong psychometric properties as a screening measure for depression (Beck et al., 1996; van der Does, 2002).

Depression specific Single Category Implicit Association Test

The Single Category Implicit Association Test (Sc-IAT; Karpinski & Steinman, 2006) is an adaptation of the original IAT (Greenwald & Farnham, 2000; Greenwald, McGhee, & Schwartz, 1998) and was used to obtain an implicit measure of specific core beliefs of depression exclusively related to the concept of the self. Instruments such as the Sc-IAT measure associations between the self and certain attributes, which are assumed to be an indication of self-esteem. The task consisted of 3 blocks of trials and started with a practice block. In the practice block, 12 positive and 12 negative words (attributes) were presented, and the task for participants was to categorize these attributes as such by pressing the appropriate key (left or right shift key marked with a blue sticker) as quickly as possible without making too many errors. After the practice block, there were two test blocks. In the test blocks, a third word category was added to the positive and negative attributes: stimuli representing the self (self-stimuli). In one test block, self-stimuli shared a response key with positive attributes, while the other key needed to be pressed for negative attributes ('pos + me' block). In the other test block, the key assignment for self-stimuli was switched. They now shared a response key with the negative attributes and the other key needed to be pressed for positive words ('neg + me' block). Each test block consisted of 12 practice trials (7 attributes and 5 self-stimuli) immediately followed by 72 test trials: 30 self-stimuli and 42 attributes. The 42 attributes consisted of 30 negative and 12 positive trials for the 'pos + me' blocks and vice versa for the 'neg + me' blocks. Each block was preceded by a set of instructions concerning the dimension(s) of the categorization task and the appropriate key response. The assignment of positive and negative attributes to the right and left key and the order of test blocks were balanced over participants. Because the Sc-IAT has only three categories, it was not possible to have both an equal number of attributes per category (pos/neg) and an equal number of stimuli per response key. However, the Sc-IAT score was computed by comparing the two phases (me-pos and me-neg) and in both of these phases the distribution over response buttons was equal for all participants. During the task, attributes (pos/neg) and self-stimuli appeared on the computer screen one by one, and in a random order unique for each participant.

The rationale behind the Sc-IAT is that the response time is expected to be faster when two compatible concepts are mapped together on the same key, as compared to when two less compatible concepts are mapped together. For example, for people with positive self-esteem, when the categories 'positive' and 'me' are mapped onto the same key, the response time is predicted to be faster than for the reversed combination ('negative' and 'me'; Karpinski & Steinman, 2006).

A total of 18 different stimuli were used; 6 positive and 6 negative attributes and 6 self-related target stimuli. The Dutch words for 'I' (ik), 'me' (me), 'self' (zelf), 'my' (mij), 'own' (eigen), and 'mine' (mijn) served as self-related target stimuli. As has been noted earlier, all attributes referred to one of three main cognitive themes in depression according to Beck (1987). Negative words that represented the concept of unlovability were the Dutch words for 'rejected' (afgewezen) and 'harmful' (schadelijk). The Dutch words for 'worthless' (waardeloos) and 'redundant' (overbodig) were selected to represent the core belief worthlessness, and the Dutch translation of 'helpless' (hulpeloos) and 'vulnerable' (kwetsbaar) served as a representation of the concept of helplessness. The positive words were chosen as counterparts of the negative words. These were the Dutch words for 'beloved' (geliefd) and 'pleasant/nice' (aangenaam) for unlovability, 'powerful' (krachtig) and 'successful' (geslaagd) for helplessness, and 'perfect' (perfect) and 'valuable' (waardevol) for worthlessness. All attribute stimuli were derived from the list of Hermans and De Houwer (1994) and were matched on subjective familiarity by ratings of 25 University employees. The final positive and negative word sets were found to be statistically equivalent with regard to familiarity (t (10) = 1.24, p = .24) but showed a small difference with regard to worth length (t (10) = 4.03, p < .01). Because in the Dutch language negative words are often composed by the positive word plus a negation (e.g. the counterpart for the positive word 'waarde-vol' is 'waarde-loos'), negative words (M = 9.33, SD = .52) consisted on average of more letters than did the positive words (M = 7.83, SD = .75). Stimulus presentation was controlled by Inquisit Millisecond Software (v. 2.0). The background colour of the computer screen was white and all stimulus words were individually displayed in the centre of the screen in black ARIAL font 48 letters. During the task the target-labels ('positive', 'negative' and 'me') were displayed at the top of the screen.

Self-Liking and Self-Competence Scale–Revised

The Self-Liking and Self-Competence Scale Revised (SLCS-R, Tafarodi & Swann, 2001; Dutch translation by Vandromme, Hermans, Spruyt & Eelen, 2007) was used as an explicit measure of self-esteem. This 16-item self-report questionnaire is developed to measure Self-Liking and Self-Competence, two important dimensions of self-esteem. The subscale Self-Liking measures the subjective evaluation of personal worth. The subscale Self-Competence measures the sense of one's capability of successful intentional goal

pursuit, derived from multiple experiences (Tafarodi & Swann, 2001). The SLSC-R contains eight items for each of the two subscales (e.g. Self-Competence; 'I am highly effective at the things I do' and Self-Liking: 'I tend to devalue myself'). Items are rated on a 5-point Likert-scale and subscale scores can range from 8 to 40, with higher scores indicating higher Self-Competence or higher Self-Liking. Following Hermans et al. (2008) we also calculated a total score by adding up the scores of the two subscales (range 16–80)[1]. Psychometric properties of the Dutch version of the SLCS-R have been described as good (Vandromme, Hermans, Spruyt, & Eelen, 2007).

Statistical analysis

Only cases with complete data were entered in the analyses. Originally, 106 depressed patients and 34 healthy controls were tested. 19 depressed participants had missing data and were therefore excluded from analyses. Furthermore, four controls appeared to have a BDI-II score of 10 or higher, which might indicate mild depressive symptoms, and were therefore also excluded. In total, data of 87 depressed patients and 30 healthy controls were used in the analyses. The Sc-IAT effect was calculated using the improved D600-score algorithm (Greenwald, Nosek, & Banaji, 2003), with higher scores indicating a tendency to associate the self more strongly with positive words than with negative words. Independent samples *t*-tests were used to examine differences between patients and controls on the SLSC-R total- and subscale scores, and the Sc-IAT. Pearson's correlations were calculated to investigate the relationship among the implicit and explicit measure.

To further examine the relationship between the explicit and implicit measure of self-esteem, their discrepancy, and depression regression analysis was used. We started with a multiple hierarchical regression model with depression severity being the dependent variable and centered total scores on the implicit and explicit measure being potential predictors. In line with previous research by for example Franck, De Raedt, and De Houwer (2007) and Schröder-Abé et al. (2007), we examined main effects of both self-esteem measures (centered scores) on depressive symptoms in step 1 and checked for the effect of the discrepancy between both variables expressed as a centered interaction score (c_implicit * c_explicit) in step 2 using the procedure proposed by Aiken and West (1991).

Furthermore, we examined the discrepancy expressed as a difference score between implicit and explicit measures of self-esteem. Following studies by for example Creemers et al. (2012) and Kesting, Mehl, Rief, Lindenmeyer, and Lincoln (2011), we subtracted the standardized score on the explicit measure from the standardized score on the implicit measure (z_implicit e z_explicit), and examined the effect of this discrepancy score on depression in a separate regression analysis[2]. In order to shed more light on the magnitude of the effect of this second method, the difference score model was compared with the main effect models obtained in step 1 of the hierarchical regression analysis.

Results

Table 1 shows the descriptive statistics of patients and controls and the differences between the groups on the self-associative measures. Analysis showed significant differences between patients and controls on the explicit measure of self-esteem: patients reported significantly lower scores as compared to healthy controls. On the implicit measure, the group as a whole (i.e., depressed patients and healthy controls) showed significantly stronger associations between positive attributes and self-related stimuli than between negative attributes and self-related stimuli (t (116) = 3.12, p < .01). No significant differences in Sc-IAT scores were found between the patients and healthy control participants. Furthermore, there was a significantly larger difference between score on the implicit and explicit measure of self-esteem in healthy controls than in patients. See Table 1 for relevant statistics.

Table 1. Mean (SD) scores and differences on the explicit and implicit measures of self-esteem between depressed patients and healthy controls.

	Patients (*n* = 87)		Controls (*n* = 30)		Independent samples T-test	
SLSC-R						
- Total Score	39.69	(8.92)	59.97	(7.88)	*t* =-11.05,	*p* =< .001
- Subscale Self-Liking	18.79	(5.20)	31.53	(4.85)	*t* =-11.77,	*p* =< .001
- Subscale Self-Competence	20.90	(4.74)	28.43	(3.87)	*t* = -7.85,	*p* =< .001
Sc-IAT	0.12	(0.45)	0.17	(0.42)	*t* = -0.56,	*p* = .58
SLSC-R total – Sc-IAT*	-39.57	(8.90)	-59.79	(7.92)	*t* = 11.02,	*p* =< .001

Note: BDI-II = Beck Depression Inventory-Second Edition; SLSC-R = Self-Liking and Self-Competence Scale Revised; Sc-IAT = Single Category Implicit Association Test D600–score; * = Mean (SD) standardized difference score (z_ SLSC-R total – z_Sc-IAT) = 0.39 (1.21) and -1.13 (1.18) for patients and controls respectively (t (115) = 5.95, p < .01).

The correlation between the total score on the implicit and explicit measure was found to be non-significant (r (117) = 0.06, p = .54). Correlation analyses for the two dimensions of the explicit measure separately showed similar results. In addition, there were no differences in correlational patterns between patients and controls. Results of the hierarchical regression analysis used to examine main effects of the self-esteem measures and their interaction on depression are displayed in Table 2. Results of step 1 indicate that lower scores on the explicit measure of self-esteem are associated with higher levels of depressive symptoms (β = .70, t (114) = 10.51, p < .001). No significant main effect was found on the implicit measure (β = .02, t (114) = 0.24, p = .81). The interaction term was

not significant, meaning that the discrepancy between the explicit and implicit measure of self-esteem score expressed as an interaction between the two variables, did not relate to depressive symptoms[3].

      With regard to the effects of the discrepancy score expressed as a difference between the standardized measures of implicit and explicit self-esteem (not displayed in Table 2), it was found that a larger discrepancy was significantly associated with symptoms of depression ($\beta = .47$, t (114) $= 5.714$, $p < .001$). In order to shed more light on the magnitude of this effect, the model was compared to the significant main effect model obtained in step 1 using the method of Steiger (1980). Analysis showed that the relation between the explicit measure of self-esteem and symptoms of depression was significantly stronger than the relation between the discrepancy score and depressive symptoms ($Z = -8.83$, $p < .01$).

Table 2. Associations of the explicit and implicit measures of self-esteem and their interaction with depressive symptoms.

4

	Beck Depression Inventory-II		
	B	SE	β
Step 1			
- SLSC-R	-0.75	0.07	-0.70*
- Sc-IAT	-0.49	2.00	-0.02
Step 2			
- SLSC-R x Sc-IAT	-0.00	0.17	0.00

Note: All independent variables are centered; SLSC-R = Self-Liking and Self-Competence Scale Revised Total Score; Sc-IAT = Single Category Implicit Association Test D600–score; $R^2 = 0.49$ in step 1 ($p = 0.01$); $\Delta R^2 = 0.49$ in step 2 ($p = 0.99$); * $= p < 0.001$.

Discussion

Main findings

The objective of the current study was to examine whether there are differences in scores on implicit and explicit measures of self-esteem between depressed patients and healthy control participants. Furthermore, we examined whether our implicit and explicit self-associative measures and their discrepancy (expressed as an interaction and difference score) were associated with symptoms of depression. It was expected that depressed patients would show reduced scores on both the implicit and explicit measure of self-esteem as compared to healthy controls, and that larger discrepancies between the two measures would be related to more severe depressive symptoms.

In line with prior research (e.g., Hollon et al., 1986; Silverman et al., 1984; Xi et al., 2007), expected differences were found on the explicit measure. However, in spite of our attempt to create larger contrasts between patients and controls by adapting our Sc-IAT specifically to measure depressive core beliefs, depressed patients did not score lower on this implicit measure of self-esteem than did healthy controls. These results resemble those of other studies that obtained implicit measures of self-esteem in depressed patients and healthy controls and also did not find differences (Creemers et al., 2012; De Raedt et al., 2006; Franck, De Raedt, & De Houwer, 2007, Franck, De Raedt, Dereu, et al., 2007; Gemar et al., 2001), and those who compared scores on both self-associative measures within depressed patient and found reduced levels of explicit, but normal levels of implicit self-esteem (Kesting et al., 2011; Valiente et al., 2011). Our results are not in line however, with studies conducted by Risch et al. (2010) and Glashouwer and de Jong (2010) who did find differences on implicit self-associative measures between depressive patients and healthy controls. Furthermore, our finding that the score on the implicit measure of self-esteem was not directly associated with the score on the explicit measure and with depressive symptoms, whereas explicit self-esteem showed a significant inverse relation with depression, is in line with previous research (e.g., Creemers et al., 2012; Kesting et al., 2011; Vázquez, Diez-Alegría, Hernández-Lloreda, & Moreno, 2008). The divergence between explicit and implicit measures was further underlined by a lack of correlations between these measures, which is also not uncommon in this field of research (see Roefs et al., 2011). In addition, even though we found that the discrepancy between the two self-associative measures expressed as an interaction score was not associated with depression, the discrepancy expressed as a difference score was significantly related to depressive symptoms. These results show that interchangeably using two different methods to assess the same construct might not only lead to confusion because it is not clear what the concept exactly means and how it is measured, but also leads to diverging results. A closer look at the significant relation between the difference score and depression showed that this relation was significantly weaker than the relation between the explicit measure of self-esteem and depression. This might shed more light on the nature of the relation between discrepancy and depression; since the discrepancy score is a composition of the scores on the explicit and implicit measure, and no relation was found between the implicit measure and depression, one can argue that the effect of the discrepancy score is solely a result from variability in the explicit measure of self-esteem made statistically weaker by subtracting the implicit score – which increased error variance.

Methodological considerations and recommendations

Although consistent with the majority of previous research, our results on the implicit measure contradict the cognitive theory of depression which states that the view of the self in general is negatively biased in depressed patients. It should be noted that findings from implicit measures often contradict the cognitive theory of the disorder (Roefs et

al., 2011). The question that remains is whether this is reason to call for an adjustment of cognitive theory or that one should doubt the validity of the implicit measures. Because it is still not clear yet in what sense procedures such as the Sc-IAT can be considered implicit and valid indices of self-esteem (Buhrmester, Blanton, & Swann, 2011; De Houwer, 2006), results obtained by implicit measures should be critically evaluated before one can start thinking about changing or rejecting theory.

First of all, one could argue that our findings are the result of the specific paradigm we used. Because self-esteem is assumed to be a complex, multi-dimensional construct, it is possible that our implicit measure of self-esteem did not capture crucial aspects of the construct, which might have biased the results. However, previous research has consistently shown positive associations with the self in depressed patients, regardless of the measurement procedure that was used (e.g. Creemers et al., 2012; De Raedt et al., 2006). Therefore, it seems unlikely that the results that we obtained are mainly a result of the specific measurement procedure that was used to obtain an implicit measure of self-esteem, especially considering that we adapted the paradigm as much as possible to reflect depressive core beliefs.

Second, the lack of a direct comparison between our measure and another existing implicit measure of self-esteem might limit the possibilities to determine if our adaptations provide an advantage over existing measures. The problem with this however, is that in the field of implicit measures there is no such thing as a gold standard yet, which makes the exact choice of paradigm to compare with arguable (Roefs et al., 2011). Furthermore, adding a second implicit measure within one sample would also create extra problems such as order- and learning effects. We therefore think that, given the current status of the field, face value comparisons of design and methodology across studies and populations is a valid option to compare results.

Third, it is generally agreed upon that the (Sc-)IAT effect is largely determined by the category labels that have been chosen (De Houwer, 2001; Fazio & Olson, 2003). Category labels are considered critical for the interpretation of stimulus items and thereby influence the implicit measure. One could therefore argue that the labels that were used in the current study ('positive' vs. 'negative') could have influenced the outcome, and that changing these category labels (for example into 'depressed' and 'elated' as in the study by Glashouwer & de Jong, 2010) could lead to different results. However, we think it is unlikely that this has driven the effect. First of all because various other studies using different category labels were also not able to show reduced levels of implicit self-esteem in depressed patients (e.g., Franck et al., 2007 studies). And second, because Risch et al. (2010) used similar labels and actually did find differences between patients and controls.

Fourth, the fact that Glashouwer and de Jong (2010) assessed a sample that was 25.5 times larger than ours (2981 vs. 117) and found hypothesized results, might indicate that differences in implicit measures of self-esteem between patients and controls do exist, but that we were not able to reveal these differences because of insufficient

power. However, we think that it is improbable that this has driven our effects, since our study has 98% power to detect a similar effect size ($d = .86$ at two-tailed $\alpha = .05$) as Glashouwer and colleagues.

More likely is the explanation that these differences are due to the choice of target words. It is remarkable that the two studies that were able to demonstrate lower levels on implicit self-associative measures (Glashouwer & de Jong, 2010; Risch et al., 2010) used target words that reflected more general automatic self-associations instead of specific depressogenic core beliefs related to self-esteem. This might even indicate that perhaps not self-esteem but other self-relevant attributes are essential in implicit processes in depression. It is therefore important to examine whether patients who implicitly associate the self with for example pessimism, sadness, lack of motivation and passiveness, are at larger risk for developing depression, or whether those who are depressed develop implicit self-depressive associations. Another important question for future research is whether these self-associations are of importance in the maintenance or relapse of depression.

Conclusions

In spite of the fact that our results do not fit with the dominant cognitive theory of depression, and despite a lack of clarity regarding the validity and implicit nature of the measurement outcome, the fact that the majority of studies failed to detect a depression specific effect involving implicit measures of self-esteem calls for interpretation. Perhaps most importantly, fundamental research on the validity and implicit character of these measures should progress, before firm conclusions can be drawn from implicit measures obtained in the domain of depression and self-esteem. Future research should focus on a further refinement of measurement procedures to obtain implicit measures. It is important to identify and select the best paradigm for the research question (De Houwer et al., 2009) by carefully evaluating whether a relative or absolute measure of associations is required, what target words fit the psychosocial attributes that are being studied best, and to what sense expected effects are thought to be implicit (Roefs et al., 2011). Only then can implicit measures prove to be a true and valuable addition to explicit measures, which have their own well-known limitations such as a reliance on introspection and sensitivity to self-presentational concerns.

Funding

Specific author contributions

AR programmed the Sc-IAT paradigm. LL and HvT carried out recruitment and data-collection. LL undertook the statistical analyses and drafted the manuscript in close collaboration with AR. All authors participated in the design of the study, contributed to the writing of the manuscript and have approved of the final manuscript.

Conflict of interest

The authors declare that they have no competing interests.

4

Footnotes

[1] Both SLSC-R subscales and the total score were reliable in terms of internal consistency; that is, Cronbach's α's were good (a's between .87 and .94). Furthermore, the two subscales were highly correlated (r (117) = 0.77, p < .01).

[2] Because the difference score is derived from the variance related to both the implicit and explicit measure (z_implicit e z_explicit), it was statistically impossible to control for the main effects of scores obtained on the explicit and implicit measures of self-esteem. In short, including both the difference score and main scores in one model would eliminate the effect of the difference score from the equation. As a result, we would then be testing a main effects model that is identical to the model that was already tested in step 1 of the regression analysis.

[3] Analysis of the SLSC-R total score and both dimensions of the SLSC-R (Self- Liking and Self-Competence) separately showed similar results.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* London: Sage.

Beck, A. T. (1987). Cognitive models of depression. *Journal of Cognitive Psychotherapy: An International Quarterly, 2,* 5–37.

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

Beck, A. T., Steer, R., & Brown, G. K. (1996). *Beck Depression Inventory II: Manual.* Boston, MA: Hartcourt Brace.

Beevers, C. G. (2005). Cognitive vulnerability to depression: a dual process model. *Clinical Psychology Review, 25*(7), 975-1002.

Bosson, J. K., Brown, R. P., Zeigler-Hill, V., & Swann, W. B. (2003). Self-enhancement tendencies among people with high explicit self-esteem: the moderating role of implicit self-esteem. *Self and Identity, 2,* 169–187.

Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: the blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79*(4), 631–643.

Buhrmester, M. D., Blanton, H., & Swann, W. B., Jr. (2011). Implicit self-esteem: nature, measurement, and a new way forward. *Journal of Personality and Social Psychology, 100*(2), 365–385.

Clark, D. A., Beck, A. T., & Alford, B. A. (1999). *Scientific foundations of cognitive theory and therapy of depression.* New York, NY: John Wiley and Sons.

Creemers, D. H. M., Scholte, R. H. J., Engels, R. C. M. E., Prinstein, M. J., & Wiers, R.W. (2012). Implicit and explicit self-esteem as concurrent predictors of suicidal ideation, depressive symptoms, and loneliness. *Journal of Behavior Therapy and Experimental Psychiatry, 43*(1), 638–646.

De Houwer, J. (2001). A structural and process analysis of the implicit association test. *Journal of Experimental Social Psychology, 37*(6), 443–451.

De Houwer, J. (2006). *What are implicit measures and why are we using them?* In R.W. Wiers, & A.W. Stacy (Eds.), The handbook of implicit cognition and addiction (pp. 11e28) Thousand Oaks, CA: Sage Publishers.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: a normative analysis and review. *Psychological Bulletin, 135*(3), 347–368.

De Raedt, R., Schacht, R., Franck, E., & De Houwer, J. (2006). Self-esteem and depression revisited: implicit positive self-esteem in depressed patients? *Behaviour Research and Therapy, 44,* 1017–1028.

Does van der, A. J.W. (2002). De Nederlandse versie van de Beck Depression Inventory second edition (BDI-II-NL): Handleiding.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*(8), 709–724.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: their meaning and use. *Annual Review of Psychology, 54,* 297–327.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I disorders (SCID-I)*. New York, NY: Biometrics Research Department New York State Psychiatric Institute.

Franck, E., De Raedt, R., & De Houwer, J. (2007). Implicit but not explicit self-esteem predicts future depressive symptomatology. Behaviour Research and Therapy, 45(10), 2448–2455.

Franck, E., De Raedt, R., & De Houwer, J. (2008). Activation of latent self-schemas as a cognitive vulnerability factor for depression: the potential role of implicit selfesteem. *Cognition & Emotion, 22*(8), 1588–1599.

Franck, E., De Raedt, R., Dereu, M., & van den Abbeele, D. (2007). Implicit and explicit self-esteem in currently depressed individuals with and without suicidal ideation. *Journal of Behavior Therapy and Experimental Psychiatry, 38*, 75–85.

Garety, P. A., & Freeman, D. (1999). Cognitive approaches to delusions: a critical review of theories and evidence. *British Journal of Clinical Psychology, 38*(2), 113–154.

Gemar, M. C., Segal, Z. V., Sagrati, S., & Kennedy, S. J. (2001). Mood-induced changes on the Implicit Association Test in recovered depressed patients. *Journal of Abnormal Psychology, 110*(2), 282–289.

Glashouwer, K. A., & de Jong, P. J. (2008). Impliciete persoonlijkheidstrekken en psychopathologie: Achtergrond, huidige inzichten en perspectieven. *Gedragstherapie, 41*(2), 121–134.

Glashouwer, K. A., & de Jong, P. J. (2010). Disorder-specific automatic self-associations in depression and anxiety: results of The Netherlands Study of Depression and Anxiety. *Psychological Medicine, 40*(7), 1101–1111.

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*(6), 1022–1038.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197–216.

Haaga, D. A., Dyck, M. J., & Ernst, D. (1991). Empirical status of cognitive theory of depression. *Psychological Bulletin, 110*(2), 215–236.

Haeffel, G. J., Abramson, L. Y., Brazy, P. C., Shah, J. Y., Teachman, B. A., & Nosek, B. A. (2007). Explicit and implicit cognition: a preliminary test of a dual-process theory of cognitive vulnerability to depression. *Behaviour Research and Therapy, 45*(6), 1155–1167.

Hermans, D., & De Houwer, J. (1994). Affective and subjective familiarity ratings of 740 Dutch words. *Psychologica Belgica, 34*, 115–139.

Hermans, D., Vandromme, H., Debeer, E., Raes, F., Demyttenaere, K., Brunfaut, E., et al. (2008). Overgeneral autobiographical memory predicts diagnostic status in depression. *Behaviour Research and Therapy, 46*(5), 668-677.

Hollon, S. D., Kendall, P. C., & Lumry, A. (1986). Specificity of depressotypic cognitions in clinical depression. *Journal of Abnormal Psychology, 95*(1), 52–59.

Ingram, R. E., Miranda, J., & Segal, Z. V. (1998). *Cognitive vulnerability to depression.* New York, NY: Guilford Press.
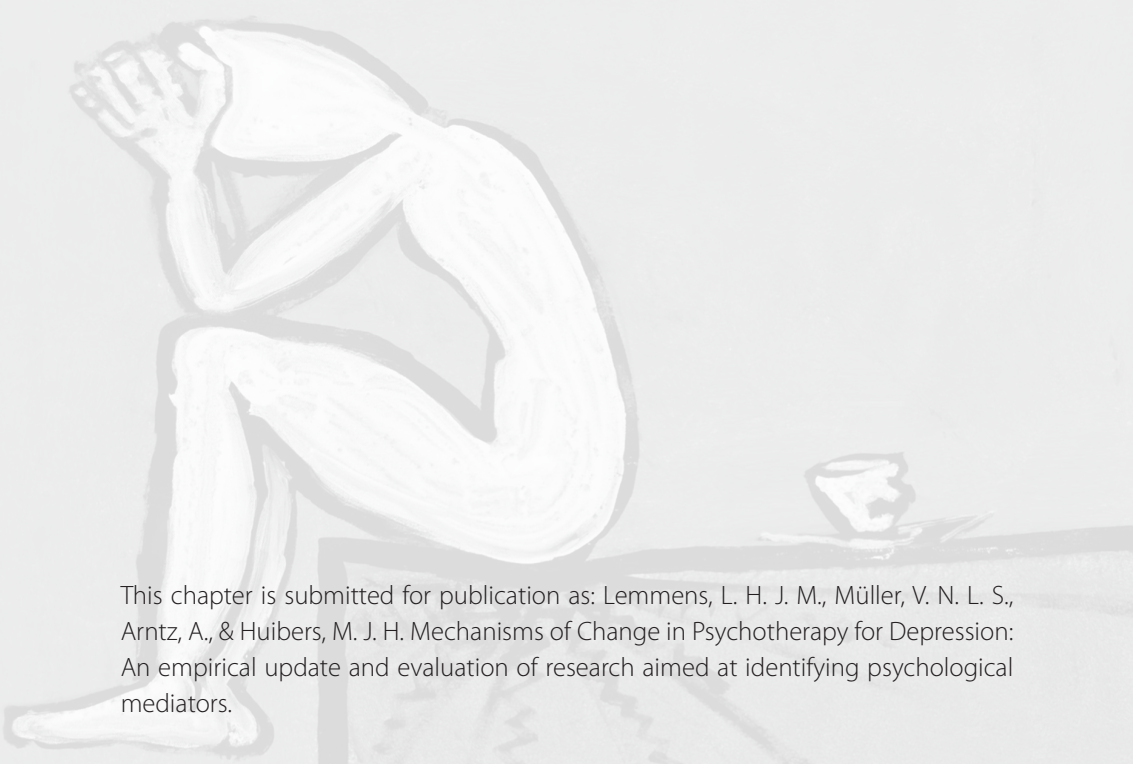
4

Karpinski, A., & Steinman, R. B. (2006). The single category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91,* 16–32.

Kesting, M.-L., Mehl, S., Rief, W., Lindenmeyer, J., & Lincoln, T. M. (2011). When paranoia fails to enhance self-esteem: explicit and implicit self-esteem and its discrepancy in patients with persecutory delusions compared to depressed and healthy controls. *Psychiatry Research, 186*(2–3), 197–202.

Lemmens, L. H. J. M., Arntz, A., Peeters, F. P. M. L., Hollon, S. D., Roefs, A., & Huibers, M. J. H. (2011). Effectiveness, relapse prevention and mechanisms of change of cognitive therapy vs. interpersonal therapy for depression: study protocol for a randomised controlled trial. *Trials, 12*(1), 150–162.

McKay, R., Langdon, R., & Coltheart, M. (2007). The defensive function of persecutory delusions: an investigation using the Implicit Association Test. *Cognitive Neuropsychiatry, 12*(1), 1–24.

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-Go Association Task. *Social Cognition, 19*(6), 625–666.

Nuttin, J. M. (1985). Narcissism beyond Gestalt and awareness: the name letter effect. *European Journal of Social Psychology, 15*(3), 353–361.

Risch, A. K., Buba, A., Birk, U., Morina, N., Steffens, M. C., & Stangier, U. (2010). Implicit self-esteem in recurrently depressed patients. *Journal of Behavior Therapy and Experimental Psychiatry, 41,* 199–206.

Roefs, A., Huijding, J., Smulders, F. T. Y., MacLeod, C. M., de Jong, P. J., Wiers, R. W., et al. (2011). Implicit measures of association in psychopathology research. *Psychological Bulletin, 137*(1), 149–193.

Schröder-Abé, M., Rudolph, A., & Schütz, A. (2007). High implicit self-esteem is not necessarily advantageous: discrepancies between explicit and implicit selfesteem and their relationship with anger expression and psychological health. *European Journal of Personality, 21*, 319–339.

Silverman, J. S., Silverman, J. A., & Eardley, D. A. (1984). Do maladaptive attitudes cause depression? *Archives of General Psychiatry, 41*, 28–30.

Steffens, M. C. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology, 51,* 165–179.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*(2), 245–251.

Steinberg, J. A., Karpinski, A., & Alloy, L. B. (2007). The exploration of implicit aspects of self-esteem in vulnerability–stress models of depression. *Self and Identity, 6*(2–3), 101–117.

Tafarodi, R. W., & Swann, W. B. (2001). Two-dimensional self-esteem: theory and measurement. *Personality and Individual Differences, 31*(5), 653–673.

Valiente, C., Cantero, D., Vãzquez, C., Sanchez, Ã., Provencio, M., & Espinosa, R. (2011). Implicit and explicit self-esteem discrepancies in paranoia and depression. *Journal of Abnormal Psychology, 120*(3), 691–699.

Vandromme, H., Hermans, D., Spruyt, A., & Eelen, P. (2007). Dutch translation of the Self-Liking/Self-Competence Scale – revised: a confirmatory factor analysis of the two-factor structure. *Personality and Individual Differences, 42*(1), 157–167.

Vázquez, C., Diez-Alegría, C., Hernández-Lloreda, M. J., & Moreno, M. N. (2008). Implicit and explicit self-schema in active deluded, remitted deluded, and depressed patients. *Journal of Behavior Therapy and Experimental Psychiatry, 39*(4), 587–599.

Xi, M.-J., Zhang, Y.-J., & Li, Y.-X. (2007). Implicit self-esteem level and stability of depressive patients. *Chinese Mental Health Journal, 21*(11), 756–758.

5

Mechanisms of Change in Psychotherapy
for Depression: An empirical update
and evaluation of research aimed
at identifying psychological mediators

Abstract

We present a systematic empirical update and critical evaluation of the current status of research aimed at identifying a variety of psychological mediators in various forms of psychotherapy for depression. We describe study characteristics and results of 32 relevant studies, and report on the extent to which they meet several important requirements for mechanism research. Our review indicates that in spite of increased attention, advances in theoretical consensus about necessities for mechanism research, and sophistication of study designs, research in this field is still heterogeneous and unsatisfactory in methodological respect. The field would benefit from a further refinement of research methods to assess the causal relation between change in the mediator and change in depressive symptoms. However, psychotherapy is a multi-dimensional phenomenon that might work through interplay of multiple mechanisms at several levels. Psychotherapeutic change might therefore be too complex to be explained in relatively simple causal models of psychological change.

5

Introduction

Many researchers in the field of clinical psychology agree that gaining a better understanding of the mechanisms underlying psychotherapeutic change is crucial for optimizing treatment outcomes for patients suffering from psychiatric disorders such as depression (Kazdin & Nock, 2003; Kraemer, Wilson, Fairburn, & Agras, 2002). Knowledge about active ingredients of therapy can assist in the verification and refinement of theories of the disorder, and allows enhancement of elements that are crucial for therapeutic change, while dismissing those found to be redundant (Garrat, Ingram, Rand, & Sawalani, 2007; Longmore & Worrell, 2007).

An important first step towards examination of mechanisms of change is the identification of mediators (Kazdin & Nock, 2003; Kraemer, Stice, Kazdin, Offord, & Kupfer, 2001; Kraemer et al., 2002). A mediator is a variable that statistically explains why and in what way a treatment has an effect on outcome, and can be seen as a potential mechanism: the actual process or event that is responsible for change (Baron & Kenny, 1986; Kazdin, 2007, 2009; Kraemer et al., 2001; MacKinnon, Fairchild, & Fritz, 2007). In other words, the mechanism is the phenomenon to reveal, the mediator can be the mean to this end.

Requirements for a Mediator

Establishing a mediator involves several requirements. For a long time, mediation solely referred to statistical mediation: to statistically demonstrate that the effect of treatment on outcome is explained by a third variable: the mediator. The most well-known and influential method to determine statistical mediation is indubitably Baron and Kenny's (1986) causal step method. With over 17.000 citations, their paper is one of the most frequently cited articles in the field of psychology (MacKinnon et al., 2007). According to Baron & Kenny, mediation is established when 1) there is a main effect of treatment (efficacy test; *path c*), 2) treatment is related to change in the mediator (intervention test; path *a*), 3) change in the mediator and change in outcome are related (psychopathology test; path *b*), and 4) the effect of treatment on outcome is absent (full mediation) or significantly weakened (partial mediation) when statistically controlling for the mediator (mediation test; path *c'*). A graphical representation of the model can be found in Figure 1. Subsequently, a Sobel test (Sobel, 1982) determines the amount of mediation – also called the indirect effect.

However, influential as it has been, the original Baron and Kenny (1986) model has significant limitations for application in social sciences and therefore also in clinical process research for disorders such as depression. For example, the method has low type I error rates and, in order to have sufficient power, requires large sample sizes and large treatment effects, both of which are not always available in this type of research (Hoyle & Kenny, 1999; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; MacKinnon, Lockwood, & Williams, 2004; Shrout & Bolger, 2002). The applicability of the model in this field is further limited by restrictions resulting from the first and fourth criterion. The first criterion (efficacy test,

5
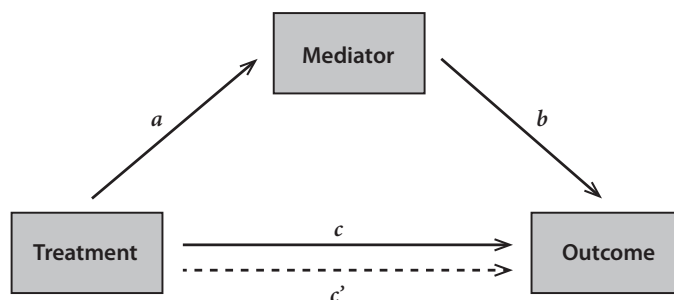
Figure 1. Statistical Mediation according to Baron & Kenny (1986).

c path) is formulated in a way that the ability to perform mediation analysis strongly depends on the presence of differential treatment effects. When two treatments turn out to be equally effective – a phenomenon that is not uncommon in the field of psychotherapy for depression (for more details see e.g. Cuijpers & van Straten, 2011; Cuijpers, van Straten, Andersson, & van Oppen, 2008; Wampold et al., 1997) – this type of mediation analysis is not possible. This is an important drawback, because especially when two treatments turn out to be equally effective it is important to examine processes of change, since this can tell us more about whether the change that is observed is reached through similar or differential pathways (MacKinnon, 2008). Moreover, given the population (depressed patients) and the nature of treatments (psychotherapy), it is ethically and practically very difficult (if not impossible) to include a substantially less powerful treatment (such as a full waiting-list control group, or a placebo intervention) to increase the contrasts between groups. And even if a third ineffective control condition would be added, it is still not possible to test differential pathways between the two equally effective treatments. The fourth Baron and Kenny (1986) criterion (mediation test) has been criticised because the tests that have to demonstrate the reduction of the effect after statistically controlling for the mediator have shown to be underpowered (MacKinnon et al., 2007).

As a result of these limitations, the criteria for statistical mediation have been modified over time to make them more applicable and suitable for treatment research. The MacArthur group (Kraemer et al., 2001, 2002) toned down the importance of the first criterion by stating that differential treatment effects are not required to establish mediation as long as there is an interaction between treatment and the mediator. This is particularly useful in clinical trials comparing two (equally) effective treatments that are likely to operate through different mechanisms. With regard to step 4, it was decided that it was sufficient to show that treatment has an effect on the mediator and that the mediator has an effect on the outcome, even after controlling for treatment, a procedure known as joint significance testing (MacKinnon et al., 2007).

Statistical Mediation is important but not sufficient

Although statistical mediation still plays a central role in addressing whether a particular construct accounts for change (Hollon & DeRubeis, 2009; Kazdin, 2007, 2009), it is not sufficient to make a case for the operation of a mediator (e.g. Johansson & Høglend, 2007; Kazdin, 2007, 2009; Laurenceau, Hayes, & Feldman, 2007). Probably the most important addition to statistical mediation is demonstrating the direction of causality. Conditions for inferring causal relations in scientific research have been outlined by e.g. Hill (1965), Kenny (1979), Schlesselman (1982), and were brought to the psychotherapy literature by Kazdin (2003, 2007, 2009). Apart from a strong statistical association between treatment, mediator and outcome, Kazdin describes six requirements for adequate evidence for causal temporal relationships. First of all, it has to be demonstrated that the treatment causes the mediator variable to change, which in turn causes the outcome, and not the other way around (Kazdin & Nock, 2003; Kraemer et al., 2002). In order to get a clear view of the shape of change and the relation between mediator and outcome, it is important that both the mediator and outcome measure are assessed at multiple time points during treatment. The importance of demonstrating *temporality* is supported by many research groups (e.g. Collins & Graham, 2002; Hollon & DeRubeis, 2009; Johansson & Høglend, 2007; Kazdin, 2007, 2009; Kazdin & Nock, 2003; Kraemer et al., 2002; Laurenceau et al., 2007; Murphy, Cooper, Hollon, & Fairburn, 2009), and has even been called the fifth step of statistical mediation analysis (Johansson et al., 2010). Second, alternative explanations for the observed relation between mediator and outcome should be ruled out. This can be done by using an *experimental approach* in which all variables are held constant across individuals in various conditions while changing only the proposed mechanism of change (Kazdin, 2007, 2009). Furthermore, Kazdin emphasizes the importance of *specificity* of the association among the intervention, proposed mediator and outcome. This means that it has to be demonstrated that the mediator plays a crucial role in one treatment, but not (or less so) in the other. In addition, inclusion of *plausible processes*, *consistency* across studies, and a *gradient*, in which larger changes in the mediator are associated with larger changes in outcome, should further enhance the evidence.

Kazdin (2007) emphasizes that each criterion is important, but that interpretations should be made based on their convergence. Examination starts with statistical tests for mediation. After that, one determines the value of the results by examining the extent to which a study meets the other criteria. Even though the satisfaction of each criterion increases the strength of the argument for the operation of a mediator – or even a mechanism – not all criteria are weighted equally important. According to Kazdin and Nock (2003), *association*, *temporality*, *specificity*, and *experiment* are considered to be the most important, whereas the remaining three should further enhance the evidence.

5

Requirements for study designs

The extended requirements for establishing mediation also called for additional features of study designs. According to the latest standards, the extent to which a process meets the requirements for mediation can only be examined properly in a theoretically well planned RCT with carefully spaced repeated measures, sufficient power and an appropriate control group (Kazdin, 2007; Kazdin & Nock, 2003; Kraemer et al., 2002; Laurenceau et al., 2007). Furthermore, it is important to experimentally manipulate the proposed mediators. In addition, mediation analyses should be performed using up-to-date definitions and state-of-the-art statistical analysis techniques (Collins & Graham, 2002; Haaga & Stiles, 2000; Haubert & Dobson, 2007; Kraemer et al., 2002; Laurenceau et al., 2007; MacKinnon et al., 2007). Moreover, depending on what the theory stipulates about processes, assessment of a single mediator might not be sufficient. It is therefore recommended to include multiple mediators to examine rival hypotheses, test alternative explanatory models, and map out interactions between theorized processes.

Research studying mediators in psychotherapy for depression

The past decades, the interest for mediators in mechanism research in depression has grown, and several research groups worldwide have studied mediators of psychotherapy. In 2007, Johansson and Høglend identified 61 studies that performed mediational analyses to identify the active ingredients of psychotherapy for several psychiatric disorders. A closer look at the literature specific for depression indicates that the majority of studies has focused on the mediational role of cognitive processes, such as automatic thoughts, dysfunctional attitudes, attributional style, and other cognitive distortions. In their influential systematic review, Garrat et al. (2007) summarize the results of 31 studies on the role of cognitive change. They conclude that research generally supports the cognitive mediation hypothesis, but that this does not necessarily needs to be specific for CT, indicating that cognitive change, no matter how it occurs, might play an important role in other psychotherapeutic interventions as well. Even though Garret et al. acknowledge that existing research increases the knowledge about the relation between cognition and depression, they emphasize that it does not permit clear-cut answers about the exact role of cognitive change as a process that facilitates symptom change in the context of psychotherapy. First of all, because of the large variety in research questions and methodology, which makes it difficult to compare results across studies and integrate findings into broader knowledge. Second, because many studies did not meet the criteria for reputable mechanism research. More specifically, Garrat et al., conclude that none of the studies that were identified in the review addressed the criteria for mediation in methodologically sound ways. They expressed their hope that this would happen in subsequent years, in studies with e.g. large sample sizes, up-to-date-statistical methods, and a broader array of measures. These difficulties are acknowledged by others in the field as well (e.g. Johansson & Høglend, 2007; Kazdin, 2007; Kraemer et al., 2001;

Laurenceau et al., 2007). A third difficulty in interpreting results from studies in this field, – not mentioned by Garrat et al. – is the fact that the term mediation is often used for multiple purposes. Apart from using it to refer to results obtained by traditional statistical tests of mediation based on Baron & Kenny (1986) – as is the case in the current paper – the term is also used to refer to potential processes of change assessed with other procedures than tests for statistical mediation (e.g. by using correlations, or prediction), or as a synonym for a mechanism. Fourth, since most research mainly focused on the role of cognitive factors, non-cognitive factors are still largely overlooked.

Aim of the current review

Eight years have passed since the Garrat et al. (2007) review and the question is whether and how the field has changed. The aim of the current review is therefore to provide an empirical update and critical evaluation of the current body of research on this topic. In a systematic literature search, we selected studies aimed at identifying psychological mediators in psychotherapy for depression. To get a comprehensive overview of the field, we did not limit ourselves to cognitive factors in CT (as Garrat et al. did), but also included non-cognitive processes and various other forms of psychotherapy for (subclinical) depression in adolescents and adults. We only selected studies that included a statistical test of mediation in the sense of Baron and Kenny (1986) or its modern extensions. We report study characteristics and results of 32 studies and assessed to which extent they met the most important requirements for mechanism research that were discussed earlier. With this we hope to learn more about the magnitude and relevance of the existing body of research and map out necessities for future research.

5

Method

Data sources and data reduction

Three different approaches were used to identify relevant studies. First, five databases (i.e. PubMed, PsychInfo, Embase, Cochrane, and Cinahl) were systematically searched for potentially relevant papers that were published in English in peer reviewed journals until summer 2014. Key terms were Depression, Psychotherapy, Mechanisms and Mediation (a full key-term scheme can be found in appendix 1). The data search yielded a total of 425 unique studies. One of us (VM) carefully read through all abstracts[1] and retained those articles that met a set of a priori generated inclusion and exclusion criteria. LL checked the generated table entries for accuracy.

To be included in the review articles needed to be empirical research reports (no reviews, theoretical essays or commentaries) examining psychological mediators over the course of treatment of various forms of evidence-based psychotherapy for patients

(adults and adolescents) with (subclinical) depression. Furthermore, studies needed to actually include statistical mediation analyses according to Baron and Kenny or its modern extensions in their analysis plan. Studies including patients diagnosed with bipolar depression were excluded, as were those that focused on other forms of psychopathology[2] or (relapse) prevention. A complete overview of the in- and exclusion criteria can be found in appendix 2.

Of the 425 articles that were identified in the literature search, 395 did not meet our inclusion criteria and were therefore excluded. Main reason for exclusion was the fact that articles did not focus on psychological mechanisms of action in depression ($n = 226$). Other papers were excluded because they were theoretical background papers (review, commentary) instead of empirical research reports ($n = 54$), or because the therapeutic intervention that was studied was not (evidence-based) psychotherapy or there was no intervention ($n = 113$). One paper was excluded because after careful reading it did not perform statistical mediation analysis (Backenstrass et al., 2006) and one because it did not include a clinical outcome measure (Johansson et al., 2010)[3]. A total of 30 articles met all inclusion criteria and were selected for further review. Subsequently, we hand searched reference lists of the 30 articles that met all inclusion criteria, and asked several experts (3 psychologists, 1 psychiatrist) with longstanding experience in the research field and clinical practice of depression to check the list that was generated. Two additional papers were added, resulting in a total of 32 studies that were further explored.

Data Assessment

Two researchers (LL and VM) carefully read the 32 articles that were selected and tabulated study characteristics and results. Furthermore, all papers were assessed by means of several important requirements for mediation research that were discussed earlier: the use of an RCT design and inclusion of a control group, a sufficient sample size (defined as $n \geq 40$), examination of multiple potential mediators within one study, the assessment of temporality (defined by 3 or more assessments in the treatment phase), and direct experimental manipulation of the mediator. Each study was rated with respect to meeting (+) or not meeting (–) each of these criteria. Differences in scoring were resolved by consensus. A qualitative analysis was conducted by summarizing, comparing and contrasting the data.

It has to be noted that *specificity* is not included in the list of features that was described above. Not because we think that examining specificity is not important (in fact, as was stated in the previous section, we think it is very important to examine whether change in two treatments is achieved through similar or differential pathways), but because we think that conceptually it does not make sense to include this as a first-order

requirement for a mediator. In our view, the primary goal in process research is to identify any factors that facilitate symptom change, regardless of their specificity to one treatment. A first priority is therefore to identify process factors that are a linking pin between treatment and outcome. A subsequent specificity analysis could then show whether this factor plays a role in only this treatment or also in other treatments. By requiring specificity as a (testable) criterion for mediation, basic information about whether or not a process facilitates symptom change is discarded when it turns out that the specificity criterion is not met. Since we consider this information important, we decided not to include specificity in our evaluation.

Results

Study characteristics and results

Table 1 (left panel) gives an overview of study characteristics and results of 32 studies that were included in the review. The majority of studies was conducted in the USA (59.4% vs. 28.1% in Europe, and 12.5% in other parts of the world), and 62.5% was published in the past five years (2010–2014). Sample sizes ranged between $n = 4$ and $n = 517$, with a mean of $n = 161$ (SD = 137.4). Patients were adults (in 24 studies) and adolescents (in 8 studies) ranging in age from 12 to 68 years (M = 40.4, SD = 8.4 for studies in adults[4] and M = 15.1, SD = 0.5 for studies including adolescents). In 90.3% of the studies the majority (> 50%) of participants were female[5].

      Cognitive (Behavioural) Therapy (C(B)T) was the most frequently researched intervention (examined in 19/32 studies), followed by Mindfulness Based Cognitive Therapy (MBCT, included in 5 studies). Other treatments were Acceptance and Commitment Therapy (ACT, 3x), Behavioural Activation (BA, 1x), Cognitive Behavioural Analysis System of Psychotherapy (CBASP, 1x), Interpersonal Psychotherapy (IPT, 2x), Non-Directive Supportive Therapy (NST, 2x), Problem Solving (Couples) Therapy (PST, 2x), Psychodynamic Therapy (1x), Psychoanalytic Therapy (1x), and Systematic Behavioural Family Therapy (SBFT, 2x). Three studies included a combined treatment.

      Common measures of depression severity were the (second edition of the) Beck Depression Inventory (BDI(-II); Beck, Steer, & Brown, 1996; Beck, Ward, Meldelson, Mock, & Erbauch, 1961), which was implemented in 18 studies, and the Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960), used in 7 studies. Seven studies (e.g. van Aalderen et al., 2012; DeRubeis et al., 1990: Warmerdam, van Straten, Jongsma, Twisk, & Cuijpers, 2010) used them both, thereby obtaining a self-report and an observer-based measure of depression.

5

Table 1. Characteristics and results of 32 identified studies aimed at identifying psychological mediators for (subclinical) depression, and the extent to which they meet requirements for process research.

| | STUDY CHARACTERISTICS AND RESULTS | | | | REQUIREMENTS FOR PROCESS RESEARCH | | | | | |
Authors	Intervention(s)	Measure(s) of Depression	Statistical Method	Potential Mediator(s) & Results	RCT	Control	n≥40	Multiple Mediators	Temporality (≥2 ass.)	Manipulation
- **v. Aalderen** **et al. (2012)** (NL)	- MBCT + TAU (n = 102) vs. - TAU only (n = 103)	- HRSD - BDI	- Preacher & Hayes (2008) - MacKinnon (2008)	- ↓ Rumination [(+)] - ↓ Worry [(+)] - Mindfulness skills [(-)]	(+)	(+)	(+)	(+)	(-)	(-)
- **Allart-v.** **Dam et al.** **(2003)** (NL)	- Group CBT (n = 68) vs. - Advice only (n = 42)	- BDI	- Baron & Kenny (1986) - MacKinnon et al. (2002)	- ↓ Negative Thoughts [(+)] - ↑ Self-esteem [(+)] - Pleasant activities [(-)] - Social skills [(-)] - Perceived support [(-)]	(+)	(+)	(+)	(+)	(-)	(-)
- **Batink** **et al. (2013)** (NL)	- MBCT (n = 64) vs. - TAU: WLC (n = 66)	- IDS-SR - HRSD	- Goodman (1960) - Sobel (1982) - MacKinnon et al. (2007)	- Δ in Mindfulness skills [(+)] - Δ Positive affect [(+)] - Δ Negative affect [(+)] - Δ Worry [(+ via pos. and neg. affect)] - Δ Rumination on sadness [(-)]	(+)	(+)	(+)	(+)	(-)	(-)
- **Beevers** **et al. (2007)** (USA)	- PHT + various forms of Psychological Treatment (n = 121)	- MHRSD	- Curran & Hussong (2002)	- Negative Cognition [(+ in relation between depression history and outcome)] - Family impairment [(predicts slower change in depression)]	(-)[1]	(-)[1]	(+)	(+)	(-)	(-)

Table 1. (continued)

| | STUDY CHARACTERISTICS AND RESULTS | | | | REQUIREMENTS FOR PROCESS RESEARCH | | | | | |
| | | Measure(s) of | | | | | | Multiple | Temporality | |
Authors	Intervention(s)	Depression	Statistical Method	Potential Mediator(s) & Results	RCT	Control	n≥40	Mediators	(≥2 ass.)	Manipulation
- Blalock et al. (2008) (USA)	- CBASP+PHT (n=179) vs. - PHT only (n=165) vs. - CBASP only (n=173)	- HRSD	- Baron & Kenny (1986) - MacKinnon (1994)	- Attributional style (+) - Coping style –Escape/Avoidance (+) –Problem solving (+) –Distancing (-) –Accepting –Responsibility (-)	(+)	(+)	(+)	(+)	(-)	(-)
- Cohen et al. (2014) (USA)	- Brief Problem Focused Therapy (n=18) vs. - WLC (n=17)	- BDI-II	- Baron & Kenny (1986) - Pitariu & Ployhart (2010) - MacKinnon (2008)	- Negative behaviour/attitude towards depression (+) - Empathic communication (-) - Support (-)	(+)	(+)	(-)	(+)	(-)	(-)
- DeRubeis et al. (1990) (USA)	- CT + PHT (n=32) vs. - PHT only (n=32)	- Composite score of 3 measures of depression[2]	- Baron & Kenny (1986) - Temporal correlations**	- Δ Attributional style (-) - Δ Neg. auto. thoughts (-) - Δ Dysfunctional attitudes (-) - Δ Hopelessness (-)	(+)	(+)	(-)	(+)	(+)	(-)

5

Table 1. (continued)

	STUDY CHARACTERISTICS AND RESULTS				REQUIREMENTS FOR PROCESS RESEARCH					
Authors	Intervention(s)	Measure(s) of Depression	Statistical Method	Potential Mediator(s) & Results	RCT	Control	$n \geq 40$	Multiple Mediators	Temporality ($\geq$ 2 ass.)	Manipulation
- **Dietz et al. (2014)** (USA)	- CBT ($n = 25$) vs. - SBFT ($n = 20$) vs. - NST ($n = 18$)	- BDI - K-SADS-P	- MacKinnon et al. (2002)	- Δ in Interpersonal Behaviour [+ but only in CT and only when maternal depressive symptoms were low or moderate] – Δ in problem solving – Δ in involvement [-] – Δ in dyadic conflict [-]	(+)	(+)	(-)	(+)	(-)	(-)
- **Fledderus et al. (2013)*** (NL)	- Self-help ACT ($n = 250$) vs. - WLC ($n = 126$)	- CES-D	- Preacher & Hayes (2008)	- ↑ psychological flexibility [+ in ACT]	(+)	(+)	(+)	(-)	(+)	(-)
- **Forman et al. (2012)*** (USA)	- CT ($n = 90$) vs. - ACT ($n = 84$)	- BSQ	- Krull & MacKinnon (2001); - Mackinnon et al. (2002, 2004, 2007)	- Cognitive and Affective Change [+ CT only] - Acceptance [+ ACT only] - Cognitive defusion [+ both] - Dysfunctional Thinking [+both] - Willingness to engage in behavioural activity [+both]	(+)	(+)	(+)	(+)	(+)	(-)
- **Gaynor & Harris (2008)** (USA)	- BA ($n = 4$)	- BDI-II	- Mueser, Yarnold, & Foy, (1991)	- ↑Activation level [+ in 2/4 patients] - ↓Dysfunctional thoughts [-]	(-)	(-)	(-)	(+)	(+)	(-)

Table 1. (continued)

| | STUDY CHARACTERISTICS AND RESULTS | | | | REQUIREMENTS FOR PROCESS RESEARCH | | | | | |
| | | Measure(s) of | | Potential Mediator(s) | | | | Multiple | Temporality | |
Authors	Intervention(s)	Depression	Statistical Method	& Results	RCT	Control	$n \geq 40$	Mediators	(≥ 2 ass.)	Manipulation
- Jacobs et al. (2009) (USA)	- CBT ($n=111$) vs. - PHT ($n=109$) vs. - CBT + PHT ($n=107$) vs. - Pill placebo ($n=112$)	- CDRS-R	- Kraemer et al. (2002)	- Perfectionism (DAS) [+ partial]	(+)	(+)	(+)	(−)	(+)	(−)
- Jacobs et al. (2014) (USA)	- CBT ($n=95$) vs. - PHT ($n=99$) vs. - CBT + PHT ($n=97$)	- CDRS-R	- Kraemer et al. (2002)	Δ in 4 Cognitive Constructs –in Cognitive distortions [+] –in Cognitive avoidance [+] –in Positive outlook [+strongest] –in Solution-foc. thinking [-]	(+)	(+)	(+)	(+)	(−)	(−)
- Kaufmann et al. (2005) (USA)	- Group CBT ($n=45$) vs. - Life skills control ($n=48$)	- BDI-II - HRSD	- Baron & Kenny (1986); - MacKinnon et al. (2002); - Muthén & Muthén (2001)	- Δ Automatic Thoughts [+] - Relaxation [-] - Social Skills [+] - Pleasant Activities [-] - Problem Solving [-] - Working Alliance [-] - Group Cohesion [-]	(+)	(+)	(+)	(+)	(−)	(−)

5

Table 1. (continued)

	STUDY CHARACTERISTICS AND RESULTS				REQUIREMENTS FOR PROCESS RESEARCH					
Authors	Intervention(s)	Measure(s) of Depression	Statistical Method	Potential Mediator(s) & Results	RCT	Control	$n \geq 40$	Multiple Mediators	Temporality (≥ 2 ass.)	Manipulation
- **Klug et al. (2012)** (GER)	- Psychoanalytic (n=35) vs. - Psychodynamic (n=31) vs. - CBT (n=34)	- BDI - SCL-90-R	- Baron & Kenny (1986)	- Positive Introject [+] - Therapeutic relationship [-]	(−)	(+)	(−)	(+)	(−)	(−)
- **Kolko et al. (2000)** (USA)	- CBT (n=36) vs. - SBFT (n=32) vs. - NST (n=35)	- BDI	- Baron & Kenny (1986); - Holmbeck (1997)	- Cognitive distortion [-] - Family dysfunction [-]	(+)	(+)	(−)	(+)	(+)	(−)
- **Kuyken et al. (2010)** (USA)	- Group MBCT (n=61) vs. - Maint. PHT (n=62)	- HRSD	- Kraemer et al. (2002)	- ↑ Mindfulness [+ in MBCT] - ↑ Self-compassion [+ in MBCT] - Cognitive reactivity [+ in MBCT]	(+)	(+)	(+)	(+)	(−)	(−)
- **Kwon & Oei (2003)** (AUS)	- Group CBT (n=35)	- BDI	- SEM - Arbuckle & Wothke (1999)	- ↓Neg. automatic thoughts [+]()	(−)	(−)	(−)	(−)	(+)	(−)

Table 1. (continued)

	STUDY CHARACTERISTICS AND RESULTS				REQUIREMENTS FOR PROCESS RESEARCH					
Authors	Intervention(s)	Measure(s) of Depression	Statistical Method	Potential Mediator(s) & Results	RCT	Control	n ≥ 40	Multiple Mediators	Temporality (≥ 2 ass.)	Manipulation
- **Lewis et al. (2009)** (USA)	- CBT (n = 79) vs. - PHT (n = 85) vs. - CBT + PHT (n = 86) vs. - Pill placebo (n = 82)	- CDRS-R	- Kraemer et al. (2002)	- Readiness to change –Action (+) –Pre-contemplation (–) –Contemplation (–) –Maintenance (–)	(+)	(+)	(+)	(+)	(–)	(–)
- **Lo et al. (2013)*** (CHI)	- C-MBCT (n = 41) vs. - WLC (n = 41)	- BDI-II	- Baron & Kenny (1986) - Sobel (1982)	- Stagnation (+)	(+)	(+)	(+)	(–)	(–)	(–)
- **Quilty et al. (2008)** (CAN)	- CBT (n = 45) vs. - IPT (n = 46) vs. - PHT (n = 41)	- BDI-II - HRSD	- Arbuckle (2005) - Shrout & Bolger (2002) - Kline (2005)	- Dysfunctional Attitudes (+ only in CT)	(+)	(+)	(+)	(–)	(–)	(–)
- **Ryba et al. (2014)** (USA)	- BA (n = 23)	- BDI-II	- Selig & Preacher (2009) - Muthén & Muthén (2007)	- ↑ Environmental Reward (+ both in relation between 1) proportion- and 2) quantity of activities completed and outcome depression)	(–)[1]	(–)[1]	(–)	(–)	(+)	(–)

5

Table 1. (continued)

| | STUDY CHARACTERISTICS AND RESULTS | | | | REQUIREMENTS FOR PROCESS RESEARCH | | | | | |
Authors	Intervention(s)	Measure(s) of Depression	Statistical Method	Potential Mediator(s) & Results	RCT	Control	$n \geq 40$	Multiple Mediators	Temporality (≥ 2 ass.)	Manipulation
- **Shahar et al. (2004)** (USA)	- Combined sample: CBT, IPT & PHT (n = 144)	- Composite score of 5 clinical measures[4]	- SEM - Hoyle & Smith (1994) - Arbuckle (1999)	- Social Network (+ in relation between perfectionism and depression) - Working Alliance (+ in relation between perfectionism and depression)	(–)[1]	(–)[1]	(+)	(+)	(–)	(–)
- **Shahar et al. (2010)** (USA)	- Group MBCT (n = 26) vs. - WLC (n = 19)	- BDI	- Preacher & Hayes (2004, 2008)	- Mindfulness (+) - Rumination: Brooding (+) - Rumination: Reflection (-)	(+)	(+)	(–)	(+)	(–)	(–)
- **Stice et al. (2010)** (USA)	- Group CBT (n = 89) vs. - Group SE-CBT (n = 88) vs. - Bibliotherapy (n = 80) vs. - Control (n = 84)	- BDI	- Stice, Presnell, Gau, & Shaw (2007)	- Negative cognitions (-) - Pleasant activities (-) - Emotional expression (-) - Loneliness (-)	(+)	(+)	(+)	(+)	(+)	(–)
- **Toth et al. (2013)** (USA)	- IPT (n = 99) vs. - ECS (n = 29)	- BDI-II - HRSD	- Cheong, MacKinnon & Khoo (2001, 2003) - Jo (2008)	- ↓ Perceived Stress (+ on both BDI-II and HRDS) - ↑ Family Social Support (+ on HRDS) - ↑ Friends Social Support (-) - ↑ social adjustment (-)	(+)	(+)	(–)	(+)	(–)	(–)

Table 1. (continued)

	STUDY CHARACTERISTICS AND RESULTS				REQUIREMENTS FOR PROCESS RESEARCH					
Authors	Intervention(s)	Measure(s) of Depression	Statistical Method	Potential Mediator(s) & Results	RCT	Control	n≥40	Multiple Mediators	Temporality (≥2 ass.)	Manipulation
- Warmerdam et al. (2010) (NL)	- Online CBT (n=88) vs. - Online PST (n=88) vs. - WLC (n=87)	- CES-D	- Baron & Kenny (1986)	- Dysfunctional attitudes (+ CBT and PST); - Worrying (+ CBT and PST); - Negative problem orientation (+ CBT and PST); - Other problem solving orientations (−); - Perceived control (+ CBT and PST)	(+)	(+)	(+)	(+)	(+)	(−)
- Watkins et al. (2011) (UK)	- R-CBT + TAU (n=21) vs. - TAU only (n=21)	- HRSD - BDI-II	- Baron & Kenny (1986)	- Δ Rumination (+)	(+)	(+)	(−)	(−)	(−)	(−)
- Watson et al. (2014) (USA)	- Combined sample: CBT & EFT-PE (n=55)	- BDI	- Baron & Kenny (1986)	- ↓ insecure attachment (+ in the relation between therapists' empathy and outcome); - ↓ negative self-treatment (+ in the relation between therapists' empathy and outcome); - ↑ secure attachment (+ in the relation between therapists empathy' and outcome); - ↑ positive self-treatment (−)	(−)[1]	(−)[1]	(+)	(+)	(−)	(−)
- Webb et al. (2013)* (USA)	- CBT (n=420)	- CES-D - BASIS	- Preacher & Hayes (2008)	- Cognitive skills (+ in expectancy outcome relation); - Behavioural skills (+ in expectancy outcome relation)	(−)	(−)	(+)	(+)	(−)	(−)

5

Table 1. (continued)

	STUDY CHARACTERISTICS AND RESULTS				REQUIREMENTS FOR PROCESS RESEARCH					
Authors	Intervention(s)	Measure(s) of Depression	Statistical Method	Potential Mediator(s) & Results	RCT	Control	n≥40	Multiple Mediators	Temporality (≥2 ass.)	Manipulation
- v/d Zanden et al. (2014) (NL)	- Online CBT (n=121) vs. - WLC (n=123)	- CES-D	- Baron & Kenny (1986) - MacKinnon (2008)	- Δ in Anxiety [+] - Δ in Mastery [+]	(+)	(+)	(+)	(+)	(−)	(−)
- Zettle et al. (2011) (USA)	- CT (n=13) vs. - ACT (n=12)	- BDI	- Baron & Kenny (1986) - Preacher & Hayes (2004, 2008)	- Cognitive defusion [+ in ACT] - Depressogenic thoughts [−] - Dysfunctional attitudes [−]	(+)	(+)	(−)	(+)	(−)	(−)

ABBREVIATIONS: Column Headings: RCT= Design is Randomized-Controlled Trial (yes/no); n≥40=Sample size per treatment arm is at least 40 (yes/no); Control=Control Group (yes/no); Multiple Mediators=Study included more than 1 potential mediator (yes/no); Temporality (ass n > 2)=Number of assessment in treatment phase (FU assessments are not included in count); Manipulation of mediator=Manipulation of Mediator (yes/no); (+)=Present/Yes; (−)=Absent/No. Countries: AUS = Australia; CAN=Canada; CHI=China; GER= Germany; NL =the Netherlands; UK = United Kingdom; USA =United States of America Interventions: ACT =Acceptance and Commitment Therapy; BA = Behavioural Activation; CBASP = Cognitive Behavioural Analysis System of Psychotherapy; CBSM= Cognitive Behavioural Stress Management Treatment; CBT = Cognitive Behavioural Therapy; CM = clinical management; C-MBCT = Compassion focused MBCT; CT = Cognitive Therapy; DT = Dynamic Therapy; ECS = Enhanced Community Standard; EFT-PE = Emotion-Focused Therapy, Process Experiential approach; IPT: Interpersonal Psychotherapy; MBCT= Mindfulness Based Cognitive Therapy; MBSR = Mindfulness Based Stress Reduction; NST =Non-Directive Supportive Therapy; PHT = Pharmacotherapy; PST = Problem Solving Therapy; R-CBT = Rumination focused CBT; SBFT = Systematic Behavioural Family Therapy; SE-CBT = Supportive-Expressive Cognitive Behavioural Therapy; TAU= Treatment-as-usual; WLC = Waiting-List Control. Outcome measures: BASIS = Behaviour and Symptom Identification Scale; BDI = Beck Depression Inventory; BDI-II = Beck Depression Inventory II; BSI = Brief Symptom Inventory; BSQ = Before Session Questionnaire (items: symptom intensity and progress towards goal); CDRS-R = Children's Depression Rating Scale Revised; CES-D (10) =Centre for Epidemiological Studies Depression Inventory (10 item version).; HRSD = Hamilton Rating Scale for Depression; IDS-SR = Inventory of Depressive Symptoms, self-rating; K-SADS-P = Schedule for Affective Disorders and Schizophrenia for School-Age children, Present Episode Version; MHRSD = Modified Hamilton Rating Scale for Depression; PFS-IF = Psychodynamic Functioning Scale – Interpersonal Functioning Subscales; SCL-90-(R) = Symptoms Checklist 90 (Revised). Statistical Method: SEM=structural equation modeling. Mediator variables: DAS= Dysfunctional Attitudes Scale. Notes: *= study sample consisted of patients diagnosed with depression (> 50%) and anxiety disorders; **= Used residualized change scores; [1] =Data come from RCT's, but studies do not make use of RCT design in analyses: Beevers et al. (2007), Shahar et al, 2004 and Watson et al. (2014) use a combined sample, and Ryba et al. (2014) only select patients allocated to BA. Therefore there is no control/comparison group available in these studies; [2] = Composite score of BDI, HRSD, Raskin Depression Scale; [3] = Description of the exact statistical procedure used in this study is lacking. Based on reading the method section one could conclude that this is Baron & Kenny (1986) and Sobel test (1982); [4] = Composite score of BDI, SCL-90, HRDS, Global Assessment Scale (GAS) and Social Adjustment Scale (SAS), as described by Blatt et al. (1996).

The identified studies examined 40 different potential mechanisms. Given the substantial number of studies that examined C(B)T, mediators were predominantly the theorized processes of this intervention, such as Negative (Automatic) Thoughts (7 studies), Dysfunctional Attitudes (6 studies), Attributional style (2 studies) and other cognitive constructs (7 constructs in 6 studies). Furthermore, six studies assessed the behavioural component of CBT. In studies in which Mindfulness-Based interventions were the choice of treatment Rumination, Mindfulness, and Worry were common process measures (included in 4, 4, and 3 studies respectively). The potential mediational role of Therapeutic Alliance was examined in 3 of 32 studies. As can be seen in Table 2 (left panel), conclusions about the mediational role of the constructs described above are mixed. Approximately half of the studies examining the same construct find evidence for mediation, whereas the other half does not find a relation between the mediator and outcome.

Table 2. Selection of significant mediators in the identified studies.

	All studies ($n = 32$)		Studies meeting ≥ 4 criteria ($n = 16$)	
	Examined	Significant	Examined	Significant
- Therapeutic Alliance	3	1×	1	1×
- Attributional Style	2	1×	2	1×
- Behavioural Concepts	6	3×	4	3×
- Dysfunctional Attitudes	6	3×	3	2×
- Mindfulness skills	4	3×	3	2×
- Negative (Automatic) Thoughts	7	4×	4	2×
- Rumination	4	3×	2	1×
- Worry	3	3×	3	3×
Total	35	21× (60%)	22	15× (68%)

Statistical methods were predominantly (an extension of) the Baron and Kenny (1986) causal step method. Early papers mainly used linear regressions to examine the four basic steps of the mediational model, and a Sobel (1982) test to examine the size of the indirect effect. However, as time passed, a range of new (more sophisticated) statistical analyses techniques was observed. For example, the four steps of mediation were now also estimated by multiple regression (ordinary least squares), logistic regression, multilevel regression and structural equation modeling, and the size of the indirect effect using e.g. joint-significance (MacKinnon et al., 2007) and bootstrapping (Preacher & Hayes, 2004, 2008).

A closer look at the statistical method of the identified studies showed that two studies (DeRubeis et al., 1990; Kolko, Brent, Baugher, Bridge, & Birmaher, 2000) could not finish their mediation analyses because the treatment conditions that were compared did

not differ significantly with regard to outcome. However, while DeRubeis et al. (1990) concluded that mediation analysis was not *possible* because group differences were absent, Kolko et al. (2000) concluded that mediational effects of the proposed mediators were *lacking*. In our view it is inappropriate to conclude that effects are absent before having a chance to properly study these effects. Re-analysing these data using the adapted guidelines of MacArthur (as discussed in the introduction) could have been a solution here.

A closer look at the value of these results

As discussed by Kazdin (2007), after completing statistical mediation analysis, one should return to the other criteria to assess the extent to which they are met. The results of the assessment of requirements for process research are presented for each individual study in the right panel of Table 1 and summarized in Table 3.

Table 3. Number (%) of studies meeting requirements for process research ($n = 32$).

Requirement	*n* studies (%)	
- RCT, yes, *n* (%)	24	(75.0)
- Control group, yes, *n* (%)	25	(78.1)
- Sample Size per condition ≥ 40, yes, *n* (%)	20	(62.5)
- Multiple mediators, yes, *n* (%)	25	(78.1)
- Assessment of temporality, yes, *n* (%)	10	(31.3)
- Manipulation of mediator/experiment, yes, *n* (%)	0	(0.00)

The majority of studies (75%) used an RCT design, and consequently included one or more comparison groups. Interventions of interest were compared to a) other active treatments (psychological and/or pharmacological; e.g. Blalock et al., 2008; DeRubeis et al., 1990; Forman et al., 2012; Jacobs et al., 2009, 2014); b) treatment as usual (e.g. Watkins et al., 2011); or c) non-active waiting-list control conditions (e.g. Fledderus, Bohlmeijer, Fox, Schreurs, & Spinhoven, 2013; Lo, Ng, Chan, Lam, & Lau, 2013; Shahar, Britton, Sbarra, Figueredo, & Bootzin, 2010). Studies with a non-active control group showed relatively more significant mediators as compared to studies with an active contrast group. More specifically, in studies with an active control group 55.6% of the investigated mediators were found to be statistically significant, whereas for studies with a non-active control this was 67.9%. Furthermore, four studies used data that originally came from RCTs, but did not make use of the RCT design in the mediational analyses. They either only selected patients allocated to one particular condition (Ryba, Lejuez, & Hopko, 2014), or merged the various intervention groups into one combined sample (Beevers, Wells, & Miller, 2007;

Shahar, Blatt, Zuroff, Krupnick, & Sotsky, 2004; Watson, Steckley, & McMullen, 2014). As a result, there was no control/comparison group available. In addition, as can be seen in Table 3, the number of studies including a control group is higher than the number of studies with an RCT design. This can be explained by the fact that one study compared two treatments in a non-randomized design (e.g. Klug, Henrich, Filipiak, & Huber, 2012).

Table 3 furthermore shows that two thirds of the selected studies included sample sizes of more than 40 participants per condition. This was even the case in several RCTs with three or four arms (e.g. Jacobs et al., 2009, 2014; Stice, Rhode, Seeley, & Gau, 2010). However, the small sample sizes in several other studies show that power can still be an issue in this type of research, also in relatively recent studies (e.g. Ryba et al., 2014; Watkins et al., 2011; Zettle, Rains, & Hayes, 2011). The smallest sample was found in a study by Gaynor and Harris (2008), who conducted single participant assessment of mediators in four depressed adolescents. One study explicitly compensated for the small sample size and low power by conducting mediation analyses with the therapy groups combined (Watson et al., 2014).

Almost 80% of studies included more than one mediator in their design. Some studies included several separate potential processes of change (e.g. Allart-van Dam, Hosman, & Hoogduin, 2003; Kaufman, Rohde, Seeley, Clarke, & Stice, 2005; Warmerdam et al., 2010), whereas others examined subscales of the same construct (Blalock et al., 2008; Lewis et al., 2009). However, even when multiple mediators were included in a study, they were often analysed individually. Only a small number of studies looked at the relative importance and collaboration between several potential mechanisms. For example, the study by Batink, Peeters, Geschwind, Os van, and Wichers (2013) indicated that even though both positive affect, as well as negative affect played a substantial mediating role in the reduction of depressive symptoms during MBCT, the effect of the first was larger compared to the latter. Shahar et al. (2010) also included several potential mediators in one model and showed that changes in mindfulness and changes in brooding both mediated the effect of MBCT on depression severity, and that they did so to the same extent.

A closer look at the aspect of temporality identified three categories of studies. First of all, there were 10 studies that assessed mediator(s) and outcome more than twice during treatment, and were therefore able to make some kind of judgment about the temporal order of change (e.g. DeRubeis et al., 1990; Fledderus et al., 2013; Jacobs et al., 2009; Kwon & Oei, 2003; Warmerdam et al., 2010). Two of these ten studies even assessed mediators and outcome on a session-by-session basis (Forman et al., 2012; Ryba et al., 2014). The second group consisted of studies that only included pre- and post-treatment assessments. By assessing processes and outcomes only at pre- and post-treatment one can say that change in a mediator indeed correlates with, explains a certain amount of variance, or predicts change in outcome, but not whether one process precedes the other. For example, Quilty, McBride, and Bagby (2008) found in their study that a decrease in dysfunctional attitudes was associated with a decrease in depression severity in CBT. This is in line with cognitive theory of depression. However, no conclusions about temporality

5

could be drawn because mediators and outcome measures were only measured twice at the same assessments. Similarly, the fact that Allart-van Dam et al. (2003) found that changes in depressive cognition and self-esteem were significant mediators of depressive symptoms following a coping with depression course, is of less value because they only used two assessment points. Other examples can be found in Table 1. A third category consisted of studies that did include more than two assessment points, but not within the active phase of treatment. For example, Kuyken et al. (2010), included a total of three assessment points, but one of them was at 15 month follow-up, leaving only 2 assessments during treatment (baseline and post-treatment). A similar approach was used by Toth et al. (2013). Even though this is very informative regarding to the knowledge on mediators of sustained treatment effects, it will not help to reveal mechanisms during treatment.

Lastly, none of the identified studies used an approach in which the proposed mediator was experimentally manipulated.

The criteria in concert

Since satisfaction of each criterion increases the strength of the argument for the operation of a mediator, further interpretation of findings should be based on concerted action between these criteria. We therefore also looked at the total number of criteria met by each study. An overview is given in Figure 2. As can be seen in the figure, not one study meets every criterion.
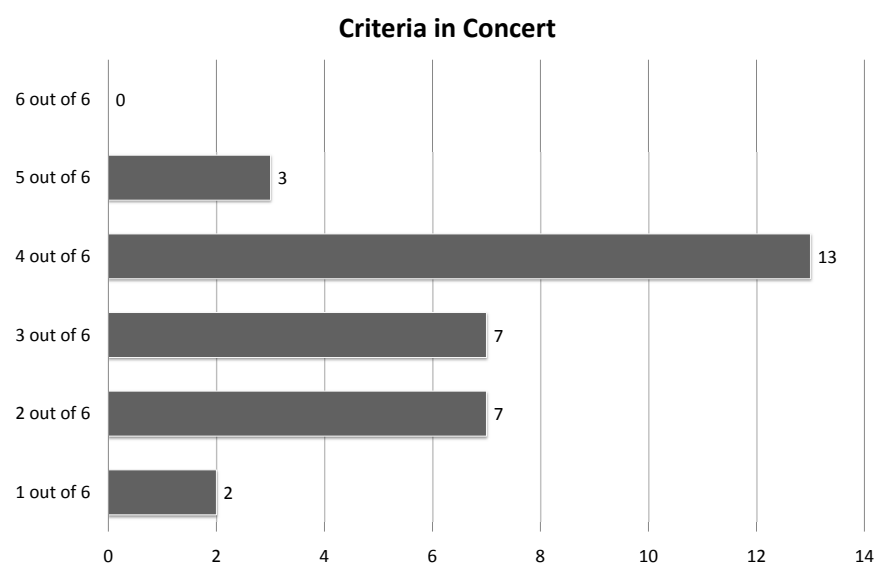
Figure 2. The criteria in Concert: Number of studies per number of criteria met.

Three studies met five out of six criteria and seem to be the most promising with regard to meeting the various criteria. Forman et al. (2012) examined the mediating role of theorized mechanisms in ACT and CT (utilization of cognitive acceptance vs. change, utilization of affective acceptance vs. change, dysfunctional thinking, cognitive defusion and committed action). They found that treatment group moderated the mediating effects of both cognitive and affective changes. More specifically, cognitive techniques (challenging and restructuring dysfunctional cognitions, distraction from unhelpful thoughts and feelings) facilitated outcome for those receiving CT, whereas utilization of psychological acceptance strategies facilitated outcome in ACT. Results of this study are promising since they are obtained in a large RCT ($n = 174$) with repeated assessments (before each session) of multiple mediators and outcomes. It has to be noted however, that they included a mixed sample of patients with anxiety and depression, and did not control for the influence diagnosis. This has to be kept in mind when interpreting the results of this study.

Stice et al. (2010) randomized 341 teens with elevated levels of depression to either group CBT, Group Supportive Expressive therapy (SET), Cognitive Behavioural Bibliotherapy or assessment-only control, and examined the mediating role of theorized processes of change of CBT and SET. For the CBT interventions processes of interest were negative cognitions and pleasant activities, and for SET therapy these were emotional expression and loneliness. Separate analyses were conducted for each of the active treatments, in which each treatment was contrasted to the non-active control. The results of the Group CBT intervention indicated the presence of a mediator: the treatment reduced depressive symptoms, negative cognitions, and increased pleasant activities. Furthermore, change in these processes predicted change in depression, and intervention effects became weaker when controlling for change in the processes. However, after examining the sequence of changes, it was found that change in depression occurred before change in the mediator. Therefore it was concluded that changes in theorized processes did not mediate the intervention effects. This illustrates the importance of including the aspect of temporality. A similar pattern was found for SET, although correlations between change in the potential mediators and outcome were not as strong as in the CBT intervention. Bibliotherapy did not significantly affect the potential mediators or depressive symptoms. Again, change in depression occurred before change in the mediators.

Warmerdam et al. (2010) studied the mediating role of dysfunctional attitudes, worry, negative problem orientation, and feelings of control in online CBT and PST for depression. A total of 263 participants were randomly allocated to one of the two active treatment conditions, or to a waiting list condition. Measures were taken at three points over the course of treatment. Mediation effects were assessed following Baron & Kenny's (1986) requirements for mediation. Similarly to Stice et al. (2010) active conditions were contrasted to the WLC condition. Warmerdam and colleagues found support for the notion that the mechanisms of interest played a mediating role in both CBT as well as in PST. Multiple mediation analysis showed that – in both groups – reduction in depression

5

was mostly explained by improvement in worrying, perceived control and a negative problem orientation. The relationship between early change in potential mediators and subsequent change in depression was examined using correlation analysis and regression. However, since most of the total improvement had already taken place before the mid-treatment assessment (5 months) – leaving only little room for later change – the authors were not able to differentiate between cause and effect. So in spite of a suitable repeated measures design, and promising results, they were not able to discern the temporal relation necessary to identify a mechanism of change.

In addition, 13 studies met 4 out of 6 criteria. As can be seen in Table 1, the combination of criteria that were met was different for the various studies. Apart from the manipulation criterion, many studies did not meet the requirement of temporality. The remaining 16 studies met less than 4 criteria. Remarkably, the two studies meeting only one criterion (Kwon & Oei, 2003; Ryba et al., 2014), met the temporality criterion, which was lacking in many of the other studies.

The question that remains is what is left of the empirical evidence when only taking high quality studies into consideration (i.e. studies that meet ≥ 4 criteria). Results of studies meeting 4 or more criteria ($n = 16$) are displayed in the right panel of Table 2. As can be seen in the table, the relative percentage of studies finding significant mediators is slightly higher than when all studies are taken into account (68% vs 60%). However, results should be interpreted with caution given the relatively small number of studies per potential mechanism.

Discussion

Eight years after the influential review by Garrat et al. (2007), we provided a systematic empirical update and critical evaluation of the current status of research aimed at identifying a variety of psychological mediators in various forms of psychotherapy for depression, to see whether – and how – the field has changed. We described study characteristics and results of 32 relevant empirical studies that were identified in a systematic literature search, and reported on the extent to which these studies met several important requirements for mechanism research. The selected studies examined a total of 40 potential mediators in 12 different treatment modalities. Conclusions about the mediational role of the various constructs that were examined across studies are mixed. Approximately half of the studies examining a similar construct found support for statistical mediation. Studies with a non-active control group showed relatively more significant mediators as compared to studies with an active control group. None of the identified studies met the highest standards for tests of treatment mediation, mainly because studies were unable to assess the temporal relationship between change in mediator and change in outcome, and because none of the studies used an approach in which the proposed mediator was experimentally manipulated.

When comparing our findings to those of Garrat et al. (2007) and their prede-cessors, it can be concluded that some advances have been made in theoretical con-sensus about necessities for this type of research, and in the degree of sophistication that researchers bring to research on mediators. More and more attention is paid to the aspect of temporality, sample size, and the inclusion of multiple processes in one study. Nevertheless, the empirical state of affairs has shown little progress in the past decade. Research is still heterogeneous and often unsatisfactory in methodological regard. As a result, after more than three decades of process research focused on depression treat-ment there is still no clear-cut empirical explanation for psychotherapeutic change.

Probably the biggest challenge in research aimed at identifying mediators is demonstrating the causal relation between change in mediator and change in de-pression severity. Demonstrating causality is difficult though, even in studies that are designed to explain therapeutic change in terms of causal processes. First of all, because determining the best timing and spacing of observations to capture the critical point of change is a difficult and delicate matter, especially when there is no prior information available about the speed and shape of change. One needs to balance the most optimal study design, with the burden for patients, and the risk of measurement artefacts when making too many demands for data (Longwell & Truax, 2005). Furthermore, research de-signs are often based on the assumption that change is gradual and linear. However, various studies have shown that change often happens sudden, rather than gradually over the course of treatment (see review of Aderka, Nickerson, Bøe, & Hofmann, 2011 for more details). If therapeutic change indeed occurs suddenly (e.g. the 'aha-experience') it might be very difficult to capture this moment, let alone to assess the temporal rela-tion between change in the mechanism and change in symptoms. From this point of view the fine graininess of analyses might never suffice to identify the critical moment in which change takes place.

Designing and conducting experimental studies – a second important require-ment for demonstrating causality – might even be a bigger challenge in the field of psychotherapy for depression research. Apart from ethical issues, psychological con-structs are difficult to manipulate. Furthermore, as Bullock, Green, and Ha (2010) point out, there are several important methodological aspects that need to be taken into ac-count in manipulation studies. Perhaps the most difficult aspect to demonstrate is that the manipulation that is used only affects one mediator without affecting the others. This is especially complicated when research is conducted in clinical settings which are usually less controlled than studies that are conducted in the laboratory. Bullock and colleagues underline that these complications cannot be overcome by purely statistical innovations, but that awareness of important assumptions of manipulation studies and their implications for research designs, is crucial for progress in the field.

Research aimed at identifying the active ingredients of psychotherapy for de-pression, would benefit from a further refinement of research methods to disentangle mechanisms of change. More specifically, future studies focus on establishing a more

5

fine grained analysis of the exact shape of change. Apart from traditional designs to examine processes of change, alternative designs including experimental manipulations and component analyses should be considered as well, since this is the only way to fully test for causality. Furthermore, it is important that researchers use modern statistical methods for the analysis of change and pay attention to the potential influence of the choice of the contrast group. In addition, researchers should invest in a uniform research language, making it easier to compare results across studies and integrate findings into broader knowledge.

However, even in the most optimal causal research designs, explaining psycho-therapeutic change remains a challenge. Psychotherapy for depression is a complex, multi-dimensional phenomenon that might work through interplay of multiple mecha-nisms on several levels (e.g. physiological, affective, behavioural and cognitive aspects). Psychotherapeutic change might therefore consist of a complicated chain of events on these different levels. In addition, it is possible that active components of therapy and their associated mechanisms of change work differently at different points in time and differ between (subgroups of) depressed patients. With this in mind, psychotherapeutic change might be too complex to be explained in relatively simple causal models of psy-chological change. If this is the case, research designs might never be able to explain all aspects of therapeutic change. However, it would make it a lot easier to understand why research so far has not led to clear-cut empirical explanations of how psychotherapy for depression works.

Footnotes

[1] If the abstract did not provide all the information necessary to assess in- and exclusion criteria, the full article was consulted.

[2] If a study used a mixed sample (e.g. depression and anxiety) but the main focus was on depression and the majority of the sample was depressed, the study was included.

[3] Because we excluded studies as soon as they did not meet one of the inclusion criteria, the number of studies meeting multiple exclusion criteria is unknown.

[4] Based on 22 studies, two studies did not report on this (Webb et al., 2013; Zettle et al., 2011).

[5] Based on 31 studies; one study did not report on this (Webb et al., 2013).

References

Aalderen van, J. R., Donders, A. R. T., Giommi, F., Spinhoven, P., Barendregt, H. P., & Speckens, A. E. M. (2012). The efficacy of mindfulness-based cognitive therapy in recurrent depressed patients with and without a current depressive episode: a randomized controlled trial. *Psychological Medicine, 42*(5), 989–1001.

Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2011). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80*(1), 93-101.

Allart-van Dam, E., Hosman, C. M. H., & Hoogduin, C. A. L. (2003). The coping with depression course: Short-term outcomes and mediating effects of a randomized controlled trial in the treatment of subclinical depression. *Behavior Therapy, 34*(3), 381–396.

Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 user's guide*. Chicago, IL: Smallwaters Corporation.

Arbuckle, J. L. (1999). *AMOS: a structural equation modeling software*. Chicago: IL: SmallWaters Corporation.

Arbuckle, J. L. (2005). *Amos 6.0 User's Guide*. Spring House, PA: Amos Development Corporation.

Backenstrass, M., Schwarz, T., Fiedler, P., Joest, K., Reck, C., Mundt, C., & Kronmueller, K.-T. (2006). Negative mood regulation expectancies, self-efficacy beliefs, and locus of control orientation: moderators or mediators of change in the treatment of depression? *Psychotherapy Research, 16*(2), 250–258.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.

Batink, T., Peeters, F., Geschwind, N., van Os, J., & Wichers, M. (2013). How does MBCT for Depression Work? Studying Cognitive and Affective Mediation Pathways. *PLoS ONE, 8*(8), 1–13.

Beck, A. T., Steer, R., & Brown, G. K. (1996). *Beck Depression Inventory II: Manual*. Boston, MA: Hartcourt Brace.

Beck, A. T., Ward, C. H., Meldelson, M., Mock, J., & Erbauch, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–571.

Beevers, C. G., Wells, T. T., & Miller, I. W. (2007). Predicting response to depression treatment: The role of negative cognition. *Journal of Consulting and Clinical Psychology, 75*(3), 422–431.

Blalock, J. A., Fouladi, R. T., Cinciripini, P. M., Markowitz, J. C., Klein, D. N., Rothbaum, B. O., . . . McCullough, J. P. (2008). Cognitive and Behavioral Mediators of Combined Pharmacotherapy and Psychotherapy of Chronic Depression. *Cognitive therapy and research, 32*(2), 197–211.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, But What's the Mechanisms? (Don't Expect an Easy Answer). *Journal of Personality and Social Psychology, 98*(4), 550–558.

Cheong, J., MacKinnon, D. P., & Khoo, S. T. (2001). A latent growth modeling approach to mediation analysis. In L. M. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 390–392). Washington, DC: American Psychological Association.

Cheong, J., MacKinnon, D. P., & Khoo, S. T. (2003). Investigation of Mediational Processes Using Parallel Process Latent Growth Curve Modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 10*(2), 238-262.

5

Cohen, S., Daniel O'Leary, K., Foran, H. M., & Kliem, S. (2014). Mechanisms of Change in Brief Couple Therapy for Depression. *Behavior Therapy, 45*(3), 402–417.

Collins, L. M., & Graham, J. W. (2002). The effect of the timing and temporal spacing of observations in longitudinal studies of tobacco and other drug use: Temporal design considerations. *Drug and Alcohol Dependence, 68*, 85–96.

Cuijpers, P., & van Straten, A. (2011). New psychotherapies for mood and anxiety disorders: Necessary innovation or waste of resources? *Canadian Journal of Psychiatry, 56*(4), 251–252.

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology, 76*(6), 909–922.

Curran, P. J., & Hussong, A. M. (2002). Structural Equation Modeling of Repeated Measures Data: Latent Curve Analysis. In D. S. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Methods and applications (pp. 59–85)* (pp. 59–85). Mahwah, NJ: Erlbaum.

DeRubeis, R. J., Evans, M. D., Hollon, S. D., Garvey, M. J., Grove, W. M., & Tuason, V. B. (1990). How Does Cognitive Therapy Work? Cognitive Change and Symptom Change in Cogntive Therapy and Pharmacotherapy for Depression. *Journal of Consulting and Clinical Psychology, 58*(6), 862–869.

Dietz, L. J., Marshal, M. P., Burton, C. M., Bridge, J. A., Birmaher, B., Kolko, D., . . . Brent, D. A. (2014). Social problem solving among depressed adolescents is enhanced by structured psychotherapies. *Journal of Consulting and Clinical Psychology, 82*(2), 202–211.

Fledderus, M., Bohlmeijer, E. T., Fox, J., Schreurs, K. M. G., & Spinhoven, P. (2013). The role of psychological flexibility in a self-help acceptance and commitment therapy intervention for psychological distress in a randomized controlled trial. *Behaviour Research and Therapy, 51*, 142–151.

Forman, E. M., Chapman, J. E., Herbert, J. D., Goetter, E. M., Yuen, E. K., & Moitra, E. (2012). Using Session-by-Session Measurement to Compare Mechanisms of Action for Acceptance and Commitment Therapy and Cognitive Therapy. *Behavior Therapy, 43*(2), 341–354.

Garrat, G., Ingram, R. E., Rand, K. L., & Sawalani, G. (2007). Cognitive Processes in Cognitive Therapy: Evaluation of the Mechanisms of Change in the Treatment of Depression. *Clinical Psychology, 14*, 224–239.

Gaynor, S. T., & Harris, A. (2008). Single-Participant Assessment of Treatment Mediators: Strategy Description and Examples From a Behavioral Activation Intervention for Depressed Adolescents. *Behavior Modification, 32*(3), 372–402.

Goodman, L. A. (1960). On the Exact Variance of Products. *Journal of the American Statistical Association, 55*(292), 708–713.

Haaga, D. A. F., & Stiles, W. B. (2000). Randomized clinical trials in psychotherapy research; Methodology, design and evaluation. In C. R. Snyder & R. E. Ingram (Eds.), *Handbook of psychological change; psychotherapy processes and practices for the 21st Century* (pp. 14–39). New York, NY: Wiley.

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

Haubert, L. C., & Dobson, K. S. (2007). Treatment of Depression and Mechanisms of Change: Strengthening the Links among Theory, Research and Practice. *Clinical Psychology: Science and Practice, 14*, 247–251.

Hill, A. B. (1965). The environment and disease: Association or causation? *Proceeding of the Royal Society of Medicine, 58*, 295–300.

Hollon, S. D., & DeRubeis, R. J. (2009). Mediating the effects of cognitive therapy for depression. *Cognitive Behaviour Therapy, 38*(S1), 43–47.

Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology, 65*(4), 599–610.

Hoyle, R. H., & Kenny, D. A. (1999). Statistical power and tests of mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research*. Newbury Park: Sage.

Hoyle, R. H., & Smith, G. T. (1994). Formulating clinical research hypotheses as structural equation models: A conceptual overview. *Journal of Consulting and Clinical Psychology, 62*(3), 429–440.

Jacobs, R. H., Becker, S. J., Curry, J. F., Silva, S. G., Ginsburg, G. S., Henry, D. B., & Reinecke, M. A. (2014). Increasing Positive Outlook Partially Mediates the Effect of Empirically Supported Treatments on Depression Symptoms Among Adolescents. *Journal of Cognitive Psychotherapy, 28*(1), 3–19.

Jacobs, R. H., Silva, S. G., Reinecke, M. A., Curry, J. F., Ginsburg, G. S., Kratochvil, C. J., & March, J. S. (2009). Dysfunctional Attitudes Scale Perfectionism: A Predictor and Partial Mediator of Acute Treatment Outcome among Clinically Depressed Adolescents. *Journal of Clinical Child and Adolescent Psychology, 38*(6), 803–813.

Jo, B. (2008). Causal inference in randomized experiments with mediational processes. *Psychological Methods, 13*(4), 314–336.

Johansson, P., & Høglend, P. (2007). Identifying mechanisms of change in psychotherapy: Mediators of treatment outcome. *Clinical Psychology & Psychotherapy, 14*(1), 1–9.

Johansson, P., Høglend, P., Ulberg, R., Amlo, S., Marble, A., Bøgwald, K.-P., . . . Heyerdahl, O. (2010). The mediating role of insight for long-term improvements in psychodynamic therapy. *Journal of Consulting and Clinical Psychology, 78*(3), 438–448.

Kaufman, N. K., Rohde, P., Seeley, J. R., Clarke, G. N., & Stice, E. (2005). Potential Mediators of Cognitive-Behavioral Therapy for Adolescents With Comorbid Major Depression and Conduct Disorder. *Journal of Consulting and Clinical Psychology, 73*(1), 38–46.

Kazdin, A. E. (2007). Mediators and Mechanisms of Change in Psychotherapy Research. *Annual Review of Clinical Psychology, 3*, 1–27.

Kazdin, A. E. (2009). Understanding how and why psychotherapy leads to change. *Psychotherapy Research, 19*(4–5), 418–428.

Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, 44*(8), 1116–1129.

Kenny, D. A. (1979). *Correlation and causality*. New York, NY: Wiley.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd ed.). New York, NY: Guilford.

5

Klug, G., Henrich, G., Filipiak, B., & Huber, D. (2012). Trajectories and Mediators of Change in Psychoanalytic, Psychodynamic, and Cognitive Behavioral Therapy. *Journal of the American Psychoanalytic Association, 60*(3), 598–605.

Kolko, D. J., Brent, D. A., Baugher, M., Bridge, J., & Birmaher, B. (2000). Cognitive and Family Therapies for Adolescent Depression: Treatment Specificity, Mediation, and Moderation. *Journal of Consulting and Clinical Psychology, 68*(4), 603–614.

Kraemer, H. C., Stice, E., Kazdin, A. E., Offord, D., & Kupfer, D. J. (2001). How Do Risk Factors Work Together? Mediators, Moderators, and Independent, Overlapping and Proxy Risk Factors. *American Journal of Psychiatry, 158*(6), 848–883.

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*, 877–883.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel Modeling of Individual and Group Level Mediated Effects. *Multivariate Behavioral Research, 36*(2), 249–277.

Kuyken, W., Watkins, E., Holden, E., White, K., Taylor, R. S., Byford, S., . . . Dalgleish, T. (2010). How does mindfulness-based cognitive therapy work? *Behaviour Research and Therapy, 48*, 1105–1112.

Kwon, S., & Oei, T. P. S. (2003). Cognitive change processes in a group cognitive behavior therapy of depression. *Journal of Behavior Therapy and Experimental Psychiatry, 34*, 73–85.

Laurenceau, J.-P., Hayes, A. M., & Feldman, G. C. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review, 27*(6), 682–695.

Lewis, C. C., Simons, A. D., Silva, S. G., Rohde, P., Small, D. M., Murakami, J. L., . . . March, J. S. (2009). The role of readiness to change in response to treatment of adolescent depression. *Journal of Consulting and Clinical Psychology, 77*(3), 422–428.

Lo, H. H. M., Ng, S. M., Chan, C. L. W., Lam, K. F., & Lau, B. H. P. (2013). The Chinese medicine construct "stagnation" in mind–body connection mediates the effects of mindfulness training on depression and anxiety. *Complementary Therapies in Medicine, 21*(4), 348–357.

Longmore, R. J., & Worrell, M. (2007). Do we need to challenge thoughts in cognitive behavior therapy? *Clinical Psychology Review, 27*(2), 173–187.

Longwell, B. T., & Truax, P. (2005). The Differential Effects of Weekly, Monthly, and Bimonthly Administrations of the Beck Depression Inventory-II: Psychometric Properties and Clinical Implications. *Behaviour Therapy, 36*, 265–275.

MacKinnon, D. P. (1994). Analysis of mediating variables in prevention and intervention research. *NIDA Research Monograph, 139*, 127–153.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Lawrence Erlbaum Associates.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation Analysis. *Annual Review of Clinical Psychology, 58*, 593–614.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A Comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83–104.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*(1), 99–128.

Murphy, R., Cooper, Z., Hollon, S. D., & Fairburn, C. G. (2009). How do psychological treatments work? Investigating mediators of change. *Behavioural Research Therapy, 47*(1), 1–5.

Muthén, L. K., & Muthén, B. O. (2001). *Mplus user's guide* (2 ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (6 ed.). Los Angeles, CA: Muthén & Muthén.

Pitariu, A., & Ployhart, R. E. (2010). Explaining change: Theorizing and testing dynamic mediated longitudinal relationships. *Journal of Management, 36*, 405–429.

Preacher, K., & Hayes, A. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*(4), 717–731.

Preacher, K., & Hayes, A. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879–891.

Quilty, L. C., McBride, C., & Bagby, R. M. (2008). Evidence for the cognitive mediational model of cognitive behavioural therapy for depression. *Psychological Medicine, 38*, 1531–1541.

Ryba, M. M., Lejuez, C. W., & Hopko, D. R. (2014). Behavioral activation for depressed breast cancer patients: The impact of therapeutic compliance and quantity of activities completed on symptom reduction. *Journal of Consulting and Clinical Psychology, 82*(2), 325–335.

Schlesselman, J. J. (1982). *Case-control studies: Design, conduct and analysis*. New York, NY: Oxford University Press.

Selig, J. P., & Preacher, K. J. (2009). Mediation Models for Longitudinal Data in Developmental Research. *Research in Human Development, 6*(2–3), 144–164.

Shahar, B., Britton, W. B., Sbarra, D. A., Figueredo, A. J., & Bootzin, R. R. (2010). Mechanisms of Change in Mindfulness-Based Cognitive Therapy for Depression: Preliminary Evidence from a Randomized Controlled Trial. *International Journal of Cognitive Therapy, 3*(4), 402–418.

Shahar, G., Blatt, S. J., Zuroff, D. C., Krupnick, J. L., & Sotsky, S. M. (2004). Perfectionism Impedes Social Relations and Response to Brief Treatment for Depression. *Journal of Social and Clinical Psychology, 23*(2), 140–154.

Shrout, P. E., & Bolger, N. (2002). Mediation in Experimental and Nonexperimental Studies: New Procedures and Recommendations. *Psychological Methods, 7*(4), 422–445.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological Methodology* (pp. 290–312). Washington DC: American Sociological Association.

Stice, E., Presnell, K., Gau, J., & Shaw, H. (2007). Testing mediators of intervention effects in randomized controlled trials: An evaluation of two eating disorder prevention programs. *Journal of Consulting and Clinical Psychology, 75*(1), 20–32.

Stice, E., Rohde, P., Seeley, J. R., & Gau, J. M. (2010). Testing mediators of intervention effects in randomized controlled trials: An evaluation of three depression prevention programs. *Journal of Consulting and Clinical Psychology, 78*(2), 273–280.

Toth, S. L., Rogosch, F. A., Oshri, A., Gravener-Davis, J., Sturm, R., & Morgan-López, A. A. (2013). The efficacy of interpersonal psychotherapy for depression among economically disadvantaged mothers. *Development and Psychopathology, 25*, 1065–1078.

5

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A Meta-Analysis of Outcome Studies Comparing Bona Fide Psychotherapies: Empirically, "All Must Have Prizes". *Psychological Bulletin, 122*(3), 203–215.

Warmerdam, L., van Straten, A., Jongsma, J., Twisk, J., & Cuijpers, P. (2010). Online cognitive behavioral therapy and problem-solving therapy for depressive symptoms: Exploring mechanisms of change. *Journal of Behavior Therapy and Experimental Psychiatry, 41*, 64–70.

Watkins, E. D., Mullan, E., Wingrove, J., Rimes, K., Steiner, H., Bathurst, N., . . . Scott, J. (2011). Rumination-focused cognitive-behavioural therapy for residual depression: phase II randomised controlled trial. *The British Journal of Psychiatry, 199*, 317–322.

Watson, J. C., Steckley, P. L., & McMullen, E. J. (2014). The role of empathy in promoting change. *Psychotherapy Research, 24*(3), 286–298.

Webb, C. A., Kertz, S. J., Bigda-Peyton, J. S., & Björgvinsson, T. (2013). The role of pretreatment outcome expectancies and cognitive–behavioral skills in symptom improvement in an acute psychiatric setting. *Journal of Affective Disorders, 149*(1–3), 375–382.

Zanden van der, R., Galindo-Garre, F., Curie, K., Kramer, J., & Cuijpers, P. (2014). Online cognitive-based intervention for depression: exploring possible circularity in mechanisms of change. *Psychological Medicine, 44*(6), 1159–1170.

Zettle, R. D., Rains, J. C., & Hayes, S. C. (2011). Processes of Change in Acceptance and Commitment Therapy and Cognitive Therapy for Depression: A Mediation Reanalysis of Zettle and Rains. *Behavior Modification, 35*(3), 265–283.

Appendix

Appendix I: Key-term scheme for database search

'Psychotherapy', 'Psychotherapies', 'Psychological Treatment(s)/Intervention(s)', 'Interpersonal (Psycho)therapy', '(Mindfulness-based) Cognitive (Behavio(u)ral) Therapy', 'Psychodynamic/analytic Therapy', 'Client-Centered Therapy', 'Behavio(u)ral Activation', 'Acceptance Commitment Therapy'; 'Mechanisms of Change/Action', 'Working Mechanisms (of psychotherapy)', 'Processes of therapy', 'Process Research', Change'; 'Mediation', 'Mediator', 'Mediating effects'; 'Depression', 'Major Depressive Disorder', 'Dysthymia', 'Dysthymic Disorder.'
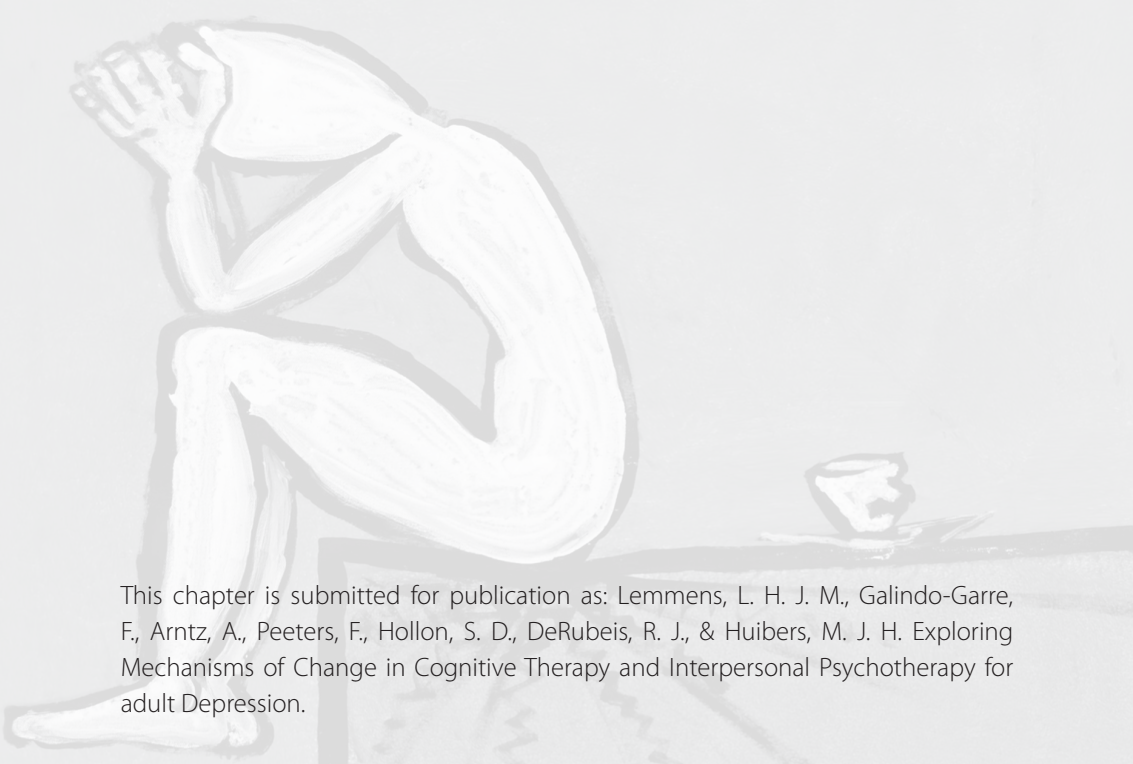
Appendix II: In- & Exclusion criteria

— Published in English in Peer-reviewed Journal.
— Empirical Research report (no review/theoretical paper).
— Focus = Psychological mechanisms of treatment for Depression.
— Population = Diagnosis/Symptoms of Depression.
— Intervention = (Evidence-based) psychotherapy.
— Including clinical outcome measure for depression.
— Performing Statistical Mediation analysis in the sense of Baron & Kenny (1986) or its modern extensions.

5

6

Exploring Mechanisms of Change in Cognitive Therapy and Interpersonal Psychotherapy for adult Depression

Abstract

The present study explored the temporal relationships between change in five candidate causal mechanisms and change in depressive symptoms in the context of a randomized comparison of individual Cognitive Therapy (CT) and Interpersonal Psychotherapy (IPT) for adult depression. Furthermore, hypotheses concerning the mediation of change in these treatments were tested. Patients were 151 depressed outpatients treated with either CT (n=76) or IPT (n=75). Depression severity was assessed with the Beck Depression Inventory-II. Potential mediators were Dysfunctional Attitudes, Interpersonal Problems, Rumination, Self-Esteem and Therapeutic Alliance. All measures were assessed at baseline, mid-, and post-treatment. Pearson's correlations and Latent Difference Score models were used to examine direct and indirect relationships between change in each of the five process measures and change in depression severity. Over the course of treatment, patients showed improvement on all process measures, with medium to large within-treatment effect sizes. No between-treatment differences were observed. Changes in process variables were strongly related to changes in symptoms. Early change in self-esteem was associated with subsequent change in depression. No other temporal relations were found. Change on the BDI-II was mediated by concurrent change in interpersonal problems. No temporal mediation was found. In spite of a well-considered research design and the use of an innovative statistical approach, we found little empirical support for the respective theoretical models of change in CT and IPT. Future tests of causal models of therapeutic change should pay special attention to the timing of assessments as well as to within-patient variance.

6

Introduction

The efficacy of Cognitive Therapy (CT) and Interpersonal Psychotherapy (IPT) for the treatment of Major Depressive Disorder (MDD) has been well established (Cuijpers et al., 2011; Cuijpers, van Straten, Andersson, & van Oppen, 2008; Hollon & Ponniah, 2010; Hollon, Thase, & Markowitz, 2002). Despite this proven efficacy, research has provided relatively little evidence that speaks to the psychological mechanisms through which CT and IPT lead to symptom change, and therefore whether they do so for reasons hypothesized in their respective theoretical backgrounds. The psychological processes that are assumed to be responsible for therapeutic change can be represented in research by measures that are proposed to represent *mediators* (Kazdin, 2007). A mediator is a variable that (statistically) explains why and in what way a treatment has an effect on outcome and can be seen as representing a potential mechanism (the actual process) through which therapeutic change is brought about (Baron & Kenny, 1986; Kraemer, Wilson, Fairburn, & Agras, 2002; MacKinnon, Fairchild, & Fritz, 2007).

Theorized mechanisms of CT and IPT

According to their respective theoretical models of change, different mechanisms are involved in CT and IPT. CT is based on Beck's (1964) cognitive theory which states that depression is caused and maintained by maladaptive information processing strategies and dysfunctional beliefs. It is assumed that interventions aimed at altering the function, content and structure of maladaptive patterns of thinking lead to symptom improvement to the extent that they succeed in changing those cognitive mechanisms. CT, therefore, encourages the patient to identify and evaluate beliefs, schemas, and interpretations associated with negative affect, and replace unrealistic thoughts with alternatives that more closely reflect reality (Beck, Rush, Shaw, & Emery, 1979). In the model that underlies IPT, depression results from significant problems in the interpersonal domain (Klerman, Weissman, Rounsaville, & Chevron, 1984). It is hypothesized that insofar as patients solve their interpersonal problems, or are able to change their emotions in relation to their problems, their depressive symptoms should resolve as well (Markowitz & Weissman, 2004).

The idea that cognitive change accounts for therapeutic change is a popular hypothesis that has motivated dozens of investigations of the role of cognitive change in CT and other treatments for depression (see reviews of e.g., Garrat, Ingram, Rand, & Sawalani, 2007; Longmore & Worrell, 2007; Whisman, 1993 for an overview). Unfortunately, the majority of the relevant studies do not address what is called the 'Achilles heel' (Kazdin & Nock, 2003) and probably the biggest challenge in contemporary mechanism research: establishing the temporal relation between changes in purported mechanisms and symptom change. As a result, it remains unclear whether changes in cognitions precede or follow from changes in depression symptom severity during treatment. Only a few studies of the mechanisms of change in psychotherapy have included efforts to detect

6

the direction of causality (e.g., DeRubeis et al., 1990; Kuyken et al., 2010; Strunk, Brotman, & DeRubeis, 2010; Warmerdam, van Straten, Jongsma, Twisk, & Cuijpers, 2010). In a recent analysis of the relevant literature, Lorenzo-Luaces, German, and DeRubeis (2015) argue that the body of existing research provides some support for the cognitive mediational model, but that the support is not strong. They conclude that the evidence best supports the idea that insofar as cognitive change is a mechanism, it is likely not specific to CT.

Few investigations of the processes of change in IPT have been published. Recently, Toth et al. (2013) examined mediators of sustained treatment effects of IPT in a sample of economically disadvantaged mothers with MDD. They found that changes in perceived stress and social support mediated treatment outcome eight months after treatment termination. However, no studies so far have examined whether changes in interpersonal functioning during the acute phase of treatment for depression mediate outcomes. Nonetheless, in several studies the relation between changes in theorized processes of IPT have been found to correlate with outcome. Bernecker (2012) concluded in her summary of this literature that reduction or resolution of interpersonal problems, reduction of attachment and anxiety avoidance, and improved marital adjustment are indeed related to treatment outcome, and therefore warrant further attention as potential mechanisms. However, in a recent study by the same group (Bernecker, Constantino, Pazzaglia, Ravitz, & McBride, 2014), changes in interpersonal functioning and outcome were not found to be related.

Common factors

Contrary to the view that treatments exert beneficial effects through their own (specific) theorized mechanisms is the idea that treatments work through common factors. One of the most frequently investigated common factors is the therapeutic alliance (Castonguay, Constantino, & Holtforth, 2006). Researchers have consistently reported that a good alliance is associated with better treatment outcomes in various types of psychological treatment, including CT and IPT (see e.g., Horvath, Del Re, Flückiger, & Symonds, 2011), thereby supporting the hypothesis that alliance is a factor that contributes to symptom improvement in psychotherapy. However, it is premature to conclude that alliance is a mechanism through which change occurs. Not only have the effect sizes reported in the literature been relatively small (overall relation of $r = .28$ in the most recent meta-analysis by Horvath et al., 2011), but also – similar to studies examining the specific mechanisms of CT and IPT – few studies have controlled for temporal confounds (Barber, 2009).

Studies in which temporal priority has been accounted for have produced mixed results. For example, Webb et al. (2011) found support for the notion that alliance – more specifically agreement on goals and tasks – facilitates subsequent symptom change in CT, whereas others have not found the alliance to be predictive of subsequent change in depression severity (e.g., DeRubeis & Feeley, 1990; Feeley, DeRubeis, & Gelfand, 1999;

Gaston, Marmar, Gallagher, & Thompson, 1991; Strunk et al., 2010). We are not aware of any studies that have controlled for the temporal aspect of measurement of the alliance in IPT. This is remarkable because previous research has shown predictive relations between alliance and outcome in treatment modalities with a strong emphasis on interpersonal processes, such as Expressive Dynamic Therapy (Barber, Connolly, Crits-Christoph, Gladis, & Siqueland, 2000) and CBASP (Klein et al., 2003). Since IPT is directly aimed at improving interpersonal functioning, one could argue that alliance might play an important role in IPT as well. Future research should therefore also focus on the causal influence of alliance on symptom change over the course of IPT. In addition, other common factors such as facilitation of self-esteem and reduction of rumination are largely overlooked as potential mechanisms of change and need attention.

Current study

Apart from the theoretical and methodological difficulties that were discussed above, progress in research on therapeutic mechanisms has been impeded by small sample sizes, and the fact that there is little consensus concerning the statistical approaches and specific methods to be used in tests of mediation. Furthermore, because many of the studies have lacked a comparison group, they have not allowed for a direct comparison of mediation patterns between interventions. As a result, there is a need for additional studies in designs suitable for mechanism research. More specifically, for studies that compare treatments head-to-head, and test for mediation in large, longitudinal research studies using up-to-date statistical analysis techniques.

Thus far, there is only one randomized comparison of CT and IPT in which mediation has been a focus (Quilty, McBride, & Bagby, 2008). Quilty and colleagues examined evidence for the cognitive mediational model and found that CT produced greater change in dysfunctional attitudes than did IPT. Furthermore, they concluded that change in dysfunctional attitudes mediated the effect of CT on depression. However, given their pre- to post-treatment design they could not address the issue of temporal precedence. Finally, Quilty et al. focused only on cognitive change. As a result, the role of interpersonal variables and common factors in a randomized comparison of CT and IPT remains to be studied.

Recently we conducted an RCT investigating the effectiveness of individual CT and IPT for the treatment of adult depression (Lemmens et al., 2011, 2015). We now examine the role of five candidate mediators of CT and IPT – specific and non-specific – over the course of treatment. A measure of dysfunctional attitudes was included to represent a theorized mechanism of CT. The interpersonal theory was represented by a measure of interpersonal problems. Rumination[1], self-esteem and the quality of the working alliance were included as potential common factors (see Figure 1)[2]. In a design suitable to test for longitudinal mediation, we examined whether scores on the various process measures changed over the course of therapy and whether changes on these measures distin-

guished the two conditions. Furthermore, for each potential mediator we examined its relation with change in depressive symptoms, both contemporaneously and as a predictor of subsequent change in the outcome variable. To conclude, we investigated whether the theorized processes mediated the relation between treatment and outcome.
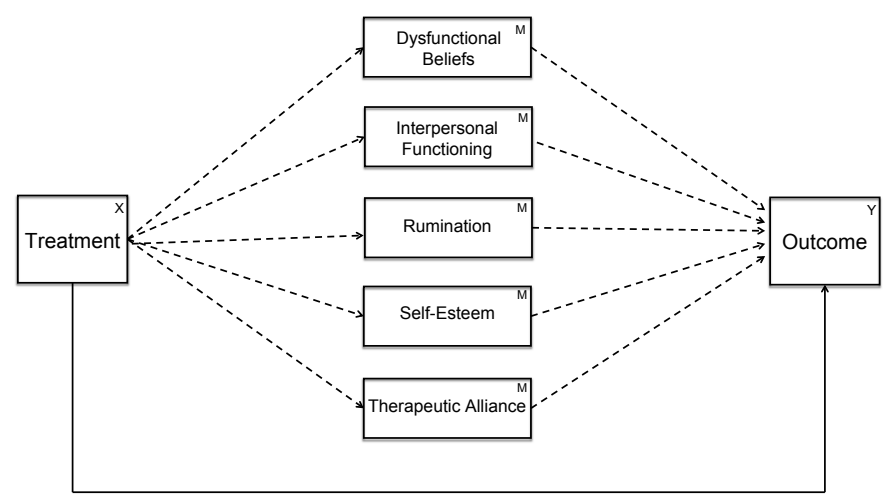
Figure 1. Theorized and Estimated models of direct and indirect effects of psychotherapy on depression severity through five potential mediators. *Note:* M = Potential Mediator.

We hypothesized that scores on the process measures would each, on average, change in an adaptive direction over the course of treatment (i.e. increases in self-esteem and working alliance, and decreases in dysfunctional attitudes, interpersonal problems and rumination). Based on theoretical underpinnings we expected the effects of CT to be mediated by cognitive change (Beck et al., 1979), and the effects of IPT by change in interpersonal functioning (Klerman et al., 1984), thereby indicating that different processes lead to similar treatment outcomes. However, since previous research (e.g., Bernecker et al., 2014; Quilty et al., 2008; Warmerdam et al., 2010) suggests that change in theorized processes may not be treatment-modality-specific, we did not rule out the possibility that both treatments might work through similar mechanisms (Hollon, DeRubeis, & Evans, 1987). Given previous reports on the alliance (Barber et al., 2000; Klein et al., 2003), we hypothesized that the alliance would have an impact in both CT and IPT.

Methods

Design and Participants

Data were collected as part of a large RCT examining the clinical effects and mechanisms of change of individual CT and IPT for depression (Lemmens et al., 2011, 2015). A total of 182 depressed outpatients were randomly allocated to CT ($n = 76$), IPT ($n = 75$), or a 2-month Waiting-List Control condition followed by treatment of choice (WLC; $n = 31$). For the purpose of the present study, we focus only on the two active conditions and data collected in the acute phase (baseline to 7 months). Details concerning the design of the study, participants, interventions, and outcome are reported elsewhere (Lemmens et al., 2011, 2015) and will therefore only be briefly summarized here.

Participants were adult outpatients (18 – 65 years) referred to the mood disorder program of the Academic Community Mental Health Centre Maastricht (the Netherlands) with a primary diagnosis of MDD confirmed by the Dutch version of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I; First, Spitzer, Gibbon, & Williams, 1997). Further inclusion criteria were: internet access, an e-mail address, and sufficient knowledge of the Dutch language. Patients receiving concomitant pharmacological or psychological treatment at baseline were excluded from the study, as were those at imminent risk for suicide. Other exclusion criteria were: bipolar or chronic depression (current episode > 5 years); substance abuse/dependence and mental retardation (IQ < 80). Treatment consisted of 16 to 20 individual 45-minute sessions (Mean = 17, SD = 2.9). The CT protocol followed the guidelines laid out by Beck et al. (1979). The IPT protocol was based on the manual by Klerman et al. (1984).

All participants provided their informed consent before they participated in the study. The study was approved by Maastricht University's Ethical Board, and was conducted in accord with relevant ethical guidelines. The study is registered at the Netherlands Trial Register, part of the Dutch Cochrane Centre (ISRCTN67561918). Both treatments led to considerable improvement in depression severity (pre-post treatment effect size $d = 1.72$ in the pooled active conditions) that was sustained up to one year. Response to therapy exceeded response in the WLC condition. No differential effects between treatments were found. Quality of therapy was rated by independent assessors as being "(very) good" to "excellent" in both conditions (Lemmens et al., 2015).

Outcome measure

Depressive Symptoms – The primary outcome was depression severity as measured with the Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996; Dutch translation by van der Does, 2002a). The BDI-II is a 21 item self-report questionnaire with items rated on

a 4-point Likert-scale (0 – 3), with higher scores indicating higher levels of depression severity (range is 0 – 63). The instrument has strong psychometric properties as a screening measure for depressive symptom severity (Beck et al., 1996; van der Does, 2002a).

Process (or mediator) variables

Dysfunctional Beliefs – The presence and intensity of dysfunctional beliefs – a core concept of Beck's cognitive theory – was assessed with the 17-item Dysfunctional Attitudes Scale (form A) Revised (DAS-A17; de Graaf, Roelofs & Huibers, 2009). In this shortened version of the original DAS-A by Weissman and Beck (1978), respondents rate the extent to which they agree with a series of dysfunctional assumptions. Items are rated on a 7-point Likert-scale (1 = fully disagree to 7 = fully agree). Higher scores reflect more negative and absolute thinking (range is 17–119). The DAS-A17 is a valid measure of dysfunctional cognitions in depressed patients (de Graaf et al., 2009).

Interpersonal Functioning – Interpersonal Functioning was assessed with the 64-item version of the Inventory of Interpersonal Problems (IIP-64, henceforth IIP, Horowitz, Alden, Wiggins, & Pincus, 2000). The IIP is a self-report measure that describes the types of problems that people experience in their relationships with others, and the level of distress associated with this. Items are rated on a 5-point Likert-scale (0 – 4) with higher scores indicating more interpersonal problems (range is 0 – 256). The IIP has been shown to have good psychometric properties (Horowitz et al., 2000; Horowitz et al., 1988).

Rumination – The Dutch version of the Ruminative Response Scale (RRS-NL; Raes, Hermans, & Eelen, 2003), originally developed by Nolen-Hoeksema and Morrow (1991), was used to obtain a measure of rumination. The RRS-NL includes 22 items describing responses to depressed mood that are focused on self, symptoms, or possible consequences and causes of mood. Participants read each item and indicate how often they think or do this when they feel sad or depressed. Items are rated on a four point Likert-scale ranging from almost never (1) to almost always (4). The total score is the sum of the 22 items (range is 22 – 88). The RRS-NL shows good reliability and satisfactory validity (Raes et al., 2003).

Self-Esteem – The Self-Liking and Self-Competence Scale Revised (SLCS-R; Tafarodi & Swann, 2001, Dutch version by Vandromme, Hermans, Spruyt, & Eelen, 2007) is a self-report measure of self-competence and self-liking, two dimensions of self-esteem. The SLSC-R contains eight items, each rated on a 5-point Likert scale, for each of the two dimensions. Subscale scores can range from 8 to 40, with higher scores indicating higher self-competence or higher self-liking. A total score is calculated by summing the score of the two subscales. Psychometric properties of the SLSC-R are described as good (Vandromme et al., 2007).

Therapeutic Alliance – The observer rated version of the Working Alliance Inventory–Short (WAI-O-S; Tichenor & Hill, 1989; Tracey & Kokotovic, 1989) was used to obtain a measure of the quality of the therapeutic alliance. The WAI-O-S is based on Bordin's

(1979) conceptualization of working alliance which states that a strong alliance forms if a therapist and client a) agree on the *goals* of therapy, b) agree on the *tasks* needed to meet those goals, and c) have a *bond* between them that will facilitate this process. Each of the 12 items of the scale (4 for each subscale) are rated on a 7-point scale (1 = never, to 7 = always), and a higher score indicates a stronger alliance. A total score is obtained by summing up all item scores (range is 12 to 84). Psychometric properties of the WAI-O-S have been found to be good (Andrusyna, Luborsky, Pham, & Tang, 2006; Horvath, 1994; Strunk et al., 2010).

Recruitment and Procedure

Self-report measures (BDI-II, DAS-A17, IIP, RRS, SLSC-R) were administered at baseline (0 months), mid-treatment (3 months) and post-treatment (7 months). Baseline and post-treatment assessments were administered at the research centre. Mid-treatment assessment (3 months) took place via the internet. Of the 151 patients that were included in this study, 3 patients (2%) did not complete the 3-month assessment. At seven months, data of 14 participants (9%) were missing.

All therapy sessions were videotaped. A selection of these tapes (three per patient) were watched by independent raters and were rated on the quailty of the therapeutic alliance. Raters were five clinical psychology master students from Maastricht University (the Netherlands), and four experts in the field of CT and IPT. Prior to the study, raters studied relevant literature. In addition, students received approximately 14 hours (7x2) of group training aimed at getting familiar with the treatment models, the working alliance concept, and the rating instrument. The professionals, who were already familiar with the treatments and the working alliance, received 4 hours of additional training in how to use the rating instrument. All raters received a detailed instruction manual. All identifying information was removed from the tapes to ensure that raters were blind to session number, treatment outcome and other patient characteristics. Due to the specific therapeutic interventions and the visual character of the study it was not possible to blind raters for treatment allocation and therapist. Ratings were made independently after watching an entire session.

In order to obtain a measure of alliance comparable to the fixed assessment points of the self-report data, we selected therapy sessions that were closest to the assessment points at baseline, 3 and 7 months. Of the 453 sessions (3*151) that were selected, tapes of 192 sessions were unavailable. This was either because patients did not give permission for videotaping their sessions (*n* = 40, resulting in 120 missing tapes), or because recordings of specific sessions were missing or damaged (*n* = 72). As a result, 261 tapes were available for analyses (*n* = 88 at baseline, *n* = 97 at 3 months, and *n* = 76 at 7 months). Other details about the recruitment procedure as well as the flow diagram for the study are given elsewhere (Lemmens et al., 2011, 2015).

6

Data analysis

Data were analysed according to the intention-to-treat principle. Descriptive statistics and correlation analyses were carried out using SPSS version 21 for Windows. Longitudinal Mediation analyses were carried out in M-Plus version 7.1. A full information maximum likelihood estimation algorithm was used for the analyses. Because this algorithm uses all available information under the assumption that the incomplete data are missing at random, missing values were not imputed. Normality and variability of the data distributions were checked by Skewness and Kurtosis statistics, and histograms and scatter-plots. Mean scores (SD) at each time-point were determined, and group differences were investigated using independent samples t-tests. Pre to post-treatment effect sizes were calculated to examine the extent to which scores on the various process measures changed over the course of treatment. Within-condition change was defined as Cohen's d = (pre-treatment mean − post-treatment mean) / ($\sqrt{}$baseline variance). Between-group effect sizes were determined by calculating the difference between the within-condition effect sizes of CT and IPT. Effect sizes were classified as being small (±0.2), medium (±0.5) and large (±0.8) (Cohen, 1988). In addition, we explored the pattern of change over the course of treatment and the degree to which there was variability in the data by inspecting descriptive statistics of change scores and by calculating Pearson's correlations between scores on each measure at different time points (e.g., IIP baseline with IIP 3 months).

Pearson's correlations were also used to explore the relation between (change in) process measures and (change in) severity of depression. First, we calculated correlations between the BDI-II scores at the different time points (0, 3, 7 months) and scores on the mediator variables at each of the three time points (cross-sectional correlations). We then estimated correlations between change scores on the BDI-II and change indexes of the various process measures (longitudinal correlations). In order to do this, we computed change scores on the BDI-II and each of the mediator variables from baseline to 3 months (pre- to mid-treatment), from 3 to 7 months (mid- to post-treatment) and from baseline to 7 months (entire course of treatment). To facilitate the interpretation of findings, change scores were calculated in such a way that a positive score always reflected improvement and a negative change score reflected worsening. In evaluating these longitudinal correlations, we made a distinction between temporal and concurrent relations. Temporal relations represent correlations between change in the mediator and change in the BDI-II that precede or follow those changes were (e.g., correlation of change in alliance from 0 to 3 months with change in depression from 3 to 7 months). In case of a concurrent relation, the correlation was estimated of a change score on one measure with change on another measure over the same time period (e.g., change in depression from 0 to 3 months associated with change in cognitions from 0 to 3 months). Correlations were classified as being small (±0.1), medium (±0.3) and large (±0.5).
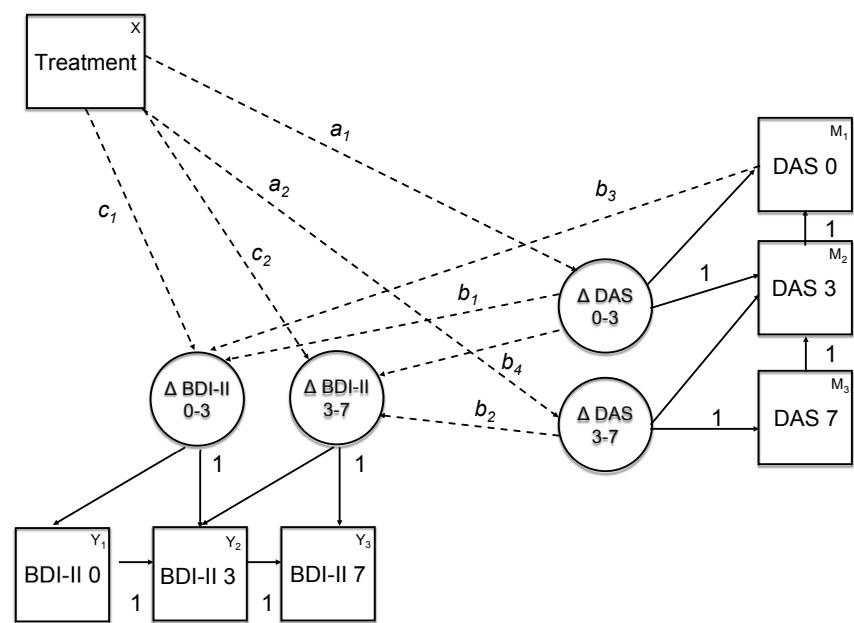
To examine the various relations between change in the process measures and change in depressive symptoms more closely, mediation analyses for longitudinal data was performed. A simple method to assess mediation in longitudinal data is to compute

pre- to post-treatment change scores in process measures and outcome, and then use these scores as variables in a single mediation model (MacKinnon, 2008). However, this method has some drawbacks for use in the current study. First of all, the model only includes pre- and post-treatment scores, and therefore ignores information collected at 3-months (mid-treatment). Second, the model does not take into consideration that the effect of the therapy might vary in different phases of treatment. Latent Difference Score (LDS) models overcome these drawbacks. LDS models use latent instead of observed difference scores to test questions about temporal relationships between variables, which are not affected by measurement errors in the pre- and post-treatment scores. Latent difference scores are defined as that part of the measurement that is not identical to the score obtained at the previous measurement. LDS models allow for examination of change per measurement interval instead of over the entire course of treatment. Therefore they were considered a better alternative for answering our specific research questions (see McArdle, 2009). LDS models are estimated under the Structural Equation Modeling (SEM) framework.

For the current study we used the LDS model proposed by MacKinnon (2008). This model tests the relation between treatment and outcome (*c* path), the relation between treatment and mediator (*a* path), and the set of possible concurrent and temporal relations between mediator and outcome (*b* paths). Figure 2 illustrates the proposed model. Y represents the outcome variable BDI-II. M represents the potential mediator (DAS-A17 in this example) and X refers to the contrast between CT and IPT. The squares in the model represent the observed raw scores at each time point. Circles represent latent variables (defined as the difference between the scores on two consecutive assessments). The b_1 and b_2 paths represent concurrent relationships between mediator and outcome for 0–3 and 3–7 months respectively. Path b_3 and b_4 illustrate temporal relationships. b_3 refers to the relationship between previous levels of the mediator and subsequent changes in the outcome. To conclude, b_4 illustrates the relationship between change in the mediator at time t-1 and change in the outcome at time t. In order to obtain the latent difference scores, two paths in the model – the paths from time 1 to time 2, and the paths from the latent difference to time 2 – were set to 1 (MacKinnon, 2008)[3].

We ran five LDS models, one for each of the process measures. In order to rule out reverse causality (i.e. symptom change facilitates change in mediator variables), we reran our analyses but now added the relation between change in symptoms from (Δ BDI-II 0–3 months) and subsequent change in the mediator in the mediator (Δ in mediator 3 – 7) to the model, and compared its fit with the fit of the model without this relation.

As a last step, the presence and magnitude of mediation was determined by estimating the various indirect effects – the product of the coefficients of paths *a * b*. Indirect effects represent the effect of treatment on outcome that goes through the mediator (e.g. when using the example in Figure 2, $a_1 * b_4$, is the effect of treatment on change in depression in the period 3–7 that is produced by change in DAS-A17 in de period 0–3; MacKinnon, 2008; Muthén & Asparouhov, 2014). We used bias corrected 95% bootstrap

Figure 2. Latent Difference Score (LDS) model to examine the various relations between treatment, process measure (in this example cognition assessed with DAS-A17), and outcome BDI-II in the acute phase of treatment (0–3–7 months). *Note:* The 1's in the model are the necessary restrictions to obtain the latent difference scores (MacKinnon, 2008, *p* 215).

confidence intervals (CI) with 1000 bootstrap samples to assess the parameter estimates. If zero was not contained in the confidence intervals we concluded that the indirect effect was significant. Indirect effects referred to effects in both treatments. When an indirect effect was found to be significant, we explored the differences between the two treatment conditions by estimating the various b paths separately for CT and IPT.

Results

Descriptive statistics

Table 1 presents the mean (SD) scores on the outcome variable BDI-II and on all potential mediator variables at each time-point in the treatment phase (0, 3, 7 months). Furthermore, effect sizes over the course of treatment are presented. All data were normally distributed, except for WAI-O-S at 3 months (Kurtosis $= 2.14$). CT and IPT showed significant baseline differences on the RRS (t (149) $= -2.4$, $p = .02$) and the WAI-O-S (t (81.2) $= 3.5$, $p < .01$)[4]. Over the course of treatment, the means of all process measures changed in the expected direction. Pre- to post-treatment (0–7 months) effect sizes were large ($d > .80$)

for BDI-II and RRS, medium to large ($0.5<d<0.8$) for DAS-A17, SLSC-R and IIP, and small to medium ($0.20<d<0.50$) for the WAI-O-S. Between-group effect sizes were small (range 0.06 to 0.19), except for the IIP ($d=0.30$). Change in process measures DAS-A-17, IIP, RRS, and SLSC mainly occurred in the second half of treatment (3–7 months; see Table 1). Correlations between measures of the same construct at different time points (e.g., correlation between DAS-A-17 score at baseline and three months) ranged from 0.65 to 0.70.

Table 1. Mean Scores and Standard Deviations (SD) on the primary outcome and all potential mediator variables at each time-point in the acute phase (total sample and stratified per condition).

	Pre-treatment	Mid-treatment	Post-treatment	Effect size *d*	
	Mean (SD)	Mean (SD)	Mean (SD)	Within-Groups[a]	CT vs. IPT[b]
Outcome variable					
- Depression Severity (BDI-II)					
Total	29.8 (9.0)	23.2 (12.0)	14.8 (12.1)	1.67	
CT	28.4 (8.9)	22.4 (11.0)	13.7 (10.8)	1.65[5]	
IPT	31.2 (8.9)	24.1 (13.0)	16.0 (13.4)	1.71	-0.06
Process (mediator) variables					
- Dysfunctional Attitudes (DAS-A17)					
Total	62.0 (16.0)	60.3 (16.8)	53.1 (16.0)	0.56	
CT	61.1 (15.4)	60.0 (15.0)	51.4 (14.7)	0.63	
IPT	62.9 (16.6)	60.7 (18.6)	54.9 (17.3)	0.48	0.15
- Interpersonal Problems (IIP)					
Total	86.4 (29.7)	81.3 (32.6)	67.4 (36.7)	0.64	
CT	83.1 (24.7)	82.8 (28.1)	62.9 (36.3)	0.82	
IPT	89.7 (33.9)	79.8 (36.9)	72.0 (36.8)	0.52	0.30
- Rumination (RRS)					
Total	50.7 (8.9)*	47.9 (10.5)	42.0 (11.5)	0.98	
CT	49.1 (9.1)	47.0 (9.7)	40.9 (10.9)	0.90	
IPT	52.4 (8.4)	48.9 (11.2)	43.2 (12.1)	1.10	-0.19
- Self-esteem (SLSC-R)					
Total	38.4 (9.7)	39.7 (9.8)	44.2 (11.2)	0.60	
CT	39.3 (8.6)	39.8 (9.4)	44.9 (10.6)	0.65	
IPT	37.6 (10.7)	39.6 (10.4)	43.4 (11.9)	0.54	0.11

Table 1. (continued)

	Pre-treatment	Mid-treatment	Post-treatment	Effect size *d*	
	Mean (SD)	Mean (SD)	Mean (SD)	Within-Groups[a]	CT vs. IPT[b]
- Alliance (WAI-O-S)					
Total	65.0 (7.4)*	66.7 (7.8)*	67.9 (10.3)*	0.39	
CT	67.8 (5.2)	68.7 (6.3)	70.4 (8.6)	0.50	
IPT	62.7 (8.1)	64.6 (8.7)	65.2 (11.3)	0.31	0.19

Note: CT = Cognitive Therapy; IPT = Interpersonal Psychotherapy; Self-report data unavailable for $n = 3$ (1 x CT, 2 x IPT) and $n = 14$ (7 x CT, 10 x IPT) at 3 and 7 months respectively (except for SLSC-R 0 & IIP 0 = 1 missing (CT), and SLSC-R 7 & IIP 7 = 15 missings (6 x CT, 9 x IPT); Observer rated data available of 88, 97 and 76 patients at baseline, 3 and 7 months respectively; * = significant difference between CT and IPT (RRS 0 → $t(149) = -2.4$, p = .02; WAI-O-S 0 →: $t(81.2) = 3.5$, $p < .01$; WAI-O-S 3 → $t(95) = 2.6$, $p = .01$; WAI-O-S 7 → $t(74) = 2.2$, $p = .03$). [a] = $(M t_0 - M t_7)/SD\ t_0$; [b] = difference in improvement effect sizes between CT and IPT.

Relationships between change in process measures and change in depression severity

Correlation analyses

Pearson's correlations between (change in) BDI-II change scores and change scores for the five potential mediator variables can be found in Table 2. Changes in the BDI-II and the various process measures were highly correlated (both cross-sectionally and longitudinally). Cross-sectional correlations show that higher levels of depression severity were associated with higher scores on the DAS-A17, IIP, and RRS (positive correlations), and lower scores on the SLSC-R, and WAI-O-S (negative correlations). It has to be noted however, that the WAI-O-S showed small, non-significant correlations with the BDI-II. In line with findings on cross-sectional correlations, improvement in depression severity was associated with improvement in the process measures (positive correlations between change scores). A further evaluation of the longitudinal correlations showed that the highest correlations were found for the concurrent relations over the entire course of treatment (0–7, cells with bold printed correlation coefficients), with correlations ranging from .33 to .60. A closer look at these correlations showed that this was mainly the result of high correlations in the second part of treatment (3–7). The temporal relations between BDI-II and the process measures were found to be small and non-significant (cells with dark grey shade).

Table 2. Pearson's correlations between (changes in) BDI-II scores and (changes in) scores on the five process measures (*n*=151).

		BDI-II					
		0 m	3 m	7 m	Δ0→3	Δ3→7	Δ0→7
DAS-A17	0 m	.43**	.32**	.21*	.00	.16	.15
	3 m	.22**	.18*	.16	-.02	.06	.04
	7 m	.26**	.29**	.52**	-.12	-.29**	-.35**
	Δ0→3	.26**	.17**	.05	**.03**	**.12**	.14
	Δ3→7	.03	-.07	-.41**	.11	**.42**	.46**
	Δ0→7	.23**	.07	-.33**	.13	.48**	**.53**
IIP	0 m	.48**	.40**	.30**	-.04	.10	.06
	3 m	.44**	.45**	.31**	-.16	.19*	.01
	7 m	.34**	.47**	.62**	-.28**	-.18*	-.39**
	Δ0→3	.01	-.12	-.05	**.17***	**-.14**	.06
	Δ3→7	.05	-.08	-.44**	.14	**.44**	.51**
	Δ0→7	.06	-.20*	-.46**	.31**	.31**	**.53**

		BDI-II					
		0 m	3 m	7 m	Δ0→3	Δ3→7	Δ0→7
RRS	0 m	.50**	.40**	.20*	-.03	.20*	.16
	3 m	.50**	.60**	.35**	-.29**	.25**	.00
	7 m	.26**	.46**	.65**	-.34**	-.23**	-.49**
	Δ0→3	-.10	-.34**	-.20*	**.34**	**-.09**	.18*
	Δ3→7	.17	.03	-.40**	.12	**.52**	.56**
	Δ0→7	.12	-.16	-.47**	.31**	.38**	**.60**
SLSC-R	0 m	-.41**	-.34**	-.16	.05	-.22*	-.17
	3 m	-.52**	-.55**	-.26**	.20*	-.34**	-.14
	7 m	-.42**	-.48**	-.53**	.22*	.05	.23**
	Δ0→3	-.14	-.24**	-.12	**.18***	**-.15**	.02
	Δ3→7	.08	.02	-.39**	.06	**.47**	.47**
	Δ0→7	-.05	-.21*	-.46**	.22*	.30**	**.45**
WAI-O-S	0 m	-.14	-.04	-.17	-.07	.16	.08
	3 m	-.10	-.16	-.30**	.12	.16	.28**
	7 m	-.22	-.29**	-.37**	.15	.13	.23*
	Δ0→3	.09	-.08	-.14	**.17**	**.04**	.24*
	Δ3→7	-.29*	-.21	-.23	-.01	**.05**	.03
	Δ0→7	-.03	-.26*	-.32*	.31*	.10	**.33**

Note: BDI-II = Beck Depression Inventory-II; DAS-A17 = Dysfunctional Attitudes Scale (form A) Revised; IIP = Inventory of Interpersonal Problems; RRS = Ruminative Response Scale; SLSC-R = Self-Liking and Self-Competence Scale Revised; WAI-O-S = Observer rated version of the Working Alliance Inventory Short; Cells with a dark grey colour represent temporal relations between process and BDI-II, cells with bold printed correlation coefficients = Concurrent relation between process and BDI-II; Change scores were calculated in such a way that a positive score always reflects improvement and a negative change score always reflects worsening. As a result, a positive correlation between change scores indicates that improvement in depression severity is related to improvement on the process measure–and vice versa. Negative correlations between change scores indicate that improvement in depression severity is related to deterioration on the process measures, and vice versa.

6

Latent Difference Score Models

Beta coefficients and Standard Errors of the various relations in the five LDS models are displayed in Table 3. The models for the DAS-A-NL and WAI-O-S showed a good fit, whereas the other models showed a poor to reasonable fit (Kline, 2005)[6]. Reflecting the fact that no main effect of treatment condition was found, standardized regression coefficients representing relations between treatment and outcome (c paths) were all non-significant. Examination of the relation between treatment and change in the mediators (a paths) revealed only a significant difference between CT and IPT on the IIP, indicating that changes on the IIP were different for CT and IPT in different phases of treatment. IPT showed a larger decrease in interpersonal problems in the first half of treatment (0–3), whereas CT showed greater change in the second half (3–7; see Figure 3). Over the entire course of treatment, these effects cancelled each other out.

Table 3. Unstandardized coefficients (standard errors in parentheses) of the five Latent Difference Score (LDS) Models.

		DAS-A17 B (SE)	IIP B (SE)	RRS B (SE)	SLSC-R B (SE)	WAI-O-S B (SE)
Direct Effects	c_1	-0.40 (1.45)	0.29 (1.52)	-0.21 (1.35)	0.16 (1.35)	-0.41 (1.81)
	c_2	0.93 (1.24)	-0.31 (1.35)	0.85 (1.27)	1.02 (1.35)	-0.11 (1.69)
	a_1	-0.31 (1.87)	-9.28 (3.74)[*]	-1.00 (1.29)	-1.54 (1.19)	0.88 (1.20)
	a_2	1.72 (2.04)	9.79 (4.48)[*]	1.59 (1.55)	1.10 (1.41)	0.63 (1.61)
	b_1	0.00 (0.07)	0.09 (0.03)[**]	0.52 (0.12)[***]	0.42 (0.13)[**]	0.28 (0.16)[+]
	b_2	0.46 (0.05)[***]	0.20 (0.03)[***]	0.57 (0.08)[***]	0.62 (0.08)[***]	0.17 (0.15)
	b_3	0.16 (0.05)[**]	0.07 (0.03)[*]	0.20 (0.10)[*]	-0.04 (0.08)	0.38 (0.16)[*]
	b_4	0.11 (0.06)[+]	0.02 (0.03)	0.05 (0.09)	0.22 (0.10)[*]	-0.09 (0.21)
Indirect Effects	a^*b_1	0.00 (0.14)	-0.84 (0.45)[§]	-0.52 (0.68)	-0.64 (0.61)	0.25 (0.40)
	a^*b_2	0.79 (0.98)	1.94 (0.92)[§]	0.91 (0.93)	0.68 (0.90)	0.11 (0.37)
	a^*b_3	-0.05 (0.31)	-0.63 (0.35)[§]	-0.20 (0.31)	0.06 (0.19)	0.34 (0.51)
	a^*b_4	-0.04 (0.26)	-0.15 (0.31)	-0.05 (0.15)	-0.33 (0.29)	-0.08 (0.33)

Note: BDI-II=Beck Depression Inventory-II; DAS-A17=Dysfunctional Attitudes Scale (form A) Revised; IIP=Inventory of Interpersonal Problems; RRS=Ruminative Response Scale; SLSC-R=Self-Liking and Self-Competence Scale Revised; WAI-O-S=Observer rated version of the Working Alliance Inventory Short; c_1 represents the relation between treatment and Δ in BDI-II from 0 to 3 months; c_2 represents the relation between treatment and change in BDI-II from 3 to 7 months; a_1 represents relation between treatment and Δ in the mediator from 0 to 3 months; a_2 represents relation between treatment and Δ in the mediator from 3 to 7 months; b_1 represents the concurrent relation between Δ in BDI-II and Δ in the Mediator from 0 to 3 months; b_2 represents the concurrent relation between Δ in BDI-II and Δ in the Mediator from 3 to 7 months; b_3 represents the temporal relation between absolute score on the mediator at 3 months and Δ in BDI-II from 3 to 7 months; b_4 represents the relation between Δ in mediator from 0 to 3 months and Δ in BDI-II from 3 to 7 months; a^*b paths (indirect effects) represent the effect of treatment on outcome that goes through the mediator (e.g., path $a_1^*b_4$ represents the effect of treatment on change in depression severity from 3–7 months that is produced by change in mediator from baseline to 3 months); $+ p < .10$, $* p < .05$; $** p < .01$; $*** p < .001$; $§$=significant indirect effect based on 95% CI with 1000 bootstrap samples (-2.01 to -0.06 for a^*b_1; 0.19 to 4.06 for a^*b_2; and -1.60 to -0.05 for a^*b_3).

With regard to the various relations between change in mediator and change in outcome (*b* paths), significant concurrent relations in both phases of treatment (represented by b_1 and b_2) were found between change in depression severity and changes in interpersonal problems (IIP), rumination (RRS) and self-esteem (SLSC-R). For the DAS-A17, only a concurrent relation in the second phase of treatment was found (b_2). The WAI-O-S did not show any significant concurrent relations. Furthermore, the absolute scores on the DAS-A17, IIP, RRS and WAI-O-S at three months were significantly related to change in depression severity in the period thereafter (b_3). In addition, a significant temporal relation between early change in the SLSC-R (0–3) and subsequent change on the BDI-II (3–7) was found, indicating that change in self-esteem might precede change in depression severity (b_4). Adding the reversed temporal relation to the model resulted in only minimal change in model fit estimates, indicating that the complex model was not better than the fit of the simple model. It was therefore concluded that the reversed path did also not show a substantial temporal relation.
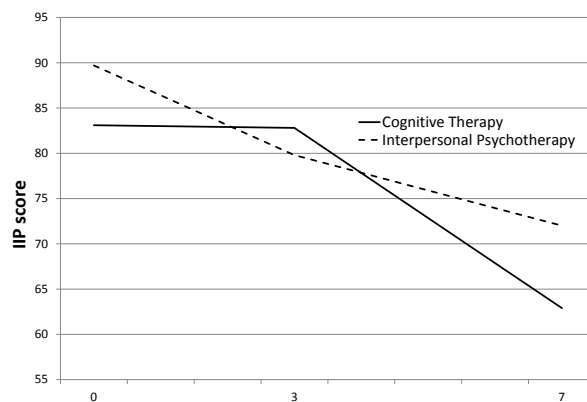
Figure 3. Estimated change in Interpersonal Functioning (IIP) for Cognitive Therapy and Interpersonal Psychotherapy.

Indirect effects

As can be seen in Table 3, we found three significant indirect effects (a*b), all on the IIP. The significant $a*b_3$ path (95% CI from -1.60 to -0.05) indicates that change on the BDI-II in the early phase of treatment (0–3), was mediated by initial levels of IIP. Furthermore, change on the BDI-II was mediated by concurrent change on the IIP both in the early phase ($a*b_1$: 95% CI from -2.01 to -0.06) and late phase of treatment (3–7 months $a*b_2$: 95% CI from 0.19 to 4.06). No temporal mediation ($a*b_4$) was found (95% CI of -0.97 to 0.52). Exploration of the various *b* paths separately for CT and IPT (Table 4) suggests that the relation between changes in interpersonal functioning (IIP) and depression sever-

ity (BDI-II) was different for these treatments. For IPT there is a concurrent relationship between change in IIP and changes in BDI from 3–7 months, and evidence for a longitudinal relationship. For CT, we only found concurrent relationships.

Table 4. Unstandardized coefficients (standard errors) of the various relations between changes in Interpersonal Problems and in Depressive Symptoms, separately for CT and IPT.

		Cognitive Therapy	Interpersonal Psychotherapy
		B (SE)	B (SE)
Effects	b_1	0.09 (0.04)[*]	0.02 (0.06)
	b_2	0.20 (0.03)[***]	0.21 (0.06)[***]
	b_3	0.02 (0.03)	0.07 (0.04)
	b_4	0.02 (0.04)	-0.02 (0.06)[*]

CT = Cognitive Therapy; IPT = Interpersonal Psychotherapy; BDI-II = Beck Depression Inventory–II; IIP = Inventory for Interpersonal Problems; b_1 represents the concurrent relation between Δ in BDI-II and Δ in IIP from 0 to 3 months; b_2 represents the concurrent relation between Δ in BDI-II and Δ in IIP from 3 to 7 months; b_3 represents the temporal relation between absolute score on the IIP at 3 months and Δ in BDI-II from 3 to 7 months; b_4 represents the relation between Δ in IIP from 0 to 3 months and Δ in BDI-II from 3 to 7 months; * $p < .05$; *** $p < .001$.

Discussion

The current study examined change in five hypothesized working mechanisms of CT and IPT – two theoretically and practically different treatments for MDD – in relation to change in depressive symptoms over the course of treatment. Furthermore, it was examined whether the pattern in the data was consistent with theorized causal mediation models. Our study is the first to examine temporal relations between various potential (specific and non-specific) mediators and depression severity in a randomized comparison of CT and IPT for depression, hereby expanding on prior (cross-sectional) research on mechanisms of change.

In line with our expectations, patients' scores on all investigated process measures changed for the better over the course of treatment, with medium to large effect sizes. Largest changes were found in depressive symptoms and rumination. The quality of working alliance showed the least change. It should be noted that alliance scores were already very high to begin with, and remained stable over the course of therapy. No differential effects in pre- to post-treatment changes were observed between the two conditions. However, change in interpersonal functioning occurred more rapidly in IPT than in CT.

In each of the treatments, concurrent changes in process variables and depressive symptoms were strongly related. More specifically, improvement in depression se-

verity was associated with concurrent improvement in self-esteem and working alliance, and a decrease of dysfunctional attitudes, interpersonal problems, and rumination. Except for a significant relation between early change in self-esteem (0–3) and subsequent change in depression (3–7), no significant relations were found between early change in the potential process variables and subsequent change in depression, or vice versa. In addition, we found that change on the BDI-II was mediated by concurrent change on the IIP in both phases of treatment, and that the relationship between interpersonal functioning and depression appeared to be different for CT and IPT. However, we could not confirm the temporal relation between change in the mediator and change in outcome. No evidence for mediation was found on the other process measures.

No overall differences between CT and IPT

The fact that patients treated with CT and IPT exhibited similar improvement on all process measures – including cognitive beliefs and interpersonal functioning – is remarkable from a theoretical point of view. According to the respective theoretical models of change, each treatment exerts a beneficial effect through its own (specific) theorized mechanisms. Since IPT, in contrast to CT, does not actively target the modification of dysfunctional cognitions, and CT does not explicitly focus on improvement of interpersonal functioning, one would expect differences between the two treatments. However, our results are in line with findings of empirical studies demonstrating that patients' scores on theorized processes improve over the course of treatment, but that this is not necessarily specific for one type of treatment (e.g., Backenstrass et al., 2006; Bernecker et al., 2014; Quilty et al., 2008; Renner et al., 2012; Warmerdam et al., 2010).

Moreover, when taking a closer look at our findings conceptually, there are fair reasons to believe that CT might lead to changes in interpersonal functioning, and IPT might facilitate cognitive change, both through direct and indirect pathways. For example, as pointed out by e.g. Bernecker et al. (2014), even though beliefs are not systematically examined and challenged in IPT, cognitive distortions within an interpersonal context are addressed in IPT. Furthermore, because IPT gives the patient the sick role – a temporary status recognizing that depression is a medical illness that keeps the patient from functioning at full capacity – self-blaming cognitions, and feelings of hopelessness and worthlessness are reframed as part of the disorder, rather than being reality-based. Similarly, cognitive interventions and behavioural experiments in CT might have a specific focus on interpersonal situations, thereby directly facilitating change in interpersonal functioning. Indirectly, specific experiences following changes in interpersonal functioning in IPT, might lead to re-adjustment of schema's and attitudes about the self, world and future. Likewise, decrease of dysfunctional attitudes in CT might also lead to change in beliefs related to interpersonal relationships, making patients more likely to improve their interpersonal functioning as well.

6

Potential explanations for the limited evidence for the theoretical models of change

Although speculative, there are several potential explanations for why we only found limited evidence for the theoretical models of change. First of all, it might have been the case that the change in depressive symptoms that was observed over the course of treatment was a result of natural course of depression and therefore unrelated to the fact that patients were treated. However, as reported in our main outcome paper (Lemmens et al., 2015), both CT and IPT exceeded response of an untreated control condition after two months, making this explanation not very likely. Second, our results could indicate that theories are incorrect, and that other mechanisms – not assessed in this study – are responsible for therapeutic change. For example, the present study did not assess change in activity level, a promising powerful mechanism of CT.

Alternatively, it might be the case that the processes that were assessed in this study actually do play a role, but that we were unable to capture their mediational role – and more specifically their temporal effects – due to flaws in our design and (selection of) measurement instruments. For example, we might have conceptualized our processes wrong or used measures that were not sensitive enough (Johansson et al., 2010). Our results could also be affected by the fact that the outcome measure BDI-II shows some overlap with the various process measures[7]. Another possibility is that temporal relations do exist, but that our assessment points were spaced too far apart to detect the temporal sequence in which change took place. A more fine-grained analysis of change over time might have given a more detailed look at the shape of change. Moreover, analyses were performed at group-level. It might be the case that mediating effects are present in certain (subgroups of) depressed patients.

Furthermore, in order for change in a potential mechanism to account for (subsequent) change in depression severity, there needs to be substantial change over the period of observation. The fact that effect sizes for (individual) mediators were smaller than for the BDI-II already indicates that current mediators could only explain a limited amount of the treatment effect. Furthermore, since the average change in process measures in the early phase of treatment was rather small, there was only little change from which to predict. This might indicate that – even though we had a fairly large sample size – the treatment provided in our study was not powerful enough, or was delivered with insufficient specificity to mobilize mechanisms. When comparing the absolute change in BDI-II over the course of treatment of our study to that of Quilty et al. (2008) – who did claim a mediating role for dysfunctional attitudes – it can be concluded that BDI change in the Quilty trial was larger than ours (overall difference of 4.0 BDI-II points in CT and 4.5 in IPT). This might indicate that Quilty et al. provided a more powerful version of CT (and IPT), thereby triggering more specific change from one or both modalities. However, it should be noted that our results resemble those of other randomized comparisons of CT and IPT (Elkin, Shea, Watkins, & Imber, 1989; Luty et al., 2007).

Finally, the possibilities for establishing statistical mediation strongly depend on the extent to which the two interventions differ. Since we compared two treatments that did not show differences on outcome- and process measures, possibilities for examining statistical mediation were limited. It is possible that if we would have compared each of the treatments with a control group that showed larger differences (e.g., a non-psychological, or full waiting-list control), we might have had more options to demonstrate a mediating role of the abovementioned variables. We therefore argue against the requirement that a mediator needs to show more change in one treatment than in the other. In our view, the primary goal in process research should be to identify any factors that are a linking pin between treatment and outcome, regardless of whether they are specific to one treatment. By including specificity as a (testable) criterion in mediation analysis, important information about processes that facilitate symptom change is discarded as soon as it turns out that this change is not specific for a given treatment modality.

Methodological considerations and recommendations

The current study was novel in that it was the first study that allowed for a direct comparison of mediating effects between CT and IPT in a large sample of adult depressed outpatients. Furthermore, whereas most studies only examine mediation, we examined temporal mediation by evaluating both temporal relations and mediation in 1 model. This is in line with recommendations by e.g. Kazdin (2007, 2009). The temporal RCT design and the use of LDS models provided a unique opportunity to evaluate whether change in proposed process measures preceded, followed from, or went together with changes in depression. Furthermore, our study is one of the few trials in the field that also examined alternative causal accounts. So even though we were not able to demonstrate temporal mediation, our method could serve as an example for future studies.

As in all research, the current study also has a number of limitations. First, despite the fact that we found some evidence for statistical mediation, we were not able to differentiate between cause and effect – thereby limiting the value of our findings. Second, several LDS models had poor fit, and only limited data were available for the assessment of therapeutic alliance. Results should therefore be interpreted with caution. Third, because we only examined single mediation models, the order of significance and potential collaboration between theorized processes remains unknown.

Further research should directly focus on testing theories of change, since this can narrow down the number of potential mechanism and provide guidance for future research questions and improvement of research designs. In doing this, studies should include both theorized processes as well as potential mechanisms that are not directly consistent with theory in order to falsify theories. Furthermore, they should pay special attention to the timing of assessments and also focus on within-person variance or homogeneous subgroups of patients, rather than on a single average for the entire sample. In addition, studies are needed that focus more specifically on the differential effects

6

between CT and IPT in order to better understand the extent to which the processes are therapy specific. Additional research is necessary to further investigate the possible explanations mentioned above.

In conclusion

The present study contributes to the developing research field on mechanisms of change of psychotherapy by examining temporal relations between various potential (specific and non-specific) mediators and depression severity in a randomized comparison of CT and IPT for depression. In spite of a well-considered research design and the use of an innovative statistical approach, we found little empirical evidence for the theoretical models of change. While it would be premature to call for revision of theoretical models based on the current data, we do not rule out the possibility that theories might need to be revised in the future, especially with regard to the assumption that a mechanism needs to be specific for one type of treatment.

Footnotes

[1] One may argue that rumination is a cognitive process, and should therefore be included as a specific factor of CT. However, given the fact that Rumination is not specifically addressed in the CT protocol, and the instrument that was used in the current study assessed the overall level of rumination, we decided to include it here as a process common to both CT and IPT.

[2] As reported in our design paper (Lemmens et al., 2011), attributional style and cognitive reactivity were assessed as well. However, because changes in these processes have not been implicated as mediator of change, but rather as risk factors that are associated with the onset of depression and risk of relapse, findings using these measures will be presented in a separate paper on relapse prevention.

[3] This was done in order to mimic the result of a subtraction (e.g., DAS 3 months = 1*DAS baseline + 1*ΔDAS baseline and 3 months) (McArdle, 2009)

[4] Because further analyses included difference scores, we did not control for baseline differences.

[5] The careful reader will notice that the effect sizes on the BDI-II that are reported here are slightly different from those reported in our main outcome paper (d = 1.71 for CT and 1.73 for IPT; Lemmens et al., 2015). This difference can be explained by the fact that effect sizes in our main outcome paper were based on the covariate corrected multilevel estimates (baseline scores on BDI/EQ5D), whereas in the current paper observed values were used.

[6] χ^2 = 5.05 (p = .17), RMSEA = .07, CFI = 1.00 for DAS-A-NL; χ^2 = 11.69 (p = .01), RMSEA = .14, CFI = 0.98 for IIP; χ^2 = 9.43 (p = .02), RMSEA = .12, CFI = 0.98 for RRS; χ^2 = 18.13 (p = .00), RMSEA = .18, CFI = 0.96 for SLSC-R; χ^2 = 3.05 (p = .38), RMSEA = .01, CFI = 1.00 for WAI-O-S.

[7] The BDI-II includes items reflecting e.g., negative cognitions (DAS-A17), rumination (RRS) and self-esteem (SLSC-R).

6

References

Andrusyna, T. P., Luborsky, L., Pham, T., & Tang, T. Z. (2006). The mechanisms of sudden gains in Supportive-Expressive therapy for depression. *Psychotherapy Research, 16*(5), 526–535.

Backenstrass, M., Schwarz, T., Fiedler, P., Joest, K., Reck, C., Mundt, C., & Kronmueller, K.-T. (2006). Negative mood regulation expectancies, self-efficacy beliefs, and locus of control orientation: moderators or mediators of change in the treatment of depression? *Psychotherapy Research, 16*(2), 250–258.

Barber, J. P. (2009). Toward a working through of some core conflicts in psychotherapy research. *Psychotherapy Research, 19*(1), 1–12.

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

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.

Beck, A. T. (1964). Thinking and depression: Theory and therapy. *Archives of General Psychiatry, 10*, 561–571.

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

Beck, A. T., Steer, R., & Brown, G. K. (1996). *Beck Depression Inventory II: Manual*. Boston, MA: Hartcourt Brace.

Bernecker, S. L. (2012). How and for whom does interpersonal psychotherapy for depression work? *Psychotherapy Bulletin, 47*(2), 13–17.

Bernecker, S. L., Constantino, M. J., Pazzaglia, A. M., Ravitz, P., & McBride, C. (2014). Patient Interpersonal and Cognitive Changes and Their Relation to Outcome in Interpersonal Psychotherapy for Depression. *Journal of Clinical Psychology, 70*(6), 518–527.

Bordin, E. S. (1979). The generalizability of the psychanalytic concept of the working alliance. *Psychotherapy: Theory, Research, Practice, Training, 16*, 252–260.

Castonguay, L. G., Constantino, M. J., & Holtforth, M. G. (2006). The working alliance: Where are we and where should we go? *Psychotherapy: Theory, Research, Practice, Training, 43*(3), 271–279.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cuijpers, P., Geraedts, A. S., van Oppen, P., Andersson, G., Markowitz, J. C., & van Straten, A. (2011). Interpersonal Psychotherapy for Depression: A Meta-Analysis. *American Journal of Psychiatry, 168*, 581–592.

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology, 76*(6), 909–922.

DeRubeis, R., & Feeley, M. (1990). Determinants of change in cognitive therapy for depression. *Cognitive therapy and research, 14*(5), 469–482.

DeRubeis, R. J., Evans, M. D., Hollon, S. D., Garvey, M. J., Grove, W. M., & Tuason, V. B. (1990). How Does Cognitive Therapy Work? Cognitive Change and Symptom Change in Cogntive Therapy and Pharmacotherapy for Depression. *Journal of Consulting and Clinical Psychology, 58*(6), 862–869.

Does van der, W. (2002). *BDI-II-NL: Handleiding; De Nederlandse versie van de Beck Depressie Inventory-II.* Enschede: Ipskamp.

Elkin, I., Shea, M. T., Watkins, J. T. & Imber, S. D. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General Effectiveness of treatments. *Archives of General Psychiatry, 46*(11), 971–982.

Feeley, M., DeRubeis, R. J., & Gelfand, L. A. (1999). The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 67*(4), 578–582.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I).* New York, NY: Biometrics Research Department New York State Psychiatric Institute.

Garrat, G., Ingram, R. E., Rand, K. L., & Sawalani, G. (2007). Cognitive Processes in Cognitive Therapy: Evaluation of the Mechanisms of Change in the Treatment of Depression. *Clinical Psychology, 14*, 224–239.

Gaston, L., Marmar, C. R., Gallagher, D., & Thompson, L. W. (1991). Alliance prediction of outcome beyond in-treatment symptomatic change as psychotherapy processes. *Psychotherapy Research, 1*, 104–113.

Graaf de, L. E., Roelofs, J., & Huibers, M. J. H. (2009). Measuring dysfunctional attitudes in the general population: the DAS-A revised. *Cognitive therapy and research, 33*, 345–355.

Hollon, S. D., DeRubeis, R. J., & Evans, M. D. (1987). Causal Mediation of Change in Treatment for Depression: Discriminating Between Nonspecificity and Noncausality. *Psychological Bulletin, 102*(1), 139–149.

Hollon, S. D., & Ponniah, K. (2010). A review of empirically supported psychological therapies for mood disorders in adults. *Depression and Anxiety, 27*(10), 891–932.

Hollon, S. D., Thase, M. E., & Markowitz, J. C. (2002). Treatment and prevention of depression. *Psychological Science and Public Interest, 3*, 39–77.

Horowitz, L. M., Alden, L. E., Wiggins, J. S., & Pincus, A. L. (2000). *IIP–Inventory of Interpersonal Problems Manual.* San Antonio, TX: The Psychological Corporation.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G., & Villaseñor, V. S. (1988). Inventory of Interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*, 885–892.

Horvath, A. O. (1994). Empirical validation of Bordin's pantheoretical model of the alliance: the Working Alliance Inventory perspective. In A. O. Horvath & L. S. Greenberg (Eds.), *The Working Alliance: Theory, Research, and Practice* (pp. 109–128). New York, NY: Wiley.

Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48*(1), 9–16.
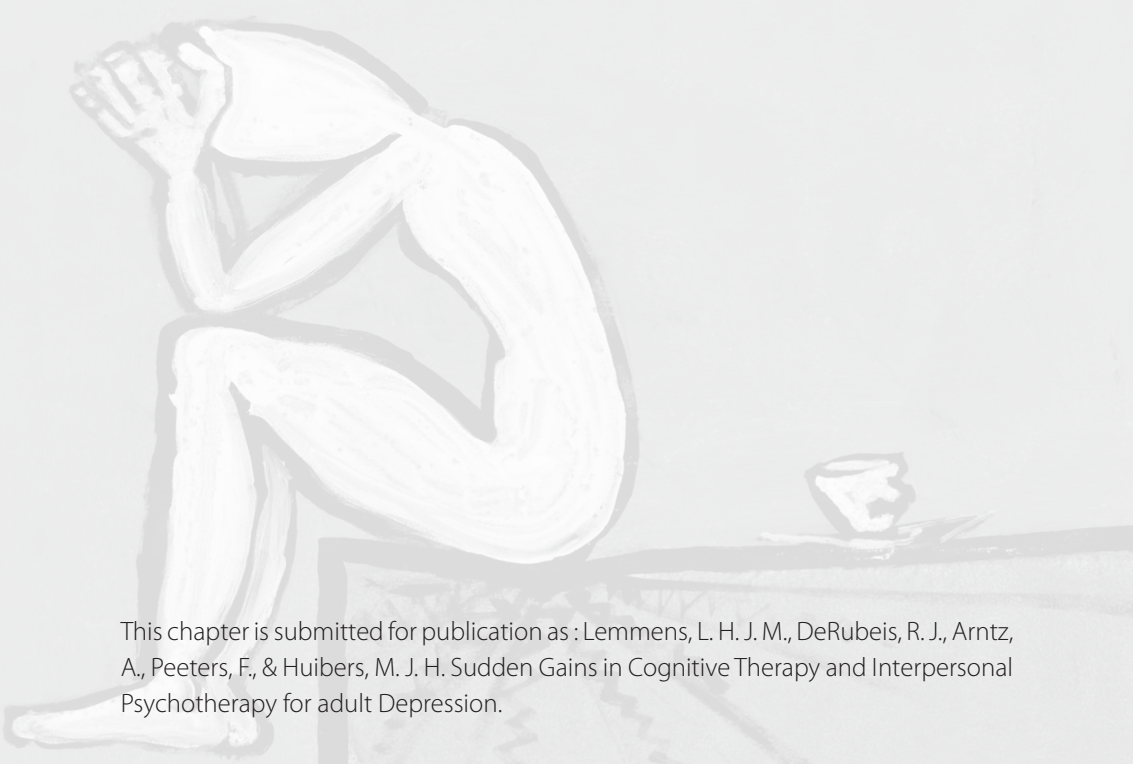
6

Johansson, P., Høglend, P., Ulberg, R., Amlo, S., Marble, A., Bøgwald, K.-P., . . . Heyerdahl, O. (2010). The mediating role of insight for long-term improvements in psychodynamic therapy. *Journal of Consulting and Clinical Psychology, 78*(3), 438–448.

Kazdin, A. E. (2007). Mediators and Mechanisms of Change in Psychotherapy Research. *Annual Review of Clinical Psychology, 3*, 1–27.

Kazdin, A. E. (2009). Understanding how and why psychotherapy leads to change. *Psychotherapy Research, 19*(4–5), 418–428.

Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, 44*(8), 1116–1129.

Klein, D. N., Schwartz, J. E., Santiago, N. J., Vivian, D., Vocisano, C., Castonguay, L. G., . . . Keller, M. B. (2003). Therapeutic Alliance in Depression Treatment: Controlling for Prior Change and Patient Characteristics. *Journal of Consulting and Clinical Psychology, 71*(6), 997–1006.

Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. S. (1984). *Interpersonal psychotherapy for depression*. New York, NY: Basis Books.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd ed.). New York, NY: Guilford.

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*, 877–883.

Kuyken, W., Watkins, E., Holden, E., White, K., Taylor, R. S., Byford, S., . . . Dalgleish, T. (2010). How does mindfulness-based cognitive therapy work? *Behaviour Research and Therapy, 48*, 1105–1112.

Lemmens, L. H. J. M., Arntz, A., Peeters, F. P. M. L., Hollon, S. D., Roefs, A., & Huibers, M. J. H. (2011). Effectiveness, relapse prevention and mechanisms of change of cognitive therapy vs. interpersonal therapy for depression: study protocol for a randomised controlled trial. *Trials, 12*(1), 150–162.

Lemmens, L. H. J. M., Arntz, A., Peeters, F. P. M. L., Hollon, S. D., Roefs, A., & Huibers, M. J. H. (2015). Clinical Effectiveness of Cognitive Therapy vs. Interpersonal Psychotherapy for Depression: Results of a Randomized Controlled Trial. *Psychological Medicine, 45*, 2095-2110.

Longmore, R. J., & Worrell, M. (2007). Do we need to challenge thoughts in cognitive behavior therapy? *Clinical Psychology Review, 27*(2), 173–187.

Lorenzo-Luaces, L., German, R. E., & DeRubeis, R. J. (2015). It's complicated: cognitive change procedures, cognitive change, and symptom change in CBT for depression. *Clinical Psychological Review*, E-pub ahead of print.

Luty, S. E., Carter, J. D., McKenzie, J. M., Rae, A. M., Frampton, C. M. A., Mulder, R. T., & Joyce, P. R. (2007). Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *British Journal of Psychiatry, 190,* 496–502.

MacKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis*. New York, NY: Taylor & Francis Group.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation Analysis. *Annual Review of Clinical Psychology, 58*, 593–614.

Markowitz, J. C., & Weissman, M. M. (2004). Interpersonal psychotherapy: principles and applications. *World Psychiatry, 3*, 136–139.

McArdle, J. J. (2009). Latent Variable Modeling of Differences and Changes with Longitudinal Data. *Annual Review of Psychology, 60*, 577–605.

Muthén, B. O., & Asparouhov, T. (2014). Causal Effects in Mediation Modeling: An Introduction with Applications to Latent Variables. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(1), 12–23.

Nolen-Hoeksema, S., & Morrow, J. (1991). A prospective study of depression and posttraumatic stress symptoms after a natural disaster: The 1989 Loma Prieta earthquake. *Journal of Personality and Social Psychology, 61*(1), 115–121.

Quilty, L. C., McBride, C., & Bagby, R. M. (2008). Evidence for the cognitive mediational model of cognitive behavioural therapy for depression. *Psychological Medicine, 38*, 1531–1541.

Raes, F., Hermans, D., & Eelen, P. (2003). Kort instrumenteel De Nederlandstalige versie van de Ruminative Response Scale (RRS-NL) en de Rumination on Sadness Scale (RSS-NL). *Gedragstherapie, 36*(2), 97–104.

Renner, F., Jarret, R. B., Vittengl, J. R., Barret, M. S., Clark, L. A., & Thase, M. E. (2012). Interpersonal problems as predictors of therapeutic alliance and symptom improvement in cognitive therapy for depression. *Journal of Affective Disorders, 138*(3), 458–467.

Strunk, D. R., Brotman, M. A., & DeRubeis, R. J. (2010). The process of change in cognitive therapy for depression: Predictors of early inter-session symptom gains. *Behaviour Research and Therapy, 48*(7), 599–606.

Tafarodi, R. W., & Swann, W. B. (2001). Two-dimensional self-esteem: Theory and measurement. *Personality and Individual Differences, 31*(5), 653–673.

Tichenor, V., & Hill, C. E. (1989). A comparison of six measures of working alliance. *Psychotherapy: Theory, Research, Practice, Training, 26*(2), 195–199.

Toth, S. L., Rogosch, F. A., Oshri, A., Gravener-Davis, J., Sturm, R., & Morgan-López, A. A. (2013). The efficacy of interpersonal psychotherapy for depression among economically disadvantaged mothers. *Development and Psychopathology, 25*, 1065–1078.

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*(3), 207–210.

Vandromme, H., Hermans, D., Spruyt, A., & Eelen, P. (2007). Dutch translation of the Self-Liking/Self-Competence Scale–Revised: A confirmatory factor analysis of the two-factor structure. *Personality and Individual Differences, 42*(1), 157–167.

Warmerdam, L., van Straten, A., Jongsma, J., Twisk, J., & Cuijpers, P. (2010). Online cognitive behavioral therapy and problem-solving therapy for depressive symptoms: Exploring mechanisms of change. *Journal of Behavior Therapy and Experimental Psychiatry, 41*, 64–70.

Webb, C. A., DeRubeis, R. J., Amsterdam, J. D., Shelton, R. C., Hollon, S. D., & Dimidjian, S. (2011). Two aspects of the therapeutic alliance: Differential relations with depressive symptom change. *Journal of Consulting and Clinical Psychology, 79*(3), 279–283.

Weissman, A. N., & Beck, A. T. (1978). Development and validation of the Dysfunctional Attitude Scale; paper presented at the annual meeting of the Association for the Advancement of Behavior Therapy.

Whisman, M. A. (1993). Mediators and moderators of change in cognitive therapy of depression. *Psychological Bulletin, 114*, 248–265.

6

7

Sudden Gains in Cognitive Therapy and
Interpersonal Psychotherapy for Adult Depression

Abstract

We examined the rates, baseline predictors and clinical impact of sudden gains in a randomized comparison of individual Cognitive Therapy (CT) and Interpersonal Psychotherapy (IPT) for adult depression. Patients were 117 depressed outpatients (mean age = 41.8; 66.7% female) who received 16–20 sessions of either CT or IPT. Session-by-session symptom severity was assessed using the Beck Depression Inventory-II. Our primary analyses examined sudden gains using the original criteria as defined by Tang and DeRubeis (1999). In a series of secondary analyses, we examined whether the duration of the between-session interval at which sudden gains were recorded affected the results obtained. There were significantly more patients with sudden gains in CT (42.2%) as compared to IPT (24.5%). There were no differences with regard to the magnitude and timing of the sudden gains. In both treatments, sudden gains were predicted by baseline quality of life score and absence of axis-I comorbidity. Furthermore, those with sudden gains reported lower levels of depression severity at post-treatment and 5 months FU. The duration of the between-session interval did not influence the results. The current study indicates differences in occurrence of sudden gains in two treatment modalities that overall showed similar results. This might reflect different mechanisms of change. Future research should examine differences between CT and IPT more closely and further examine precipitants and predictors of sudden gains.

7

Introduction

For a long time in investigations of the course of change during psychotherapy it had been assumed that the typical trajectory of change was gradual and linear (Kazdin & Nock, 2003; Kraemer, Wilson, Fairburn, & Agras, 2002; Laurenceau, Hayes, & Feldman, 2007). However, studies of the temporal course of change during treatment have revealed that a substantial portion of a patient's total reduction in symptom intensity often occurs suddenly, rather than gradually over the course of treatment. Analyses of individual change patterns in two large trials of Cognitive Therapy (CT) for depression (Elkin, Shea, Watkins, & Imber, 1989; Hollon et al., 1992) showed that almost 50% of the patients experienced half of their total symptom improvement between two consecutive therapy sessions (Tang & DeRubeis, 1999a). These large improvements in symptoms during a single between-session interval were termed "sudden gains" (Tang & DeRubeis, 1999b). In the initial study that described this phenomenon, Tang and DeRubeis found that patients who experienced sudden gains showed more therapy improvement and had higher rates of recovery compared to those without sudden gains, and they remained so 18 months later. Further research showed that patients with sudden gains were up to 75% less likely to experience relapse and recurrence in the 24 months after treatment termination (Tang, DeRubeis, Hollon, Amsterdam, & Shelton, 2007). The finding that sudden gains during treatment are associated with more favourable treatment outcomes, both in the short run and in the long term, has important implications for the prognosis of depression, given that for many patients it can follow a chronic or recurrent course.

Since its introduction in 1999, sudden gains have been examined in at least 10 studies, covering at least 7 forms of psychotherapy for depression (see meta-analysis of Aderka, Nickerson, Bøe, & Hofmann, 2011 for an overview). In the six studies that focused on individual CT for adult depression, sudden gain rates ranged from 33% to 45%. In each study, those with sudden gains showed lower levels of depressive symptomatology at post-treatment and/or follow-up as compared with those who did not experience sudden gains (Busch, Kanter, Landes, & Kohlenberg, 2006; Hardy et al., 2005; Tang & DeRubeis, 1999b; Tang, DeRubeis, Beberman, & Pham, 2005; Tang et al., 2007; Vittengl, Clark, & Jarrett, 2005). It should be noted, however, that the criteria for identifying sudden gains differed across the studies, making between-study comparisons of the rates problematic.

Interpersonal Psychotherapy (IPT) is, next to CT, the most commonly practiced and extensively researched psychotherapy for depression (Cuijpers et al., 2011; Hollon, Thase, & Markowitz, 2002). Kelly, Cyranowski, and Frank (2007) examined the occurrence and impact of sudden gains during IPT in female patients with recurrent depression. They observed sudden gains at a similar rate as has been reported in CT (34%), but found no evidence that sudden gains were associated with better outcomes at post-treatment or during follow-up. This is, to our knowledge, the only study of sudden gains during IPT for depression.

Furthermore, patient pre-treatment characteristics that distinguish those who will go on to have sudden gains from those who will not, have only been explored

7

in a few studies in the context of CT and Behavioural Activation. In these studies, variation in depressive symptom severity, and measures of negative cognitions and interpersonal functioning at baseline have not been found to predict the occurrence of sudden gains (Hunnicutt-Ferguson, Hoxha, & Gollan, 2012; Kelly, Roberts, & Ciesla, 2005; Vittengl et al., 2005). The pre-treatment characteristics that might predict sudden gains in IPT however, have yet to be examined, as has the possibility that different characteristics would be associated with sudden gains in CT versus IPT.

The aim of the present study was therefore to replicate and extend previous research on sudden gains in individual psychotherapy for adult depression. Data come from a recently published randomized comparison of CT and IPT for depression in which we found no significant differences on outcome in the acute phase (Lemmens et al., 2011; 2015). We identified the patients who met criteria for sudden gains and examined the relation between sudden gain status and end of treatment symptom severity, as well as symptom severity at five-months follow-up. In addition, we explored baseline patient characteristics that might predict the occurrence of sudden gains, both within each treatment as well as across the two treatments. In a series of secondary analyses, we examined whether the duration of the between-session interval at which sudden gains were recorded affected the results obtained.

We expected that sudden gains would appear in both CT and IPT at a similar rate (approximately 40%) as in other studies of treatment for depression, (Busch et al., 2006; Hardy et al., 2005; Kelly et al., 2007; Tang & DeRubeis, 1999b; Tang et al., 2005; Tang et al., 2007; Vittengl et al., 2005). Also in line with previous research, we expected that individuals who experienced sudden gains would experience superior outcome, both at treatment termination and through the follow-up period (Tang & DeRubeis, 1999b; Tang et al., 2007). With regard to the pre-treatment factors associated with sudden gains, we expected that lower levels of depressive symptomatology and dysfunctional processes would be predictive of sudden gains.

Methods

Data source

The study sample consisted of 151 adult outpatients, aged 18–65, who participated in an RCT examining the effectiveness and mechanisms of change of individual CT ($n = 76$) and IPT ($n = 75$) for depression (Lemmens et al., 2011; 2015)[1]. Participants' primary diagnosis was Major Depressive Disorder, as determined by the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I; First, Spitzer, Gibbon, & Williams, 1997). Potential participants were excluded if they were currently taking antidepressant medication, if the duration of their current episode of depression exceeded 5 years, and/or if they met criteria for Bipolar Disorder, Substance Abuse or Dependence, or Mental Retardation

(IQ < 80). All participants provided written informed consent before entering the study. Treatment consisted of individual sessions, each lasting 45 minutes. Depending on the patient's progress, 16 to 20 sessions were provided. The CT protocol followed Beck, Rush, Shaw, and Emery (1979). The IPT protocol was based on Klerman, Weissman, Rounsaville, and Chevron's (1984) manual. The primary outcome measure, the Beck Depression Inventory-II (BDI-II: Beck, Steer, & Brown, 1996), was completed by patients immediately prior to each therapy session, at the end of the treatment phase (7 months), and monthly throughout the 5 month follow-up period (8–12 months). Further details concerning the design of the study, measures, interventions, participants flow and, main outcomes can be found in Lemmens et al. (2011; 2015). The Medical Ethics Committee of Maastricht University approved the study protocol. The study is registered at the Dutch Cochrane Centre through the Netherlands Trial Registry (ISRCTN67561918).

Data preparation

Similar to Tang and DeRubeis (1999b), we included for the purpose of examining sudden gains only those patients who received 8 or more sessions of therapy and whose BDI-II score at the first session was available and was 15 or higher. Of the 76 CT patients, 12 were excluded (6 received fewer than 8 sessions, 2 had BDI-II scores that were missing, and 5 had BDI-II scores that were lower than 15 at the first session). The numbers for the IPT condition were 9, 4, and 10, respectively. As some patients were excluded for more than one reason, this resulted in sample sizes of 64 in the CT condition and 53 in the IPT condition.

Patients in this sample ranged in age from 18 to 63 (M = 41.8, SD = 12.1). Two-thirds (66.7%) were female, 59.8% were educated at a medium level (low: 18.8%, high: 21.4%), and 60.7% were married or in a committed relationship. The average score on the BDI-II at baseline was 30.6 (SD = 8.5), and 47% of the sample was diagnosed with recurrent depression. More than half of the patients (58.1%) suffered from "severe depression" (BDI-II score ≥ 29; Beck et al., 1996) but the majority (59.0%) was still actively employed. The 34 patients who were excluded from analyses did not differ significantly from those who were included on any of the baseline variables described above. As expected, significant differences were found on variables that were used to exclude patients from the study: mean number of sessions attended (10.0 (SD = 7.0) vs. 15.9 (SD = 3.7)) and mean first session BDI-II score (20.6 (SD = 12.4) vs. 29.2 (SD = 8.0)).

7

Statistical analyses

A total of 1667 between-session intervals were analysed for identification of sudden gains using the original criteria of Tang et al. (1999b; 2005). The sudden gain must be (*a*) large in absolute magnitude (at least 7 BDI-II points), (*b*) large in relative magnitude (representing at least 25% of the pre-gain session's BDI-II score) and (*c*) large relative to symptom fluctuation (significantly higher mean in the three pre-gain sessions than

in the three post-gain sessions in order to rule out random fluctuation in treatment)[2]. As in previous studies, reversals were defined as having occurred whenever a patient gave up 50% of the symptom improvement that constituted the sudden gain (Tang & DeRubeis, 1999b). We mapped out rates and characteristics of sudden gains and reversals, and compared baseline sociodemographic characteristics and pre-treatment clinical measures of patients with and without sudden gains. Furthermore, differences between the two interventions were examined.

In order to explore the predictive value of baseline characteristics on the occurrence of sudden gains during treatment, using univariate logistic regressions we identified baseline variables with promising associations with the dependent variable ($p < 0.20$). The variable measuring the presence or absence of one or more sudden gains (*gain*) was the dependent variable. As potential predictors we selected several demographic variables (gender, age, marital status, level of education, employment status), depression specifiers (severity, recurrence), and measures of general functioning and psychological distress. Measures of cognitive processes (assessing dysfunctional attitudes, cognitive reactivity, rumination, and attributional style) were also included, as were measures of interpersonal functioning and self-esteem. A complete listing of these measures can be found in Data Supplement I. Full descriptions of the measures are reported elsewhere (Lemmens et al., 2011). To examine the effects of these potential predictors on outcome, the selected variables were entered in a multivariate regression model and were eliminated in a manual backward procedure until the model consisted only of significant predictors ($p < 0.05$).

To determine whether patients with sudden gains benefitted more from treatment as compared to patients without sudden gains, we conducted a series of analyses of covariance (ANCOVAs), with sudden gain (present vs. absent) as the independent variable and BDI-II scores (separate analyses at 7 months and at 12 months) as the dependent variables. Pre-treatment severity (BDI-II baseline) was included as a covariate in all analyses. In an additional model at 12 months we controlled for post-treatment BDI-II score as well. In order to examine the differential effects of sudden gains on outcome for CT vs. IPT, we added condition and the interaction of gain*condition to the models. To ensure that patients who experienced sudden gains were not simply treatment responders, whereas patients without sudden gains did not respond to treatment, we selected only those patients whose BDI-II decreased by at least 9 BDI-II points over the course of treatment and reran all models in this subsample.

We note that the original sudden gain criteria do not include a specification of the maximum duration of the between-session interval and therefore do not ensure the suddenness of the sudden gain. This was not a problem in the original studies (Tang et al., 1999; 2002; 2005), since therapy sessions took place once or twice per week. As a result, the between-session intervals in those studies were less than or equal to 7 days. However, in the Dutch healthcare system and therefore in our trial, sessions are offered weekly, with the possibility to lower frequency toward the end of treatment. This can result in longer

between-session intervals such that in some instances, decreases in symptoms noted between two consecutive sessions should not be considered 'sudden'. We therefore performed a series of secondary analyses in which we examined the effects of the duration of the between-session interval. More specifically, we excluded all between-session intervals longer than 14 days (the most common alternative to weekly sessions in the Netherlands) and re-examined frequencies, characteristics and baseline predictors of sudden gains in both conditions. Furthermore, to examine whether sudden gains that occurred within or outside this 14-day interval were differentially associated with outcome, we repeated the analyses of covariance, but with *gain* as three-level variable (no sudden gain, sudden gain within 14 day interval, sudden gain outside interval), and inspected the contrast of sudden gains within vs. sudden gains outside the 14 day interval.

Results

Frequency and characteristics of Sudden Gains

A total of 27 patients (42.2%) in CT and 13 (24.5%) in IPT[3] met the full set of sudden gains criteria (see Table 1). There were no significant differences on baseline sociodemographic characteristics and pre-treatment clinical measures between those who were and were not identified as having experienced a sudden gain (see Table 2).

Table 1. Frequency of Sudden Gains and Reversal rates stratified according to treatment condition ($n = 117$).

		Large drop (a & b)	Sudden Gains (a, b, & c)		Reversals
		n patients (%)	*n* gains	*n* patients (%)	*n* patients (%)
CT	($n = 64$)	43 (67.2)	35	27 (42.2)	6 (22.2)
IPT	($n = 53$)	40 (75.5)	17	13 (24.5)	2 (15.4)

Note: CT = Cognitive Therapy; IPT = Interpersonal Psychotherapy; Criterion *a* = large drop in absolute magnitude (≥ 7 BDI-II points); Criterion *b* = large drop in relative magnitude (≥ 25% of the pre-gain session's BDI-II score); Criterion *c* = drop is large relative to symptom fluctuation (BDI-II in three pre-gain sessions > BDI-II in three post-gain sessions); Reversal = giving up 50% of the symptom improvement that constituted the gain.

7

The difference in the proportion of sudden gainers in CT versus IPT was significant (χ^2 (1) = 4.0, $p = .045$). There were no differences between the two conditions with regard to the magnitude and timing of the sudden gains. The majority of sudden gain patients (70%) had only one sudden gain (19 in CT; 9 in IPT); the others experienced two sudden

gains[4]. The average magnitude of the gain for the total sample was 11.0 BDI-II points (SD = 4.3) and the median pre-gain session was 9 (see Table 3 for results stratified per condition). Of the 40 patients who experienced sudden gains, 8 (20.0%) experienced a reversal before the end of therapy (6 in CT, 2 in IPT; see Table 1). Thus, the percentage of patients who experienced one or two sudden gains and did not experience a reversal of their sudden gain(s) was 32.8% in CT and 20.8% in IPT (see Table 3).

Table 2. Sociodemographic characteristics and pre-treatment clinical measures for patients with and without Sudden Gains.

| | Gainers | Non-gainers | | |
	(*n* = 40)	(*n* = 77)	Test Statistic (d.f.)	*p*
Sociodemographic variables				
- Female sex, *n* (%)	24 (60.0)	54 (70.1)	$\chi^2 = 1.22$ (1)	.27
- Age in years, mean (SD)	40.0 (11.7)	42.8 (12.3)	$t = 1.20$ (115)	.23
- Education, *n* (%)				
–Low	6 (15.0)	16 (20.8)		
–Medium	25 (62.5)	45 (58.4)	$\chi^2 = .58$ (2)	.75
–High	9 (22.5)	16 (20.8)		
- Partner, yes, *n* (%)	25 (62.5)	46 (59.7)	$\chi^2 = .08$ (1)	.77
- Active Employment*, yes, *n* (%)	26 (65.0)	43 (55.8)	$\chi^2 = 1.26$ (1)	.26
Pre-treatment clinical measures				
- BDI-II score, mean (SD)	31.0 (10.2)	30.4 (7.5)	$t = -.35$ (62)	.73
- Recurrent depression, *n* (%)	21 (52.5)	34 (44.2)	$\chi^2 = .74$ (1)	.39
- Severe depression**, *n* (%)	24 (60.0)	44 (57.1)	$\chi^2 = .09$ (1)	.77

Note: BDI-II = Beck Depression Inventory Second Edition; * = data unavailable for CT (*n* = 1); ** = BDI-II score ≥ 29.

Table 3. Comparison of results between relevant studies examining Sudden Gains in individual CT and IPT for adult unipolar depression.

	Cognitive Therapy						Interpersonal Psychotherapy	
	Tang & DeRubeis	Tang et al.,	Hardy et al.,	Busch et al.,	Tang et al.,	Current	Kelly et al.,	Current
	1999	2005	2005	2006	2007	Study*	2007	Study*
- Sessions, n	20	12–20	8–20	20	20	16–20	12	16–20
- Sample Size, n	61	46	76	38	60	64	185	53
- Patients with SG's, %	39.3%	43.0%	41.0%	42.1%	40.0%	42.2%	33.5%	24.5%
- Magnitude, BDI-II, M (SD)	11.2 (4.4)	10.2 (2.8)	13.3 (5.0)	9.3 (2.9)	11(-)	10.3 (3.7)	13.0 (5.3)	12.4 (5.3)
- Pre-gain session, m	5	8	5	10	5	9	-	9
- Patients with reversals, %	16.7%	40.0%	32.3%	43.8%	37.5%	22.2%	53.2%	15.4%
- Patients that maintained SG's, %	32.8%	26.1%	27.6%	23.7%	25.0%	32.8%	15.7%	20.8%

Note: CT = Cognitive Therapy; IPT = Interpersonal Psychotherapy; SG = Sudden Gain; BDI-II = Beck Depression Inventory-II; *=original sudden gain criteria defined by Tang and DeRubeis (1999b).

7

A closer look at the differences between CT and IPT

Given that the number of CT patients with sudden gains was twice what was observed in IPT, we went on to compare the rates of the individual sudden gain criteria between the two conditions. We began by comparing the number of patients who did not meet the criteria for eligibility for sudden gain analyses in this study (BDI-II at first session < 15, or n sessions < 8). Significantly more IPT patients were excluded on these grounds (22/75 in IPT vs. 12/76 in CT; χ^2 (1) = 3.97, p = .046). Thus, only 53 IPT patients, compared to 64 CT patients, were included in the sudden gain analyses. Nonetheless, the absolute numbers of patients who met sudden gain criteria a and b, which index the magnitude of symptom improvement, were nearly identical in the two conditions (43 in CT vs. 40 in IPT; see Table 1). Differences between the conditions emerged only when criterion c, which indexes the stability of the drop, was applied. In CT, 62.8% of patients who met criteria a and b also met criterion c, whereas only 32.5% of the IPT patients did so (see Table 1). These results indicate that BDI-II scores around large reductions in symptoms tended to fluctuate more in IPT, thereby resulting in fewer patients who met the full set of criteria for sudden gains.

Baseline Predictors of Sudden Gains

Univariate logistic regressions identified 4 variables with a promising association with the dependent variable (see data supplement I). The selected variables were entered into a multivariate regression model (see Table 4 for full model), and non-significant variables were eliminated in a manual backward procedure until the model contained only the two significant predictors of sudden gains. As shown in table 4, patients were more likely to experience a sudden gain if they entered treatment with a higher Quality of Life Utility score or if comorbidity on Axis I was absent. Tests of the interactions of these two variables with treatment condition were non-significant.

Table 4. Full and Final model of the Multivariate logistic regression analysis (n = 117).

| Full Model	OR	95% CI	p
- Quality of Life Utility score (EuroQol)[1]	1.20	[0.98 –1.47]	.08
- Absence of Comorbid Axis I disorder	2.88	[1.22 –6.76]	.02
- Level of Impairment: Physical Functioning	1.01	[0.99 –1.03]	.37
- Hopelessness: feelings about the future	1.33	[0.98 –1.78]	.06
Final Model*	**OR**	**95% CI**	**p**
- Quality of Life Utility score (EuroQol)[1]	1.24	[1.04 –1.49]	.02
- Absence of Comorbid Axis I Disorder	2.64	[1.15 –6.08]	.02

Note: Dependent Variable = Gainer (yes/no); [1] = Adjusted score (EuroQol Utility * 10); * = Order of elimination: 1) Level of impairment: Physical functioning; 2) Hopelessness: feelings about the future.

Sudden Gains and Outcome

Patients with sudden gains showed more improvement during treatment as compared to those without sudden gains, and consequently reported lower post-treatment BDI-II scores (10.9 vs. 18.4; $F(1) = 13.12$, $p < .01$). Furthermore, at the end of the treatment-phase, 44.4% of the sudden gain patients met criteria for remission (BDI-II < 9) versus 25.0% of those without sudden gains ($\chi^2(1) = 4.21$, $p = .040$). A similar pattern emerged at 5 months FU (mean BDI-II: 10.0 vs. 18.4; $F(1) = 11.6$, $p < .01$; remission rates of 54.3% vs. 33.8%; $\chi^2(1) = 4.00$, $p = .045$). A detailed graphical overview of the course of depression for patients with and without sudden gains after treatment termination, stratified by treatment condition, is displayed in Figure 1. As can be seen in Figure 1, BDI-II scores of patients with sudden gains are fairly similar in CT and IPT, but it appears that they differ between conditions among patients who did not experience a sudden gain. However, tests of the gain*condition interaction were non-significant at both post-treatment ($F(1) = 0.51$, $p = .476$) and at 5 months FU ($F(1) = 1.30$, $p = .256$).
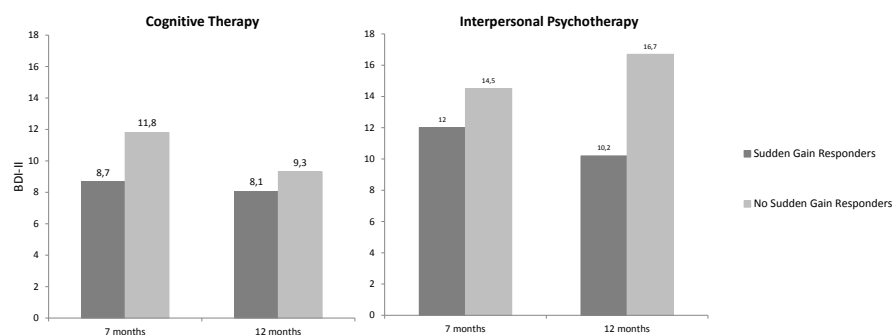
Sudden Gains and Response to treatment

At the end of the treatment phase, 81 of the 117 patients in CT or IPT met the criterion for "response" (a decrease of at least 9 BDI-II points over the course of treatment). Response rates were similar for CT and IPT (70.3% vs. 67.9%; $\chi^2(1) = 0.08$, $p = .781$). Among the 81 responders, 32 patients (39.5%) experienced a sudden gain. Overall, sudden-gain responders reported lower BDI-II scores at post-treatment and 5 months FU as compared to no-
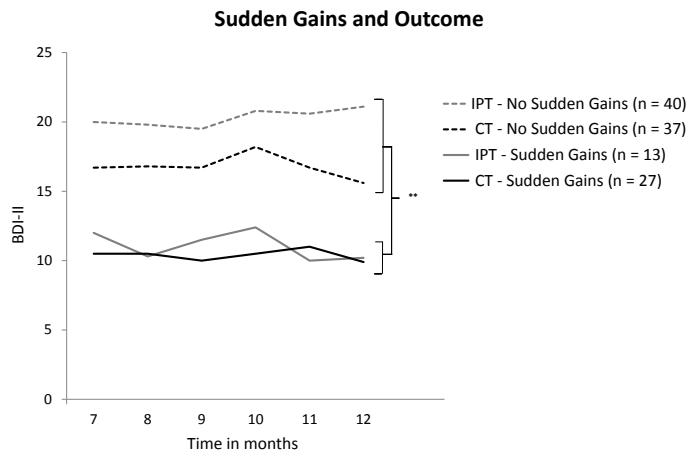
Figure 1. Course of Depression (observed BDI-II scores) for patients with and without Sudden Gains from treatment termination to five months FU, stratified according to treatment condition ($n = 117$). *Note*: ** = significant at $p < .01$; Data unavailable for $n = 9, 14, 13, 13, 14, 14$ patients at 7, 8, 9, 10, 11, 12 months respectively.

sudden-gain-responders (9.8 vs. 13.2 at post-treatment; 8.7 vs. 13.4 at 5 months FU). The difference was significant at post-treatment (F (1) = 6.53, p = .013) but not at the 5 month FU (F (1) = 2.01, p = .160). Figure 2 gives the BDI-II scores in CT and IPT at post-treatment and 5 months FU for sudden-gain-responders and non-sudden-gain-responders. Although it appears that the difference between the two patient groups is especially large in IPT at the 5 month FU, tests of the gain*condition interactions were non-significant at both time points.

Figure 2. Outcomes of Sudden-Gain Responders vs. No-Sudden Gain responders as measured by the Beck Depression Inventory-II (BDI-II) at post-treatment and five months FU, stratified according to treatment condition (n = 81; all differences are non-significant).

Effects of the duration of the between-session interval

Excluding sudden gains that occurred outside the acceptable between-session interval (>14 days) reduced the number of patients that were identified as sudden gainers from 27 to 19 (29.7%) in CT and 13 to 8 (15.1%) in IPT. The difference in rates between the two conditions became a non-significant trend (χ^2 (1) = 3.5, p = .062). Characteristics (e.g., magnitude, timing and reversals) and baseline predictors of these sudden gains were similar to those observed when we allowed for longer intervals between the gains (see data supplement II). Furthermore, contrasting patients with sudden gains that occurred within the 14-day interval (n = 27) to sudden gainers with gains that were recorded on a longer between-session interval (n = 13), showed no significant differences on outcome, both at post-treatment (adjusted mean difference = 0.72, SE = 3.76, p = .848) and 5 months FU (adjusted mean difference = -1.89, SE = 4.52, p = .676; also see data supplement II).

Discussion

The present study was the first to examine the occurrence, baseline predictors and clinical impact of sudden gains in a randomized comparison of individual CT and IPT for adult depression. Among those who met inclusion criteria for our analyses, a higher proportion of patients in CT experienced sudden gains, as compared to patients allocated to IPT. In both conditions, sudden gains were predicted by baseline quality of life score and the absence of axis-I comorbidity, suggesting that relatively "good health" at baseline predicts sudden gains. Furthermore, in line with what is reported in the literature (Aderka et al., 2011), patients with sudden gains were less depressed at post-treatment than were those without sudden gains, and they remained so 5 months after treatment termination. Considering only those patients whose sudden gains occurred within a between-session interval of only one or two weeks, which reduced the number of patients identified as sudden gainers, did not affect the CT/IPT ratio, the characteristics and predictors of sudden gains, or the advantage observed in outcome for those with sudden gains.

The frequency of sudden gains observed in the CT condition is within the range reported in other relevant studies in the field (see Table 3 for an overview). Sudden gain rates observed in the IPT condition would appear to be low when compared with those obtained in a previous study of IPT by Kelly et al. (2007). However, high reversal rates were reported in the Kelly et al. study, whereas the reversal rates in the present study indicate that the vast majority of sudden gains were maintained throughout treatment. When taking this into account and thus focusing only on those patients who maintained their sudden gains throughout therapy, the proportion of sustained sudden gains in the present study was no longer lower than Kelly et al.'s (see Table 3).

The fact that none of the pre-treatment measures that address CT- or IPT-relevant processes predicted the occurrence of sudden gains is in line with findings of Hunnicutt-Ferguson et al. (2012) and Kelly et al. (2005), who also did not find a relation between pre-treatment variables and the occurrence of sudden gains. These findings lend further support to the idea that the causes of sudden gains are not likely to be found with measures obtained long before the gain occurs, but rather with assessments of processes that occur over the course of therapy, in close temporal proximity to the gain.

Cognitive Therapy vs. Interpersonal Therapy

Although our results are in line with those obtained in other studies, the question that remains unanswered is why sudden gains are less likely to occur in IPT. One explanation could be that these large and sudden symptom improvements are indeed mainly driven by cognitive processes (as suggested by Tang and DeRubeis, 1999b). Although non-cognitive interventions may also lead to cognitive changes (e.g., Weissman, Markowitz, & Klerman, 2000), the emphasis on cognitive processes in CT may lead to more sudden gains in comparison to therapeutic approaches such as IPT that rely more on

7

non-cognitive interventions. However, even though we see merit in this explanation, we think it is unlikely that this fully explains our findings. Results indicate that the difference between the two treatment modalities emerged not because CT facilitated a higher number of large symptom improvements between two consecutive sessions, relative to IPT, but rather was due to the relative lack of stability in the pre- and post-gain sessions in IPT. This points to possible differences in the therapeutic process in CT and IPT. Because IPT focuses on solving difficulties in an interpersonal context, it may be more common in IPT for progress to be affected by the vagaries of patient's interactions with others. For example, to solve a problem or reach a specific goal in IPT the patient will likely engage with significant others. Whether these interpersonal encounters will lead to improvement in depressive symptoms depends on experiences during and the outcome of these encounters. As a result, levels of depressive symptoms may be more unstable during IPT than during CT. Second, CT includes homework between sessions, whereas IPT does not. This might promote the consolidation of the skills that are learned in treatment and therefore to more stable symptom improvement. A third possibility is that differences in therapists' attitude affect the stability of session-by-session BDI-II scores. The CT therapist is directive as he or she leads the patients through the protocol, whereas the IPT therapist is supportive as he or she helps the patient explore options in an interpersonal context, thus allowing for more instances of trial and error. These possibilities should be explored in future research on the effects of CT and IPT for depression.

Methodological considerations and recommendations

The current study was novel in that it allowed for a direct comparison between CT and IPT in the occurrence and nature of sudden gains in a large sample of adult depressed outpatients. Furthermore, it is the second study of sudden gains in IPT. By using a more heterogeneous group than Kelly et al. (2007) (males as well as females with first-episode and recurrent depressions) we increased the generalizability of previous findings. In addition, the inclusion of a separate responders analysis diminished uncertainty as to whether patients who experienced sudden gains were nothing more than treatment responders. Moreover, by explicitly examining the impact of the duration of the between-session interval, we paid attention to an important aspect of sudden gains that has been largely overlooked in the literature. To our knowledge there is only one other study that explicitly ensured the suddenness of the gain by excluding those observed over intervals that were considered too long (Kelly et al., 2007).

One limitation of the current study is that the percentage of patients who met criteria for sudden gains may have been too low, especially in IPT, to support powerful tests of potential differences between the two treatments in predictors and the relation of sudden gains with short and long-term measures of outcome. A second limitation

is that although the findings of pre-treatment variables that predict sudden gains add to an understanding of this phenomenon, they do not speak to the processes that occur over the course of therapy and elicit sudden gain

Future studies should therefore examine the differences between CT and IPT more closely, both in the short-term and in the long run, paying special attention to the impact of the duration of the between-session interval at which sudden gains are recorded. Furthermore, studies are needed that focus on therapy-specific as well as common factors that precede sudden gains by using observer-rated measures to capture relevant phenomena in the pre-gain session. A better understanding of sudden gains and their associated factors may provide insight into the mechanisms of change in psychotherapy and, further, may suggest areas for increased therapeutic focus and thereby contribute to the improvement of treatments for adults suffering from depression.

Acknowledgements

Conflict of Interest

The authors declare that they have no competing interests.

Funding

Footnotes

7

[1] As outlined in Lemmens et al. (2011), the full design also includes a third arm; a 2-month waiting-list control (WLC) condition followed by treatment of choice ($n = 31$). However, for the purpose of the present paper, we focused our attention only on sudden gains occurring in the two active conditions CT and IPT.

[2] The c criterion requires at least two BDI-II data points before and after the symptom changes. This automatically meant that large symptom changes that occurred very early or very late in treatment (within two sessions of initiating or terminating treatment) were excluded from further analysis. However, as outlined by Tang and DeRubeis (1999b) the first session is very different in nature from other sessions and therefore first-session sudden gains might be different phenomena than other sudden gains and should

be treated as such (see Tang & DeRubeis, 1999 for a detailed description of differences). Aderka et al. (2011) found further support for this notion by demonstrating that including these first-session sudden gains did not affect effect sizes of sudden gains. This line of reasoning can also be applied for the final treatment session.

[3] Following Kelly et al. (2007) we also examined the occurrence of sudden gains in IPT in recurrently depressed women only and found that 22.2% of this subsample met the full set of criteria for sudden gains.

[4] Of the 12 patients with two sudden gains, 4 (3 in CT and 1 in IPT) experienced a reversal between the two sudden gains.

References

Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2011). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80*(1), 93–101.

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

Beck, A. T., Steer, R., & Brown, G. K. (1996). *Beck Depression Inventory II: Manual*. Boston, MA: Hartcourt Brace.

Busch, A. M., Kanter, J. W., Landes, S. J., & Kohlenberg, R. J. (2006). Sudden Gains and Outcome: A Broader Temporal Analysis of Cognitive Therapy for Depression. *Behavior Therapy, 37*(1), 61–68.

Cuijpers, P., Geraedts, A. S., van Oppen, P., Andersson, G., Markowitz, J. C., & van Straten, A. (2011). Interpersonal Psychotherapy for Depression: A Meta-Analysis. *American Journal of Psychiatry, 168*, 581–592.

Elkin, I., Shea, M. T., Watkins, J. T., & Imber, S. D. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry, 46*(11), 971–982.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I)*. New York, NY: Biometrics Research Department New York State Psychiatric Institute.

Hardy, G. E., Cahill, J., Stiles, W. B., Ispan, C., Macaskill, N., & Barkham, M. (2005). Sudden Gains in Cognitive Therapy for Depression: A Replication and Extension. *Journal of Consulting and Clinical Psychology, 73*(1), 59–67.

Hollon, S. D., DeRubeis, R. J., Evans, M. D., Wiemer, M. J., Garvey, M. J., Grove, W. M., & Tuason, V. B. (1992). Cognitive therapy and pharmacotherapy for depression: Singly and in combination. *Archives of General Psychiatry, 49*(10), 774–781.

Hollon, S. D., Thase, M. E., & Markowitz, J. C. (2002). Treatment and prevention of depression. *Psychological Science and Public Interest, 3*, 39–77.

Hunnicutt-Ferguson, K., Hoxha, D., & Gollan, J. (2012). Exploring sudden gains in behavioral activation therapy for Major Depressive Disorder. *Behaviour Research and Therapy, 50*(3), 223–230.

Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, 44*(8), 1116–1129.

Kelly, M. A. R., Cyranowski, J. M., & Frank, E. (2007). Sudden gains in interpersonal psychotherapy for depression. *Behaviour Research and Therapy, 45*(11), 2563–2572.

Kelly, M. A. R., Roberts, J. E., & Ciesla, J. A. (2005). Sudden gains in cognitive behavioral treatment for depression: when do they occur and do they matter? *Behaviour Research and Therapy, 43*(6), 703–714.

Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. S. (1984). *Interpersonal psychotherapy for depression*. New York, NY: Basis Books.

7

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*, 877–883.

Laurenceau, J.-P., Hayes, A. M., & Feldman, G. C. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review, 27*(6), 682–695.

Lemmens, L. H. J. M., Arntz, A., Peeters, F. P. M. L., Hollon, S. D., Roefs, A., & Huibers, M. J. H. (2011). Effectiveness, relapse prevention and mechanisms of change of cognitive therapy vs. interpersonal therapy for depression: study protocol for a randomised controlled trial. *Trials, 12*(1), 150–162.

Lemmens, L. H. J. M., Arntz, A., Peeters, F. P. M. L., Hollon, S. D., Roefs, A., & Huibers, M. J. H. (2015). Clinical Effectiveness of Cognitive Therapy vs. Interpersonal Psychotherapy for Depression: Results of a Randomized Controlled Trial. *Psychological Medicine, 45*, 2095-2110.

Tang, T. Z., & DeRubeis, R. J. (1999a). Reconsidering Rapid Early Response in Cognitive Behavioral Therapy for Depression. *Clinical Psychology: Science and Practice, 6*(3), 283–288.

Tang, T. Z., & DeRubeis, R. J. (1999b). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 67*(6), 894–904.

Tang, T. Z., DeRubeis, R. J., Beberman, R., & Pham, T. (2005). Cognitive Changes, Critical Sessions, and Sudden Gains in Cognitive-Behavioral Therapy for Depression. *Journal of Consulting and Clinical Psychology, 73*(1), 168–172.

Tang, T. Z., DeRubeis, R. J., Hollon, S. D., Amsterdam, J., & Shelton, R. (2007). Sudden Gains in Cognitive Therapy of Depression and Depression Relapse/Recurrence. *Journal of Consulting and Clinical Psychology, 75*(3), 404–408.

Vittengl, J. R., Clark, L. A., & Jarrett, R. B. (2005). Validity of Sudden Gains in Acute Phase Treatment of Depression. *Journal of Consulting and Clinical Psychology, 73*(1), 173–182.

Weissman, M. M., Markowitz, J. C., & Klerman, G. L. (2000). *Comprehensive guide to interpersonal psychotherapy*. New York, NY: Basic Books.

Data supplements

Data Supplement I. Results of Univariate Logistic Regression analysis (n = 117).

DEMOGRAPHICS	OR	95% CI	p	QUALITY OF LIFE	OR	95% CI	p
Gender, Female	1.57	[0.70–3.48]	.27	Utility score (EuroQol)*	1.24	[1.04–1.48]	.02#
Age	0.98	[0.95–1.01]	.23	Impairment (Rand 36)			
Partner, yes	1.12	[0.51–2.46]	.77	- Physical functioning	1.02	[1.00–1.04]	.08#
Level of Education, medium	0.67	[0.19–2.31]	.52	- Social functioning	0.99	[0.97–1.02]	.56
Level of Education, high	0.99	[0.38–2.56]	.98	- Role functioning: physical	1.00	[0.99–1.01]	.69
Active Employment, yes	1.58	[0.71–3.53]	.26	- Role functioning: emotional	1.00	[0.99–1.02]	.67
DEPRESSION	**OR**	**95% CI**	**p**	- Mental health	1.00	[0.97–1.03]	.98
Severity I (BDI-II)	1.01	[0.96–1.06]	.70	- Vitality	1.02	[0.99–1.05]	.20
Severity II (QIDS)	0.99	[0.90–1.09]	.83	- Pain	1.01	[0.99–1.02]	.37
Recurrent Depression	0.72	[0.33–1.54]	.39	- General health	0.99	[0.97–1.02]	.50
Hopelessness (BHS)	1.04	[0.96–1.13]	.35	**COMORBIDITY**	**OR**	**95% CI**	**p**
- Feelings about future	1.21	[0.92–1.60]	.18#	No Comorbidity on Axis I	0.38	[0.17–0.85]	.02#
- Loss of motivation	1.03	[0.88–1.21]	.74	No Comorbidity on Axis II	1.49	[0.68–3.24]	32
- Future expectations	1.06	[0.74–1.52]	.75				
PSYCHOLOGICAL DISTRESS	**OR**	**95% CI**	**p**	**GENERAL FUNCTIONING**	**OR**	**95% CI**	**p**
Total BSI score	1.00	[0.99–1.01]	.89	Social Functioning (WSAS)	1.02	[0.97–1.08]	.47
- Somatic Complaints	0.98	[0.91–1.06]	.57	**TREATMENT EXPECTATIONS**	**OR**	**95% CI**	**p**
- Cognitive Problems	0.99	[0.91–1.08]	.81	Treatment will be successful	1.12	[0.83–1.50]	.48
- Interpersonal Sensitivity	1.02	[0.91–1.13]	.77	**PROCESS MEASURES**	**OR**	**95% CI**	**p**
- Depression	0.98	[0.91–1.07]	.70	Dysfunctional attitudes	1.02	[0.99–1.04]	.26
- Anxiety	1.00	[0.92–1.08]	.99	Interpersonal Problems	1.00	[0.99–1.02]	.67
- Hostility	1.06	[0.95–1.18]	.28	Explicit Self-Esteem	1.01	[0.97–1.05]	.76
- Phobic Anxiety	0.95	[0.85–1.05]	.33	Implicit Self-Esteem	0.78	[0.28–2.17]	.63
- Paranoid Symptoms	1.04	[0.95–1.14]	.44	Cognitive Reactivity	1.02	[0.99–1.04]	.26
- Psychoticism	0.96	[0.83–1.08]	.50	Rumination	0.99	[0.95–1.03]	.61
				Attributional style	1.08	[0.75–1.54]	.69

Note: # = p < .20 * = Adjusted score (EuroQol Utility * 10).

Data Supplement II. Frequency, Characteristics, Baseline Predictors and Clinical Impact of Sudden Gains occurring within 14-day between-session interval ($n = 27$).

Frequency and Characteristics	CT ($n = 64$)	IPT ($n = 53$)	
- Patients with SG's, n (%)	19 (29.7)	8 (15.1)	$\chi^2 (1) = 3.5, p = .06$
- Number of SG's, n	25	9	
- Magnitude	10.4 (4.0)	10.6 (3.2)	
- Median pre gain session	7	7	
- Patients with reversals, n (%)	5 (26.3)	1 (12.5)	
- Number of reversals, n	5	2	
- SG Patients maintaining SG's, n (%)	14 (21.9)	11 (20.8)	
Baseline Predictors	**OR**	**95% CI**	**p**
*Full Model**			
- Female sex	0.31	[0.10–0.96]	.04
- BSI: Depression	1.07	[0.90–1.26]	.45
- BSI: Psychoticism	0.82	[0.63–1.07]	.13
- BSI: Phobic Anxiety	1.01	[0.86–1.29]	.64
- Quality of Life Utility score (EuroQol)[1]	1.36	[1.03–1.80]	.03
- Impairment in Social Functioning	1.00	[0.96–1.03]	.82
- Absence of Comorbid Axis I disorder	2.94	[0.88–9.83]	.08
- Absence of Comorbid Axis II disorder	0.37	[0.12–1.19]	.10
- Treatment expectations	1.53	[0.95–2.47]	.08
- Dysfunctional Attitudes (DAS-A-17)	0.99	[0.94–1.05]	.81
- Interpersonal Problems (IIP)	1.00	[0.98–1.03]	.90
- Self-Esteem (SLSC-R)	1.01	[0.94–1.09]	.80
- Attributional style (ASQ)	1.10	[0.61–1.98]	.76
*Final Model***			
- Quality of Life Utility score (EuroQol)[1]	1.38	[1.08–1.76]	< .01
- Absence of Comorbid Axis I Disorder	3.93	[1.40–11.00]	< .01
Sudden Gains and Outcome	**Gainer** **($n = 27$)**	**Non-Gainer** **($n = 90$)**	
- BDI-II score 7 months, M (SD)	10.5 (8.2)	17.4 (12.3)	$F (1) = 6.51, \; p < .01$***
- Remission Rates 7 months, %	45.8%	27.4%	$\chi^2 (1) = 2.95, p = .09$
- BDI-II Score 12 months, M (SD)****	8.7 (8.2)	17.5 (14.2)	$F (1) = 7.77, p < .01$****
- Remission Rates 12 months, %	60.9%	35.0%	$\chi^2 (1) = 4.95, p = .03$

BDI-II = Beck Depression Inventory II; BSI = Brief Symptom Inventory; CT = Cognitive Therapy; IPT = Interpersonal Psychotherapy; SG = Sudden Gain; [1] = Adjusted score (EuroQol Utility * 10); * = Dependent Variable = Sudden Gain(s) (yes/no); ** = Order of elimination: Interpersonal Problems (IIP); Dysfunctional Attitudes (DAS-A-17); Impairment in Social Functioning (RAND36); Self-Esteem (SLSC-R); Attributional style (ASQ); General Psychological Distress (BSI): Depression; General Psychological Distress (BSI): Phobic Anxiety; General Psychological Distress (BSI) Psychoticism; Absence of Comorbid Axis II disorder; Female sex & Treatment expectations; *** = Gain*Interaction condition at 7 months = $F (1) = 0.68, p = .41$; **** = Gain*Outcome interaction at 12 months = $F (1) = 0.24, p = .63$.

8

Revealing the dynamic network structure
of the Beck Depression Inventory-II

Abstract

Structured interviews and questionnaires are important tools to screen for Major Depressive Disorder. Recent research suggests that, in addition to studying the mean level of total scores, researchers should focus on the dynamic relations among depressive symptoms as they unfold over time. Using network analysis, this paper is the first to investigate these patterns of short-term (i.e., session to session) dynamics for a widely used psychological questionnaire for depression – the Beck Depression Inventory–II (BDI-II). With the newly developed vector autoregressive (VAR) multilevel method we estimated the network of symptom dynamics that characterizes the BDI-II, based on repeated administrations of the questionnaire to a group of depressed individuals who participated in a treatment study of an average of 14 weekly assessments. Also the centrality of symptoms and the community structure of the network were examined. The analysis showed that all BDI-II symptoms are directly or indirectly connected through patterns of temporal influence. In addition, these influences are mutually reinforcing, with "loss of pleasure" being the most central item in the network. Community analyses indicated that the dynamic structure of the BDI-II involves two clusters, which is consistent with earlier psychometric analyses. The network approach expands the range of depression research, making it possible to investigate the dynamic architecture of depression and opening up a whole new range of questions and analyses. Regarding clinical practice, network analyses may be used to indicate which symptoms should be targeted, and in this sense may help in setting up treatment strategies.

8

Introduction

Major Depressive Disorder (MDD) is a complex and burdensome mental health disorder made up of a wide variety of symptoms (World Health Organization (WHO), 2001; Kessler et al., 2003; Hardeveld, Spijker, de Graaf, Nolen, & Beekman, 2010; American Psychiatric Association (APA), 2000). Structured interviews and questionnaires, such as the Hamilton Depression Rating Scale (HDRS; Hamilton, 1960) and Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961; BDI-II; Beck, Steer, & Brown, 1996b) are important and commonly used tools to screen for, study, and follow the course of MDD (Beck, Steer, Ball, & Ranieri, 1996a). In longitudinal studies, a total score, which results from simply adding all symptom scores, is often used as a measure of changes in depression severity. Relatively few studies use a more fine-grained analysis, in which the reduction of depression severity is studied by examining specific (clusters of) symptoms of depression instead of using the total score (e.g., Bhar et al., 2008; Fournier et al., 2013; Stewart & Harkness, 2012).

What all the above studies have in common is that they are based on the latent variable model. According to this model, symptoms of a given disorder are assumed to share an essential property; namely, their causal dependence on a latent variable, from which all symptoms arise (Borsboom, 2008; Kendler, Zachar, & Craver, 2011). In this perspective, symptoms experienced by patients are merely effects of the relevant latent variable (in this case depression). Standard models assume the symptoms to be statistically independent given the latent variable, and as a result, symptom associations are viewed to be spurious (Borsboom, 2008). Specifically, in the standard model, symptoms are not considered to have autonomous influence on one another. The latent variable approach, as utilized in standard models, is therefore not suitable to examine the dynamic relations between symptoms. The recently developed network approach (Cramer, Waldorp, van der Maas, & Borsboom, 2010) steps away from this latent variable model by proposing that research should no longer focus exclusively on the *mean level* of symptoms or change therein (e.g., an overall score or a reduction of symptoms). Instead, this approach emphasizes that clinical research should also focus on the relation between individual symptoms from one time point to another, which we denote here as *short-term dynamics*[1].

The importance of examining these short-term dynamics is supported by various sources. First, theories of treatments for depression tend to focus on the short-term symptom dynamics when describing their proposed mechanisms of change. For example, according to Beck's (1979) cognitive theory, change in cognitive processes (e.g. negative thinking) leads to changes in symptoms such as affect (Beck, 1964; Rush, Kovacs, Beck, Weissenburger, & Hollon, 1981). Second, in clinical practice it is commonly observed that if patients experience relief in one symptom (for example sleeping problems), other symptoms start to wane as well, indicating the start of recovery. This is especially notable when symptoms are systematically assessed at the start of each therapy session, as is the case in cognitive therapy (Beck, Rush, Shaw, & Emery, 1979). Third, recent studies indicate that depression risk factors and stressful life events have differential effects on depressive

8

symptoms (Cramer, Borsboom, Aggen, & Kendler, 2012; Fried, Nesse, Zivin, Guille, & Sen, 2013). As Cramer et al. (2012) showed, correlations between symptoms were directly influenced by the stressful life events and could not be explained by changes in an underlying common cause, in this case the risk to develop depression. This further supports the idea that symptoms have an autonomous influence on one another. Being able to objectively describe such symptom-by-symptom interactions can give important clues for clinical research and practice.

Apart from their substantive plausibility, network approaches open up a new range of research questions. For example, estimating a network of symptoms from depression questionnaires allows for an objective assessment of the centrality of symptoms (Opsahl, Agneessens, & Skvoretz, 2010; Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006). Symptoms with a central position in the network are probably the most important or influential ones and are therefore likely to cause the symptom spread to continue. Studying these central symptoms can give clues for further clinical research. One could investigate, for instance, the commonly held assumption that anhedonia (loss of pleasure and interest) and depressed mood are central symptoms of depression as stated by the most prevalent diagnostic systems DSM-IV-TR (APA, 2000) and ICD-10 (WHO, 2008). In addition, once the network has been estimated, the community structure of the network can be examined (Girvan & Newman, 2002). A community is present if some clusters of symptoms are more strongly interconnected with each other than with symptoms that are not part of the cluster. In this way, the dynamic architecture of depression can be investigated.

This paper will be the first to investigate the short-term dynamics of one of the most widely used psychological questionnaire for depression: the Beck Depression Inventory–II (BDI-II: Beck et al., 1996b). Inspired by the possibilities of the network approach, we will apply a novel method developed by Bringmann et al. (2013) that is able to explore these symptom dynamics, and infer a network structure of the BDI-II symptoms. Until recently, it was not possible to infer these kinds of directed and weighted networks from clinical questionnaires since two important requirements for studying short-term dynamics – intensive longitudinal data in which a set of symptoms is measured frequently across time, and a suitable statistical method – were lacking. Intensive longitudinal data is still sparse, but in a recent study of Lemmens et al. (2011; 2015) such data for the BDI-II were collected. Second, the newly developed vector autoregressive (VAR) multilevel method, which is a combination of multilevel (hierarchical) and time series models, is suited for analysing these kinds of clinical longitudinal data. These data have rather short time series (ca. 20 time points) for a large sample of patients (Bringmann et al., 2013). Note that since only few studies have investigated single specific (clusters of) symptoms of the BDI-II or even interactions between symptoms in general, all analyses are exploratory.

The structure of the paper is as follows: First, we will infer the network(s) representing the short-term dynamics of the BDI-II symptoms. Second, we will study the centrality of symptoms. Based on the DSM-IV and ICD-10, one would expect the BDI-II items that are intuitively most closely related to the main symptoms, anhedonia and depressed mood

(namely items: "loss of interest", "loss of pleasure" and "sadness"), to be the most central ones in the network(s). In the third and last part, we will analyse whether communities are present in the BDI-II network(s). Since the network(s) consist(s) of a fair number of symptoms (i.e., 21), we expect the emergence of new clusters of symptoms or community structures.

Method

Data

The data in the current study come from a large randomized clinical trial (RCT), which examined the effectiveness, relapse prevention and mechanisms of change of Cognitive therapy (CT) vs. Interpersonal Psychotherapy (IPT) for depression (Lemmens et al. 2011, 2015). In this study, 182 patients (age between 18 and 65) with a DSM-IV diagnosis of Major Depressive Disorder (MDD) were randomly allocated to one of three conditions: (a) CT ($n = 76$), (b) IPT ($n = 75$), or (c) an 8-week waiting-list control (WLC) condition followed by treatment of choice (CT or IPT; $n = 31$). In the current study, we did not differentiate between patients who started therapy immediately and who started after 8 weeks. This resulted in a sample size of 99 for the CT condition (mean age of 40 years and SD = 12; 80% female) and a sample size of 83 for the IPT condition (mean age of 41 years and SD = 12; 64% female). There were no significant differences in demographic and clinical characteristics between the groups. Each patient participated in 3 to 20 weekly individual sessions, depending on the progress of the patient or due to drop out. On average, patients completed 14 sessions (SD = 5)[2]. The BDI-II was administered before each session to assess depression severity. Of the 2661 sessions, 2.5% of the BDI-II data were missing. Further details concerning the design of the trial and effectiveness of the interventions have been fully reported elsewhere.

Beck Depression Inventory-II (BDI-II)

The Beck Depression Inventory-II (BDI-II; Beck et al., 1996b; Dutch translation by van der Does, 2002) is one of the most widely used and empirically validated questionnaires for screening depression. The BDI-II is a self-report questionnaire measuring the severity of depression with 21 items. Each item is rated on a 4 point Likert-scale ranging from 0 to 3. The total score, ranging from 0 to 63, is constructed by adding the item scores, with higher scores reflecting more severe depressive symptomatology.

Interventions

CT and IPT are two of the most empirically validated psychotherapies used for treating depression (e.g. Cuijpers, van Straten, Andersson, & van Oppen, 2008; Cuijpers et al., 2011;

8

Hollon, Thase, & Markowitz, 2002). CT is based on Beck's (1979) cognitive theory, which states that depression results from maladaptive information-processing strategies that are maintained by dysfunctional behavioural responses. CT focuses on identifying and changing dysfunctional cognitions, schemas and attitudes in order to treat depression. In IPT, the interpersonal model of depression is central (Klerman, Weissman, Rounsaville, & Chevron, 1984). According to this model, major disturbances in the interpersonal domain may cause and maintain depression. It is assumed that depressive symptoms can be reduced through the improvement of interpersonal functioning.

Statistical analysis

The BDI-II network

First, we inferred the BDI-II network by analysing the short-term dynamics between the 21 symptoms across the 20 weeks of therapy with a modified version of the multilevel-Vector Autoregressive (VAR) method (Bringmann et al., 2013)[3]. In the multilevel-VAR method, the time dynamics between the 21 symptoms of the BDI-II from one moment to the other are represented by a VAR model (see also Tschacher, Zorn, & Ramseyer, 2012 for a similar approach). In the VAR model, the dependent variable (e.g., symptom "sadness"; item 1) at time point t (e.g., session 2) is regressed on the lagged t-1 (e.g., session 1) versions of the independent variables (Box, Jenkins, & Reinsel, 1994; Walls & Schafer, 2006)[4]. The independent variables in this study are all the symptoms of the BDI-II, measured at the previous time point (in this case the previous session). To account for differences between patients, all regression coefficients were assumed to be normally distributed at the population level. As a consequence, we obtained a multilevel model consisting of fixed (average) and random (individual) effects[5]. Each BDI-II symptom was used as a criterion variable once, which means that 21 multilevel-VAR models were estimated.

In order to estimate a multilevel-VAR model, data need to be stationary. An implication of this assumption is that that the variables will fluctuate around the same mean over time (Lütkepohl, 2005). Since the BDI-II symptoms decreased over the course of treatment (Lemmens et al., 2015), the means changed significantly, which indicates a non-stationary process. For this reason, a linear trend in the multilevel-VAR model was included, making the data trend stationary (Hamaker & Dolan, 2009). This implies that the short-term dynamics or the session-to-session fluctuations of the symptoms (as represented by the network) and the decrease of symptoms across the sessions (as represented by the linear trend) are modelled separately. Therefore, change in the short-term dynamics is in principle unrelated to change in the *mean level* of the BDI-II symptoms. Note further that stationarity also implies the assumption that the effects of symptoms on other symptoms are stable across time.

In order to obtain the BDI-II network, the estimated fixed effects of the multilevel-VAR analyses were used (Snijders & Bosker, 2012). Fixed effects represent the average con-

nection strengths of the arrows in the network among the 21 symptoms and indicate whether the symptoms are positively or negatively related to each other. The fixed effects represent either autoregressive effects (self-loops) or cross-regressive effects (connections between different variables) in the network. Note that the network only represents the dynamic relations between the symptoms (the slopes of the multilevel-VAR model) and not the mean scores (the intercepts of the multilevel-VAR model) of the symptoms.

The estimated fixed effects or connections of the network resulted in a directed weighted network structure of the BDI-II, which was visualized using *qgraph* (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012), a package for the statistical programming language *R*. Arrows or connections in the network represent more than mere associations between symptoms: because symptoms are measured over time, the connections can be viewed as an approximation of causality, resembling Granger-causality (Granger, 1969; Tschacher et al., 2012). The network analyses were based on all the connections of the network. However, for reasons of clarity, we only visually present the strongest connections in the inferred network; that is, those connections which surpass the significance threshold (5%) using the False Discovery Rate (FDR) method (Benjamini & Hochberg, 1995; Benjamini, Krieger, & Yekutieli, 2006). In the visually presented network, symptoms that are more strongly related to each other tend to be closer together in the figure (this is a result of the node placement algorithm; see Fruchterman & Reingold, 1991; Epskamp et al., 2012).

Since the current study included two different therapy groups (CT and IPT), it is possible that two different network structures give rise to the data. We tested this in two ways: First, we fitted a model with the two networks separately and cross-correlated their estimated network links. Secondly, we compared a model in which we included two networks with a model that had one common network; for this purpose, we used the Bayesian Information Criterion (BIC; Schwarz, 1978). The model with the lowest BIC is the preferred model[6].

Centrality analysis

In the second analysis, the inferred network was further analysed by estimating the centrality of the BDI-II symptoms. In a centrality analysis, one can determine the relative importance or influence of a symptom in the network. We performed three types of centrality analyses: outdegree, indegree and betweenness centrality (see Opsahl et al., 2010)[7]. Outdegree centrality indicates how many outgoing arrows or how much information a symptom sends to other symptoms it is directly connected to. In the same way, indegree centrality indicates how many incoming arrows a symptom receives from the directly connected symptoms. Betweenness centrality takes into account both the direct and indirect connections of a symptom. A symptom with a high betweenness centrality is a symptom located on many paths between other symptoms and thus is a symptom through which the information in the network has to pass often. Therefore, a symptom with a high betweenness centrality is important in funneling the information flow or the symptom spread in the network.

8

Community structure analysis

As a third analysis, we performed a community structure analysis. In complex networks, new structures of clusters can often be found. An example of such a cluster is a *community*, in which groups of symptoms are densely interconnected among each other, but sparsely connected to the overall network. We used the *Walktrap* algorithm, which is suited for weighted networks (Pons & Latapy, 2005). This algorithm does not take directions of the arrows into account, so we summed the connection strengths (arrows) between two symptoms to have an appropriate undirected network suitable for analysis. The Walktrap algorithm uses random walks on the network to find communities or densely interconnected symptoms. The algorithm reveals how many groups can be found and also to which group a symptom of the network belongs to. All the analyses were done in the statistical software *R* (R Core Team, 2014).

Results

The BDI-II network

Figure 1 shows the inferred network of the dynamics between the 21 BDI-II symptoms. The analysis of cross-correlations and the model fitting approach using the BIC indicated that the network structure did not differ across the two therapy groups ($r = 0.86$, $p < 0.0001$; one network: BIC= 77367.65 versus two networks: BIC= 80574.09). Therefore, only one network was needed, representing both treatment groups. From the figure, it is evident that the strongest connections between symptoms are all positive in sign. Thus, when a symptom score increases, it is likely that other symptom scores also increase the next session, leading to an increase in the severity of symptoms in general. For example, if a participant reports feelings of guilt ("guilty feelings", item 5) in one session, that participant is more likely to report feelings of failure about the past ("past failure", item 3) the next session. Similarly, if a symptom score decreases, it is likely that other symptom scores decrease as well. The strength of the relation between symptoms translates into the thickness of the arrows in the figure: the stronger the symptoms are related the thicker the arrow between two symptoms, and the closer the symptoms tend to be together in the figure. This is expressed in, for example, the placement of the symptoms "past failure" (item 3) and "worthlessness" (item 14).

Apart from the connections between the symptoms, self-loops can contain important information. For example, the self-loop of the symptom "loss of interest in sex" (item 21) is clearly the strongest connection of the network, meaning that when a participant reports loss of interest in sex one session, he or she is highly likely to report this in the next session as well. Furthermore, self-sustaining loops are apparent in the network. For example, "worthlessness" (item 14) and "guilty feelings" (item 5) seem to

mutually influence each other. It should be mentioned that there are negative connections in the complete network as well (i.e. increase in one symptom is associated with decrease in another – and vice versa). However, since these are rather weak, they did not pass the threshold for visualization in Figure 1. Note, however, that all connections are taken into account in the further analyses[8–9].
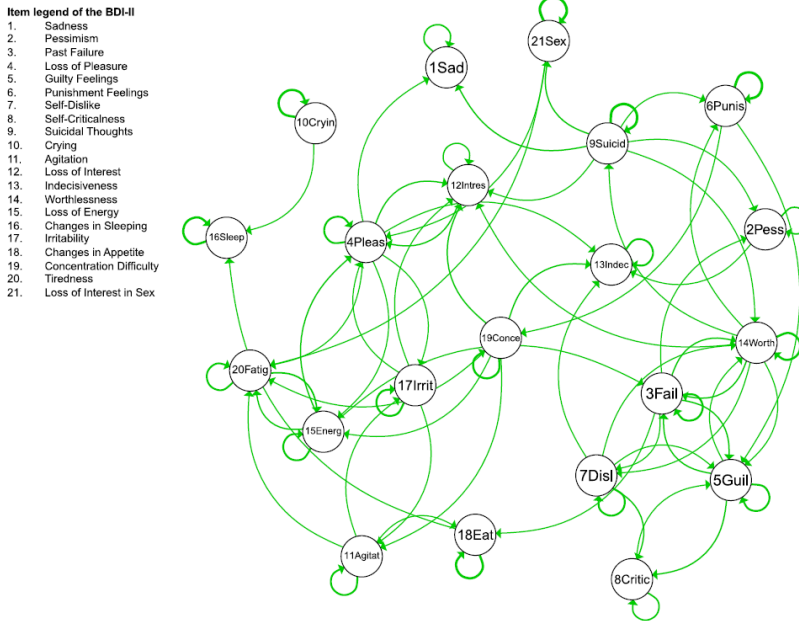
Figure 1. The BDI-II network. *Note:* In this network, the connections between the 21 symptoms that surpass the significance threshold are visualized. Because of multiple hypotheses testing, we did not use the traditional 0.05 cut-off for *p*-values as the standard (which would inflate the number of unimportant links to be visualized); instead, we control the False Discovery Rate or FDR at 5% (Benjamini & Hochberg, 1995; Benjamini et al., 2006). Here, the 75 connections that pass the FDR threshold are visualized.

Centrality analysis

Figure 2 presents the results of the centrality analysis. The left panel of the figure indicates that the symptom "loss of pleasure" (item 4) has one of the highest outdegrees, meaning that when one reports loss of pleasure in one session, it is likely that one will also report an increase in other symptoms in the next session. This is in contrast to, for instance, the symptom "changes in sleeping patterns" (item 16), which is less likely to directly affect other symptoms the next session.

8

The middle panel indicates that the symptoms "indecisiveness" (item 13), "loss of interest" (item 12), "past failure" (item 3) and "sadness" (item 1) feature higher indegrees and thus receive a lot of information from other symptoms. This is in contrast to "suicidal thoughts" (item 9): this symptom is unlikely to be influenced by other symptoms, and is more likely to influence other symptoms (see also the first panel again).

The right panel indicates that the symptoms "loss of pleasure" (item 4) and "past failure" (item 3) feature the highest betweenness centralities, but they also have one of the highest outdegree ("loss of pleasure") and indegree ("past failure") centrality scores, respectively. Thus, the symptoms "loss of pleasure" and "past failure" are important in funneling the activation flow or symptom spread in the network.
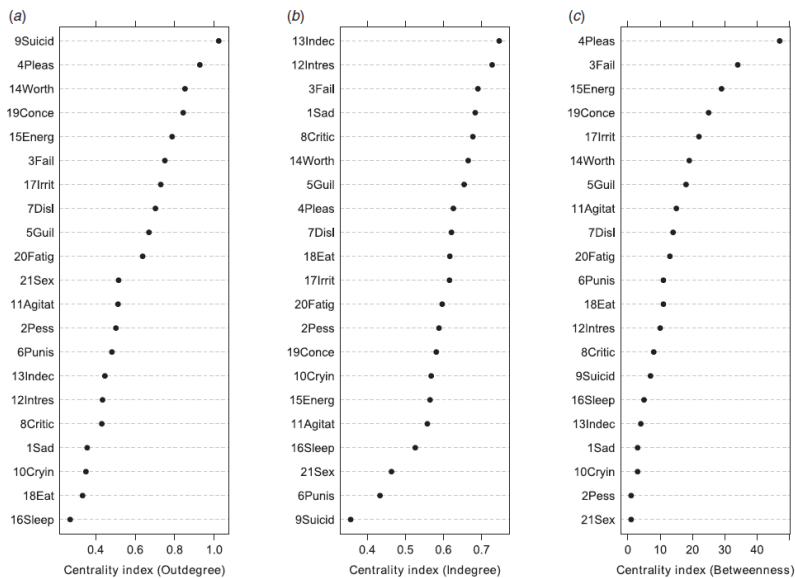

Figure 2. Outdegree, Indegree and Betweenness centrality for all BDI-II symptoms. *Note:* The black dots are the model-based estimates of outdegree, indegree, and betweenness centrality. The higher the centrality index score the more central the symptom is in the network.

Community structure of the BDI-II network

The community structure analysis using the Walktrap algorithm indicated a two-cluster solution (see Figure 3)[10]. This community structure means that symptoms in one cluster are more densely interconnected among themselves and more sparsely connected to symptoms in another cluster. The green cluster in Figure 3 consists of the symptoms "guilty feelings" (item 5), "past failure" (item 3), "self-dislike" (item 7), "self-criticalness" (item 8), "worthlessness" (item 14), "punishment feelings" (item 6) and "pessi-

mism" (item 2), which are often described as cognitive symptoms. Items in the yellow cluster mainly consist of physical and affective symptoms of depression that appear related to loss of energy and pleasure.
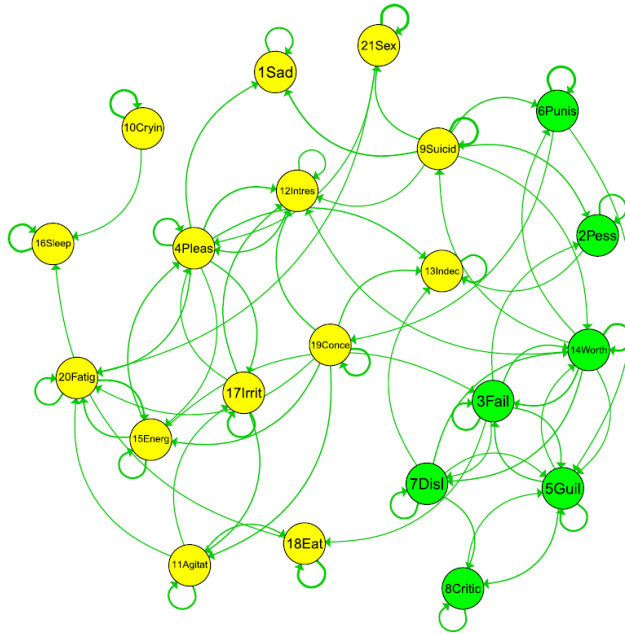
Figure 3. Community structure of the BDI-II network with the two clusters indicated in two different colours.

Discussion

In this paper, we derived for the first time a network that represents the session-to-session dynamics of one of the most widely used and empirically validated self-report measures for assessing the severity of depression: the BDI-II (Beck et al., 1996b). Results indicate that, in this network, all BDI-II symptoms are directly or indirectly connected. In addition, the strongest connections between the symptoms are uniformly positive, indicating that, in general, when a symptom changes in severity, other symptoms tend to change in the same direction. This pattern of symptom dynamics is independent of the overall decrease in symptom scores as this trend was modelled separately; hence, the pattern does not reflect the common influence of treatment or recovery. In addition, for each symptom-symptom relation, we controlled for the effect of all other symptoms in the network. Thus, the evidence from this study points to the conclusion

8

that direct effects among symptoms of the BDI-II are prevalent, and in fact connect all symptoms assessed in the questionnaire. In other words, symptoms of depression form a network of direct interactions.

Centrality analyses of the symptoms suggested that some symptoms are likely to have a larger influence on the symptom spread than other symptoms. As one may expect based on e.g. the DSM-IV, the symptom "loss of pleasure" (item 4) is one of the most central items in the symptom network and thus has a relatively large effect on the enduring of depressive symptoms in general. Somewhat more surprisingly, the symptoms "sadness" (item 1) and "loss of interest" (item 12) have a high indegree centrality, but quite a low outdegree and betweenness centrality, which means that they tend to increase in severity as other symptoms become more severe, but do not play a large role in funneling the symptom spread themselves. Thus, these symptoms may serve a mainly reactive role in the short-term dynamics of depression. Additional studies are needed to confirm these results, preferably engaging different depression questionnaires, such as the Hamilton Depression Rating Scale (HDRS; Hamilton, 1960).

Based on theory, one may expect a difference in symptom dynamics for subjects receiving CT and IPT because both treatments are assumed to work through different mechanisms. We did not observe such differences. A potential explanation for our findings could be that the dynamics between symptoms are similar when the treatments that are being compared are equally effective in reducing pathology, a fact that has been well established for CT and IPT for depression (Cuijpers et al., 2008; Hollon & Ponniah, 2010). Alternatively, it might be the case that differences between CT and IPT actually do exist, but that we did not capture them in the current study because the BDI-II is, due to its design, insensitive for the differences between the two treatments. For example, even though the BDI-II includes items on several cognitive components (key elements of CT), items referring to problems in the interpersonal domain (core of IPT) are lacking. Further research involving other questionnaires is necessary to indicate if there are differences in symptom dynamics between therapies. A final possibility is that the difference between CT and IPT does not lie in symptom-symptom interaction, as studied in this paper, but in differences that arise in, e.g., stepwise changes in symptomatology. In this case, therapy effects might be detected in the way symptoms decrease or increase from one time point to another. Models that may be used to analyse such differences, while accounting for the network of symptom-symptom interaction, are currently unavailable. However, non-linear statistical network inference techniques that may be used to model such processes are within reach, and could be used to investigate this issue in the future.

In the present study, community analyses revealed two groups of symptoms. The result appears to accommodate emerging evidence from the biomedical literature, which points to two types of depression: melancholia and atypical depression (Lamers et al., 2010, 2012). The current community clusters resemble these different depression types, as the green cluster in Figure 3 has similarities to melancholic type, whereas the yellow cluster resembles atypical depression. It is also interesting to note that the com-

munity structure result, based on multiple time points, is similar to the two-factor solution of the BDI-II, based on pooling across subjects at one time point (as found in e.g. Beck et al., 1996b; see also Arnau, Meagher, Norris, & Bramson, 2001; Steer, Ball, Ranieri, & Beck, 1999). Except for "suicidal thoughts", all other symptoms in the green cluster of Figure 3 are the same as in the cognitive dimension of the two-factor solution of the BDI-II, whereas the yellow cluster could be interpreted as the somatic-affective or non-cognitive dimension. Although it is a good sign that the results we find are consistent with what one typically finds using factor analysis, our approach leads to a different way of thinking, different strategies for intervention, and to very different conclusions. In the latent variable approach, there are just two clusters of symptoms, which is a static result. In the network view, the result concerns the communication between symptoms that is denser within the cluster than with symptoms that are not in the cluster, leading to new hypotheses on how interventions should be operationalized, namely focusing on the interaction between symptoms. Thus, the existence of such patterns of influence is not a replication of the results of factor analysis on individual differences; rather, it may be seen as a potential explanation for these results (Wichers, 2014).

Several findings of this paper suggest further research. One important issue is how our results, which only involved participants with a diagnosis of depression, compare to results from unaffected individuals. For example, it is important to investigate whether a similar network characterizes healthy individuals. One hypothesis would be that there are no distinct symptom clusters in healthy subjects, but that instead all symptoms are similarly (and weakly) connected. Such a network would be more resilient, since activation would not spread as easily, and it would be less likely to 'get stuck' in a cluster of symptoms. Another important topic for future research involves the difficult question of how to relate different time scales (Boker, Molenaar, & Nesselroade, 2009). This is because the symptoms that characterize depression are likely to influence each other in different time windows. For example, sleep problems are likely to exert effects in a pattern of a day-to-day variation, whereas mood states are much quicker and may affect each other within minutes. The question of how the dynamics of these different time scales interact with each other is, in our view, one of the main puzzles to be solved in the study of symptom dynamics.

Regarding clinical practice, the relevance of the methodology and results of our approach may lie in opportunities to determine symptom centrality. For example, network analyses may be used to indicate which symptoms should be targeted first, and in this sense may help in setting up treatment strategies. Ideally, such analyses should be based on person-specific analyses (cf. Molenaar & Campbell, 2009). Unfortunately, at the moment such analyses are not computationally feasible for large networks of 21 symptoms. However, future development of the multilevel-VAR method, combined with a higher frequency of within subject assessment, should make it possible to take this procedure a step further, which may eventually lead to person-specific therapeutic interventions. Information about person-specific network centrality would not necessarily

8

require pre-treatment assessment, and the high frequency assessment could be informative at any point, even if started during therapy. For example, if a centrality analysis of an individual network reveals that for that specific person "loss of pleasure" is the most central symptom, therapy that intervenes on this symptom would be more effective than treatment that intervenes on non-central symptoms; for other persons, different interventions may be preferable. In a similar vein, one could hypothesize that if "suicidal thoughts" is the most central symptom for a given person, this may signal acute need for care. Furthermore, since "suicidal thoughts" has a high outdegree, and is thus likely to trigger other symptoms, but a low indegree, and is thus not likely to be influenced by the other symptoms, interventions should be directly targeted at this symptom. Given the increased opportunities for assessing highly intensive time series within individuals, person-specific treatment protocols based on networks of symptom dynamics are rapidly becoming a realistic possibility. Thus, the network perspective is a promising new research field, which can give guidance to research on depression and to psychological research in general.

Acknowledgments

Footnotes

[1] Note that short-term dynamics refers to the dynamics between time points that are close to one another (e.g. time point t-1 and t). This in contrast to looking at changes in average values (mean levels), which can also be seen as a long-term dynamics. However, we do not use the latter term, since the term "change in mean level" is more intuitive.

[2] Analyses indicated that there were no differences in demographic and clinical characteristics between subjects that dropped out and those who finished therapy (at least 12 sessions).

[3] Note that we deviate in this paper slightly from the procedure as proposed by Bringmann et al. (2013). With 21 items, it is not computationally possible to include all 21 random effects in the multilevel-VAR model simultaneously. Instead, we included only 5 random effects (including the autoregressive coefficient and the intercept) at the same time in a stepwise matter. Simulations indicated that the fixed effects could still be estimated precisely with this number of subjects and time points, which means that this is a feasible approach for estimating the current average network.

[4] Theoretically further lags are also possible. For example, a lag 2 model would indicate how symptoms are related to all symptoms experienced two sessions and one session ago. However, model comparison indicated that lag 1 was a more likely model than a lag 2 model (BIC lag 1: 71162, BIC lag 2: 71539).

[5] Simulations (not reported here) have indicated that because it is computationally not possible to include all 21 random effects at once in the multilevel-VAR model, the variance components (random effects vari-

ances) cannot be estimated accurately enough. For this reason, they will not be discussed further in the paper. The random effects should not be left out of the model though, because their inclusion leads to a more precise estimate of the fixed effects.

[6] The BIC was calculated by taking the average of the BICs of the separate univariate models.

[7] Since we want to estimate the centrality between the symptoms, self-loops are not taken into account in the centrality analyses. However, in all other analyses self-loops are taken into account.

[8] We also confirmed that the connections in the network and thus the relationships between symptoms are not driven by differential variability. Standardizing the data per patient and per symptom led to a network that was highly similar to the original network; the correlation between parameters in the original and standardized network was 0.99. As a result, the conclusions of this paper are robust with respect to standardization of the data and are unlikely to reflect differential symptom variability.

[9] Proportional odds logistic regression (POLR), which is a regression model for ordinal response variables, also showed highly similar results; the correlation between parameters in the original and POLR network was 0.96 and led to similar centrality and community cluster results.

[10] A hierarchical cluster analysis on the sum of the weighted links gave highly similar results.

8

References

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., Text Revision). Washington, D.C.: American Psychiatric Association.

Arnau, R. C., Meagher, M. W., Norris, M. P., & Bramson, R. (2001). Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients. *Health Psychology, 20,* 112–119.

Beck, A. T. (1964). Thinking and depression: II. Theory and therapy. *Archives of General Psychiatry, 10,* 561–571.

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

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. (1996a). Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. *Journal of Personality Assessment, 67,* 588–97.

Beck, A. T., Steer, R., & Brown, G. K. (1996b). *Beck Depression Inventory II: Manual.* Boston, MA: Hartcourt Brace.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of general psychiatry, 4,* 561–71.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B Methodological, 57,* 289–300.

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate.*Biometrika, 93*(3), 491–507.

Bhar, S. S., Gelfand, L. A., Schmid, S. P., Gallop, R., DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., & Beck. A. T. (2008). Sequence of improvement in depressive symptoms across cognitive therapy and pharmacotherapy. *Journal of Affective Disorders, 110(1–2),* 161–166.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D-U. (2006). Complex networks: structure and dynamics. *Physics Reports, 424,* 175–308.

Boker, S. M., Molenaar, P., & Nesselroade, J. R. (2009). Issues in intraindividual variability: individual differences in equilibria and dynamics over multiple time scales. *Psychology and aging, 24,* 858–862.

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology, 64,* 1089–1108.

Box, J. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control.* Englewood Cliffs, NJ: Prentice Hall.

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PloS One, 8,* e60188.

Cramer, A. O. J., Borsboom, D., Aggen, S. H., & Kendler, K. S. (2012). The pathoplasticity of dysphoric episodes: differential impact of stressful life events on the patterns of depressive symptom inter-correlations. *Psychological Medicine, 42,* 957–965.

Cramer, A. O. J., Waldorp, L. J., van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: a network perspective. *Behavioral and Brain Sciences, 33,* 137–193.

Cuijpers, P., Geraedts, A. S., van Oppen, P., Andersson, G., Markowitz, J. C., & van Straten, A. (2011). Interpersonal psychotherapy for depression: a meta-analysis. *American Journal of Psychiatry, 168*, 137–193.

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical psychology, 76,* 909–922.

Does van der, W. (2002). *BDI-II-NL: Handleiding; De Nederlandse versie van de Beck Depressie Inventory-II.* Enschede: Ipskamp.

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). Qgraph: network representations of relationships in data. R package version 0.4.10 (http://CRAN.R-project.org/package=qgraph).

Fournier, J. C., DeRubeis, R. J., Hollon, S. D., Gallop, R., Shelton, R. C., & Amsterdam, J. D. (2013). Differential Change in Specific Depressive Symptoms during Antidepressant Medication or Cognitive Therapy. *Behaviour research and therapy, 51,* 392–398

Fried, E. I., Nesse, R. M., Zivin, K., Guille, C., & Sen, S. (2013). Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors. *Psychological Medicine, 44,* 2067–2076.

Fruchterman, T. M. J, & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software–Practice & Experience, 21,* 1129–1164.

Girvan, M., & Newman, M. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences, 99,* 7821–7826.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37,* 424–438.

Hamaker, E. L., & Dolan, C. V. (2009). Idiographic Data Analysis: Quantitative Methods—From Simple to Advanced. In J. Valsiner (Eds.), *Dynamic Process Methodology in the Social and Developmental Sciences*. New York, NY: Springer.

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

Hardeveld, F., Spijker, J., de Graaf, R., Nolen, W. A., & Beekman, A. T. F. (2010). Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatrica Scandinavica, 122,* 184–191.

Hollon, S. D., & Ponniah, K. (2010). A review of empirically supported psychological therapies for mood disorders in adults. *Depression and anxiety, 27,* 891–932.

Hollon, S. D., Thase, M. E., & Markowitz, J. C. (2002). Treatment and prevention of depression. *Psychological Science in the Public Interest, 3,* 39–77.

Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders? *Psychological Medicine, 41,* 1143–1150.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., Rush, A. J., Walters, E. E., & Wang, P. S. (2003).

The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *Journal of the American Medical Association, 289,* 3095–3105.

8

Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. S. (1984). *Interpersonal Psychotherapy for depression*. New York, NY: Basic Books.

Lamers, F., de Jonge, P., Nolen, W. A., Smit, J. H., Zitman, F. G., Beekman, A. T., & Penninx, B. W. J. H.(2010). Identifying

depressive subtypes in a large cohort study: results from the Netherlands Study of Depression and Anxiety (NESDA). *The Journal of clinical psychiatry, 71,* 1582–1589.

Lamers, F., Rhebergen, D., Merikangas, K. R., de Jonge, P., Beekman, A. T. F., & Penninx, B. W. J. H. (2012). Stability and transitions of depressive subtypes over a 2-year follow-up. *Psychological Medicine, 17,* 1–11.

Lemmens, L. H. J. M., Arntz, A., Peeters, F., Hollon, S. D., Roefs, A., & Huibers, M. J. H. (2011). Effectiveness, relapse prevention and mechanisms of change of cognitive therapy vs. interpersonal therapy for depression: Study protocol for a randomised controlled trial. *Trials, 12,* 150–162.

Lemmens, L. H. J. M., Arntz, A., Peeters, F., Hollon, S. D., Roefs, A., & Huibers, M. J.H. (2015). Clinical Effectiveness of Cognitive Therapy vs. Interpersonal Psychotherapy for Depression: Results of a Randomised Controlled Trial. *Psychological Medicine, 45*, 2095-2110.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin: Springer.

Molenaar, P. C., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18,* 112–117.

Opsahl, T., Agneessens, F., & Skvoretz, F. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks, 32,* 245–251.

Pons, P., & Latapy, M. (2005). *Computing communities in large networks using random walks. In Computer and Information Sciences-ISCIS*. Berlin: Springer.

R Core Team (2014). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing (http://www.R-project.org).

Rush, A. J., Kovacs, M., Beck, A. T., Weissenburger J., & Hollon, S. D. (1981). Differential effects of cognitive therapy and pharmacotherapy on depressive symptoms. *Journal of Affective Disorders, 3,* 221–229.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464.

Snijders, T., & Bosker, R. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications.

Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1999). Dimensions of the Beck depression inventory-II in clinically depressed outpatients. *Journal of clinical psychology, 55,* 117–128.

Stewart, J. G., & Harkness, K. L. (2012). Symptom specificity in the acute treatment of major depressive disorder: a re-analysis of the treatment of depression collaborative research program. *Journal of Affective Disorders, 137,* 87–97.

Tschacher, W., Zorn, P., & Ramseyer, F. (2012). Change Mechanisms of Schema-Centered Group Psychotherapy with Personality Disorder Patients. *PLoS ONE, 7,* e39687.

Walls, T. A., & Schafer, J. L. (2006). *Models for Intensive Longitudinal Data*. Oxford: Oxford University Press.

Wichers, M. (2014). The dynamic nature of depression: a new micro-level perspective of mental disorder that meets current challenges. *Psychological medicine, 44,* 1349–1360.

World Health Organization (2001). *The world health report. Mental health: new understanding, new hope.* Geneva: World Health Organization.

World Health Organization (2008). *ICD-10: International statistical classification of diseases and related health problems* (10th Rev. ed.). New York, NY: World Health Organization.

8

9

General Discussion

In the present dissertation, the clinical effectiveness and mechanisms of change of Cognitive Therapy (CT) and Interpersonal Psychotherapy (IPT) for depression were examined in the context of a large RCT. The dissertation started with a general introduction (chapter 1), followed by a description of the study design (chapter 2). Chapter 3 described the clinical effectiveness of CT and IPT in the acute phase and reported on the short-term follow-up (FU). Most of this dissertation (chapters 4 – 8) was aimed at increasing our understanding of the underlying processes that facilitate symptom change. Mechanisms were studied from different perspectives, and in the light of the most recent developments in the field. In chapter 4, the value of an implicit self-associative measure specific to core beliefs of depression was explored. Chapter 5 provided a systematic overview and critical evaluation of the empirical literature on psychological mediators in psychotherapy for depression, the most commonly used method to examine mechanisms of change. In chapter 6, the results of an empirical study on the mediational role of various therapy specific and non-specific potential mediators in CT and IPT were reported. Chapter 7 examined the occurrence and clinical impact of sudden gains, and explored baseline characteristics associated with these large and sudden symptom improvements. Chapter 8 explored the dynamic relations between individual symptoms of depression over the course of treatment.

This final chapter summarizes the main findings presented in previous chapters and integrates them in the broader knowledge base with a critical discussion. Furthermore, methodological considerations relevant to this dissertation will be addressed, and implications for clinical practice and future research will be provided. The chapter ends with an overall conclusion of the dissertation.

Summary of main findings

In this section, the main findings of the dissertation will be summarized. A more detailed explanation of the results can be found in the specific chapters of this dissertation.

The clinical effectiveness study (**Chapter 3**) suggested that within our power ranges and observation period, CT and IPT do not differ in the treatment of depression in the acute phase beyond. Both treatments exceeded response in the WLC condition and led to considerable improvement in depression severity, quality of life and social functioning. No differential effects among therapists were found. Patients with severe depression did not seem to benefit more from either CT or IPT. Improvements that were noted at treatment termination persisted at least until 5 months follow-up. This might suggest lasting benefits instead of temporarily symptom relief. However, an exploration of longer-term follow-up effects is needed to see whether this is actually the case.

Examination of differences in explicit and implicit measures of self-esteem between depressed patients and healthy controls (**Chapter 4**) indicated that explicit self-esteem in patients was significantly lower as compared to healthy controls, but that patients and controls reported similar (positive) implicit associations with the self. Further-

more, our results suggested that only the explicit measure of self-esteem, and not the implicit, was related to depression. Thus, in spite of our attempt to create larger contrasts between patients and controls by adapting our implicit measure specifically to measure depressive core beliefs, we did not detect a specific effect of depression on the implicit measure of self-esteem. Even though it would be tempting to interpret these findings as an indication that beliefs about the self in depression are only affected on a conscious – explicit – level instead of on a deeper rooted implicit level, it seems to be too early to translate findings obtained with implicit measures into attitudes and beliefs. Procedures such as the SC-IAT are relatively new, and it is not clear yet in what sense the obtained measurements outcomes can be considered implicit and valid indices of self-esteem. It was therefore decided to not further include our implicit measure as a proxy of an underlying self-related schema.

The theoretical review of the body of literature on psychological mediators in psychotherapy for depression (**Chapter 5**) illustrates the gap between the theoretical and empirical status of research on mechanisms. In spite of increased attention for mechanisms of psychotherapy and theoretical consensus about necessities for proper mediation research, research in this field is limited, heterogeneous, and only a few studies meet the highest standards for tests of treatment mediation. Probably the biggest challenge in research aimed at identifying mediators that became apparent in this review is modelling the temporal relation between change in mediator and change in depression severity.

The empirical study on mediators (**Chapter 6**) addressed this issue by evaluating both mediation and temporal relations in one model. Over the course of treatment, patients' scores on all investigated process measures (dysfunctional attitudes, interpersonal problems, rumination, self-esteem and the quality of the therapeutic alliance) changed for the better, with medium to large effect sizes. Yet, change in process measures was smaller than change in symptoms. Similar to results on the clinical outcome measures, no differential effects in pre- to post-treatment changes were observed between the two conditions. However, change in interpersonal functioning occurred more rapidly in IPT than in CT. Exploration of the temporal relationships between change in five candidate working mechanisms and change in depressive symptoms indicated that change in process variables and concurrent change in depression were strongly related. Except for a significant relation between early change in self-esteem and subsequent change in depression, no significant temporal relations were found. Furthermore, there was no evidence of mediation (in either direction) in the sense that change in purported mechanisms did not mediate change in subsequent depression and early change in depression did not mediate subsequent change in the purported mediators. We did however find a concurrent mediational effect of interpersonal functioning in both the early (0–3 months) and late phase (3–7 months) of treatment. Again, no differential effects between treatments were observed.

In spite of similar pre- to post-treatment changes on all clinical outcomes and process measures in CT and IPT, results of the Sudden Gain study (**Chapter 7**) showed that there were significantly more patients with sudden gains in CT (42.2%) compared to IPT (24.5%). This might reflect a different pattern through which change is brought about. In both treatments, sudden gains were predicted by baseline quality of life score and absence of axis-I comorbidity. Furthermore, those with sudden gains reported lower levels of depression severity at post-treatment and 5 months FU – regardless of the duration of the between-session interval.

Exploration of the dynamic relations between individual symptoms of depression over the course of treatment (**Chapter 8**), indicated that all symptoms measured with the BDI-II were directly or indirectly connected to each other. The strongest connections were positive, indicating that improvement in one symptom was related to improvement in the other – and the vice versa for deterioration. CT and IPT showed similar networks that could be subdivided into two clusters: a cognitive and a somatic-affective cluster. These clusters correspond to Beck's two-factor structure and the biomedical model. Centrality analysis indicated that the symptoms 'Suicidal Ideation', 'Loss of Pleasure' and 'Feelings of Failure' were the most central in the network and were important for further symptom spread. 'Sadness' and 'Loss of Interest' were less central in the model. Thus, our results call into question the DSM classification of depression which emphasizes the importance of sadness and loss of interest as hallmarks of depression.

General discussion

Reflecting on the observed treatment effects

Findings from the various studies presented in this dissertation suggest that the treatments that were provided in our trial had beneficial effects for many depressed patients across a variety of outcome measures. The following paragraphs will further reflect on these findings.

Clinical impact of reduction in depressive symptomatology

Analysis on our main outcome BDI-II (chapter 3) indicated that the majority of patients responded to treatment. Pre- to post-treatment improvement effect sizes (ESs) were found to be large, and within range of those of the other randomized comparisons of CT and IPT (Elkin, Shea, Watkins, & Imber, 1989; Luty et al., 2007; Quilty, McBride, & Bagby, 2008). However, as reported by e.g. Cuijpers et al. (2014) standardized effect sizes provide only limited clinical information. In order to learn more about the clinical relevance of our findings it is important to take a closer look at the absolute (reduction in) BDI-II score and remission rates at the end of the treatment phase (7 months), and reflect on our findings in terms of the clinical impact for patients.

Over the course of treatment, the average BDI-II score in the active conditions declined from 29 to 14, reflecting an overall reduction of 15 BDI-II points (50%). A search through the literature indicates that a reduction of 50% and post-treatment score of about 15 on the BDI-II is not uncommon in the field of psychotherapy research for depression. Cuijpers et al. (2014) examined the effects across 21 CT studies and 3 IPT trials in terms of absolute change on the BDI-II, and found a mean post-treatment score of 15.3 BDI-II points for CT and 17.64 in the IPT conditions, corresponding with an average reduction in symptoms of 45.7% and 44.5% for CT and IPT respectively.

When looking at our results in terms of categories of depression severity (Beck, Steer, & Brown, 1996), it can be concluded that patients in our study on average improved from being borderline severely depressed (BDI-II ≥ 29) at baseline, to a depression that can be categorized as mild (BDI-II between 14 and 19) at the end of the treatment phase. As outlined in the DSM-IV, patients with mild depression only meet the minimal criteria to make the diagnosis. Clinically this suggests that even though day to day functioning may still be hard work and might feel like a struggle, the depressive symptoms are manageable and result in only minor impairments in social and occupational functioning. Many patients are able to resume work and participate in social activities. The average improvement to a mild depression level at end of treatment is an important improvement compared to severe depression in which the intensity of symptoms is seriously distressing, and the depression markedly interferes with social and occupational functioning.

Moreover, 34% of the patient met criteria for remission (BDI-II < 10) at the end of the treatment phase, indicating that they only experienced minimal symptomatology. A recent meta-analysis that examined remission rates specifically for CT and IPT (Cuijpers et al., 2014) found that – depending on the cut-off for remission – 42 to 49 % of the patients in CT were in remission at the end of the treatment phase. For IPT this was 38 – 44% in IPT. Compared to these findings, the results of our study are substantially lower. It should be noted that these remission rates are all based on scores on the Hamilton rating scale for Depression (HRSD) – a clinician rated scale to assess the severity of depression–and not on research using the BDI-II. Cuijpers and colleagues (2010) have shown that clinician-rated measures often show larger effects as compared to self-report. However, also when compared to the few studies that report remission rates in terms of BDI-II changes (e.g. Elkin et al., 1989 (49% CT, 56% IPT) Luty et al., 2007 (31% CT, 43% IPT)), our study has smaller remission rates. This is remarkable especially because our treatment phase was somewhat longer than in other trials (7 months *vs.* 4–5 months in other studies), and the average number of sessions was relatively high (17 sessions *vs.* 13 in Luty and Elkin), giving patients more time and opportunities to recover. A potential explanation for this might be the fact that we did not specifically instruct our therapists to continue treatment until patients met criteria for remission in terms of the BDI-II (as is often done in clinical studies). Instead, decisions about treatment termination were based on clinical relevance, without taking BDI-II score into account. Therefore, it could be that patient and therapist agreed on termination because

the depression was in remission from a clinical perspective, and patient and client were satisfied with the progress in treatment, but that the patient did not meet criteria for remission in terms of BDI-II score.

Even though the treatments that were provided in our study led to considerable and clinically relevant improvements in depressive symptoms, the majority of patients still experienced (some extent of) residual symptoms. Since incomplete recovery has been associated with increased vulnerability for relapse e.g. (Paykel, 2008) this is an important point of attention for further research. Our long-term follow-up data (24 months) should shed more light on the relation between residual symptoms at post-treatment and relapse and recurrence on the long term.

CT and IPT tap into more than just depressive symptoms alone

The studies presented in this dissertation furthermore suggest that patients treated with CT and IPT experience changes in other areas than depressive symptoms alone. Over the course of treatment, patients reported considerable improvement in quality of life and social functioning (chapter 3). These are important additional dimensions in the evaluation of treatments, because they belong to a broader definition of psychological health than just the absence of disease. Also from a patient perspective it might be relevant to evaluate treatment effects in terms of mental, physical and social well-being rather than on symptomatic improvement alone. For example, it might be more important for patients to resume work, engage in social interactions, and improve home management skills – potentially even while experiencing some symptoms – than to have a low score on the BDI-II.

Additional to changes in quality of life and social functioning, significant changes were observed on the various process measures that were examined. As reported in chapter 6, patients' scores on all investigated process measures changed for the better over the course of treatment. This suggests that processes that have been associated with the onset, maintenance and recovery of depression, do indeed change over the course of CT and IPT. However, there are two comments that should be made here. First, the pre- to post-treatment effect sizes for individual process measures were not as large as those of depressive symptom severity. This suggests that change mainly occurs on the level of symptoms. Even though this is to be expected from symptom-oriented treatments such as CT and IPT, it might point towards another explanation for why relapse rates in depression are so high. Perhaps long-term effects can only be established if patients not only meet criteria for remission on clinical outcomes, but also end up within the non-clinical range on measures of underlying processes associated with depression. Unfortunately norm-scores for healthy controls are not always available. It is therefore important to gain more insight in 'healthy' scores on processes associated with depression, and to examine the relationship between post-treatment scores on theorized mechanisms and future relapse more closely. If it turns out that scores on these process measures that fall within the clinical range are indeed associated with relapse (or more

9

than that turn out to be stronger predictors for relapse than the post-treatment depression severity score), it would be extremely valuable to examine whether a more intense treatment (e.g. treatment continued until patients reach the non-clinical cut-off) would lead to better results on the long-term.

Second, a closer look at the shape of change in process measures indicated that the total treatment effect (change from pre- to post-treatment) was mainly a result of changes in the second phase of treatment (3–7 months). Especially the process measures that were included to assess the specific theorized mechanisms of CT and IPT showed only little change in the initial phase[1]. In a way it is not surprising that change in hypothesized processes occurs later in therapy; change requires time and practice. Furthermore, it is likely that interventions that characterize the early phase of treatment – such as getting to know the patient, and explaining the rationale – yield initial symptom relief, without immediatly leading to changes in processes associated with depression. However, our findings do argue against the idea that change in process measures is a necessity for symptom change.

Taken together, our findings suggest that even though CT and IPT are symptom oriented treatments, patients experience changes in other areas as well. However, change in process measures was found to be considerably smaller than change in symptoms, especially in the first phase of treatment. This speaks against the idea that change in process measures is a necessity for symptom change. Furthermore, since the processes that were assessed in this study have been associated with the vulnerability for depression, post-treatment scores on process measures that fall within the clinical range might increase chances for future relapse.

Lack of differences between CT and IPT

Another finding that deserves further discussion here is the fact that we found no differences between CT and IPT in pre- to post treatment change on any of the outcome measures that were examined throughout the chapters of this dissertation. The finding that CT and IPT yield similar clinical effects is in line with our expectation, with findings of the other randomized comparisons of individual CT and IPT (Elkin et al., 1989; Luty et al., 2007; Quilty et al., 2008), and with meta-analyses comparing the effects of CT, IPT and various other forms of psychotherapy for depression (e.g. Cuijpers, Andersson, Donker, & van Straten, 2011).

Even though the finding that patients treated with CT and IPT exhibited similar improvement on all process measures – regardless of whether a process was actively targeted during treatment – is also in line with previous research suggesting that change in theorized processes is not necessarily specific for one type of treatment (see chapter 5), it calls for interpretation. To begin with, our findings could indicate that the treatments that were provided in our study were not administered according to protocol, and therefore cannot be considered separate treatments. If treatments are indistinguishable, it is not surprising that they show similar changes. However, the results of our integrity

check (chapter 3) make this explanation rather unlikely. Second, our findings could indicate that the various specific interventions that are used in CT and IPT tap into the same underlying processes. In other words, patients learn similar things, but in a different language. For example, in a way, the sick role (a specific IPT intervention) can be seen as an example of reframing dysfunctional thoughts as the product of a disease rather than as a reflection of reality, which is a core feature of CT. Even though in CT the origin of the task is in the automaticity of these processes and not in the content, it requires the same cognitive structure for patients. Similarly, cognitive interventions and behavioural experiments in CT often specifically focus on beliefs in the interpersonal domain (e.g. 'I am inferior' or 'I will be rejected in social situations'), thereby directly acting on change in interpersonal functioning.

Alternatively, it might be the case that CT and IPT initially tap into different aspects of depression, but that change in one process causes change in another, because all processes are related to each other. The network study presented in chapter 8 provides some empirical evidence for this interconnectedness by demonstrating that all symtoms of depression that are assesed with the BDI-II are directly or indirectly connected to each other, and that change in one symptom likely causes change in another. Even though the BDI-II has a strong cognitive focus, and items referring to e.g. the interpersonal domain are lacking, there are good reasons to believe that this interconnectedness is also present between several other domains that were assessed in this study. For example, the specific experiences following changes in interpersonal functioning in IPT might lead to re-adjustment of schema's and attitudes about the self, the world and the future. Likewise, decrease of dysfunctional attitudes in CT might also lead to change in beliefs related to interpersonal relationships, making patients more likely to improve their interpersonal functioning as well. If this is the case, the overall change from pre – to post-treatment might be similar, but the pathway through which change is brought about is different. It should be noted though that change does not necessarily need to start with change in specific factors. Perhaps factors that are common for all types of therapy (such as motivation and working alliance) are the initial reason for change, and lead to change in other processes as well. However, in this scenario, CT and IPT would be expected to show relatively similar patterns of change, especially in the early phase of treatment. Unfortunately, because we only assessed our process measures three times over the course of treatment, we were not able to examine the exact shape of change in process measures, and compare patterns between CT and IPT. However, the idea of differential pathways is supported on the symptom level by the findings of our sudden gain study (chapter 7) which shows that there are significantly more sudden, large and stable symptom drops in CT as compared to IPT. To the extent that symptom change is driven by specific mechanisms, this might indeed reflect a different pattern through which change is brought about.

A fourth explanation for the lack of differences between CT and IPT might be related to the specific methods that were used to analyse change. For example, our analyses were based on group means. It might be the case that specific (groups

of) patients respond differently, but that this effect is cancelled out when taking the group mean into account. Furthermore, we mainly used total scores to define the amount of change on process measures. Even though exploratory subscale analyses did not show differences between conditions either, it could be possible that difference emerge on a more detailed level. In order to increase our knowledge on how CT and IPT compare, future studies should include more frequent repeated measures of specific and common processes over the course of treatment, and analyse the data on subgroup- and subscale level as well.

Limited evidence for the theoretical models of change

As described in the general introduction of this dissertation, one of our main study aims was to shed light on the extent to which CT and IPT work according to their respective theoretical models. In order to do this we examined change in various (specific and non-specific) theorized processes of CT and IPT over the course of treatment, and explored the direct and indirect (temporal) relationships between changes in these potential mechanisms and change in depressive symptoms. Analyses showed that patients' scores on the mechanism measures (assessing dysfunctional attitudes, interpersonal problems, rumination, self-esteem and therapeutic alliance) showed considerable improvement over the course of treatment. Furthermore, changes in depressive symptoms and changes in hypothesized mechanisms were highly correlated, indicating that these processes are indeed associated with one another. However, in spite of a temporal research design, careful selection of processes and measurement instruments, and the use of various innovative (statistical) approaches, we only found limited evidence for temporal relations and mediational effects. We were therefore not able to demonstrate firm empirical evidence for the cognitive model or for the interpersonal theory.

Several explanations for our findings have been discussed in chapter 6. They can be roughly divided into four categories. First, our findings might imply that change that was observed over the course of treatment was a result of the natural course of depression, and was therefore unrelated to the fact that patients were treated. However, – as reported in chapter 3 – CT and IPT exceeded response of an untreated control condition after two months, making this explanation not very likely. Second, our result could indicate that theories are incorrect, and that other mechanisms – not assessed in this study – are responsible for therapeutic change. Third, it might be the case that the treatments provided in our study were not powerful or specific enough to mobilize mechanisms. Fourth, it is possible that processes that were assessed in this study actually do play a role, but that we were unable to demonstrate their role as a mechanism because of flaws in our design and quality of measurement instruments that were used. Though speculative, some explanations seem more plausible than others, and deserve further discussion.

Timing of assessments

A first explanation for our findings that will be highlighted here is the timing of assessments. As pointed out by e.g. Laurenceau, Hayes, and Feldman (2007), and Collins and Graham (2002), the temporal design in a longitudinal study can have important effects on results of the study, and the conclusions drawn from the study. More specifically, the assessment points in our study (baseline, 3, and 7 months) might have been spaced too far apart to capture the timing in which change took place. If observations are spaced too far apart, it is impossible to observe the temporal relation between process and outcome, because by the time the assessment takes place change in both process and effect has already occurred, erroneously leading to the conclusion that temporal relations between processes and outcomes (or differential effects between treatments) do not exist. A more fine-grained analysis of change over time could have been helpful here. However, the selection of time span and number of assessments is a complicated matter. How can one determine the expected speed and shape of change before actually measuring it (chicken or egg)? And what is the best definition of change? Is this the moment that patients actually change their thoughts or interpersonal functioning, or should this be the point when patients allow themselves to doubt their existing views and tolerate schema incongruent information? Apart from the drawbacks of spacing assessments too far apart, it is also undesirable to have them spaced too close together. Not only because this can be a burden for participants, but also because of the risk of measurement artefacts when making too many demands for data (Longwell & Truax, 2005).

It might be clear that it is difficult to find a good balance between the optimal study design, the burden for patients, and risk of measurement artefacts. Appropriate timing of assessments therefore remains a challenge for future research. Experience Sampling Methods (ESM) might be promising in this regard. By designing studies in which patients are asked to frequently answer short questions about their mood and several psychological processes that are assumed to be related to treatment, and only asking them to fill out more detailed questionnaires when they show large changes and/or meet certain thresholds, detailed information about the shape of change can be obtained while minimalizing the burden for patients and the risk of measurement artefacts.

However, even with these fine-grained measurements, psychotherapy is a complex multifaceted phenomenon that might be too complex to be captured in relatively simple causal research designs. If therapeutic change indeed occurs suddenly (as suggested by the presence of sudden gains, see chapter 6), rather than gradually over the course of treatment, it might be very difficult to capture the critical moment of change, which will differ between patients. Furthermore, it might be even more difficult to assess the temporal relation between change in the mechanism and change in symptoms. Especially from the perspective of a network approach (chapter 8), it is conceivable that change in one node affects the other nodes in such short notice that it is virtually impossible to assess the spread in time. From this point of view there is a chance that the fine graininess of our analysis might never suffice to identify the critical moment in which change takes place.

9

The potential influence of behavioural change

Another potential explanation for the limited evidence on mechanisms is the possibility that therapeutic change is caused by factors that were not assessed in our study. One example of a promising potential mechanism that is addressed in treatment but has not been assessed in our trial is (change in) activity level. There are good reasons to believe that behavioural activation (BA) and positive reinforcement might be strong components in the process of therapeutic change. Several studies that examined the effects of BA indicate that activation might be as effective in reducing depression as the full package of CBT (Cuijpers, van Straten, & Warmerdam, 2007; Ekers, Richards, & Gilbody, 2008; Jacobson et al., 1996). A study by Dimidjian et al. (2006) even found that BA outperformed CBT, especially in patients who were more severely depressed.

The potential of behavioural change as a factor that might facilitate change is further supported by preliminary findings of an ongoing empirical study in which we examine processes related to sudden gains (Lemmens, DeRubeis, Tang, & Huibers, in preparation). In this study we aim to identify the processes that elicit and explain sudden gains by meticulously analysing the content of the sessions preceding and following the gain. Changes are rated in three domains: cognitive change, interpersonal change, and behavioural change. Even though we only found relatively little change in the pre-grain sessions (raters were instructed to only rate progress if this was explicitly acknowledged by the client, and were not allowed to infer), the change that was observed was mainly found in the behavioural domain. For example, patients firmly decided on increasing pleasant activities in the session preceding the gain, and/or made specific plans on how to do so. Even though these findings should be interpreted with caution because they are based on preliminary analyses in a small sample, they underline the potential of behaviour as a process related to symptom change. It would therefore be beneficial for future trials to also include behavioural measures (objective and subjective) to examine the extent to which these behavioural changes are responsible for therapeutic change.

Was treatment powerful enough?

The third and final possibility that will be briefly brought into focus here is the possibility that – even though our treatments led to considerable and clinically relevant improvements in depressive symptoms (see earlier in this chapter) – they were not powerful enough or were delivered with insufficient specificity to mobilize mechanisms. As reported in chapter 6, effect sizes for (individual) process measures were smaller than for the BDI-II. This already indicates that our mechanism measures could only explain a limited amount of the treatment effects. Furthermore, the fact that there was only little change in the initial phase of treatment, may have limited the possibilities for statistical analysis of the temporal relations. In order for change in a potential mechanism measure

to account for subsequent change in depression severity, there needs to be substantial change over the period of observation. Unfortunately, the average amount of change in process measures in the early phase of treatment was relatively small, leaving little change from which to predict (no change in the mediator means no mediation).

Our findings suggest that patients might benefit from more intense treatment, especially in the initial phase. The question is how to do this without changing the specific content and features of CT and IPT too drastically? One answer might be found in the area of session frequency. In the Netherlands, psychotherapy is provided once a week. However, there are indications that higher session frequency in the initial phase of treatment (e.g. two times per week) might lead to larger treatment effects (Cuijpers, Huibers, Daniel Ebert, Koole, & Andersson, 2013). The idea behind this is that higher frequency might improve patient's recall of the previous session, hereby enhancing the process of learning. Recently, several members of our research group started a large multicentre RCT to examine this claim more closely in CT and IPT for depression (Freq-Mech trial, Bruijniks et al., submitted). Another option might be to continue treatment to the maximum of 20 sessions as long as scores on the hypothesized process measures are within the clinical range – even though patients might be (close to) remission in terms of clinical outcomes.

Identifying mechanisms of change: trying to find a needle in a haystack?

After almost three decades of process research in psychotherapy for depression, we have to conclude that in spite of increased attention for mechanisms, advances in theoretical consensus about necessities for this type of research, and progress in the degree of sophistication that researchers bring to research on mediators, there is no clear-cut empirical explanation for psychotherapeutic change. This calls for taking a step back to evaluate whether the research designs that are (and will be) used in the field are actually suitable to answer questions like this. More specifically, one could wonder whether it is realistic to try to explain therapeutic change in terms of relatively simple causal research designs. Psychotherapy is a complex, multi-dimensional phenomenon that might work through interplay of multiple mechanisms on several levels (physiological, affective, behavioural and cognitive). Psychotherapeutic change might therefore consist of a complicated chain of events on these different levels. Furthermore, the relationships between mechanisms and outcome might differ at various points in time and between specific (subgroups of) patients. From this perspective, trying to explain psychotherapeutic change in terms of psychological processes that are assessed at limited points across treatment and analysed on group-level seems like looking for a needle in a haystack. Although discouraging, it makes it a lot easier to understand why the progress in this field is so slow, and why existing research so far was not successful in providing clear-cut empirical explanations of how psychotherapy for depression works.

Generalisability of findings to other Dutch mental health clinics

Another important point to reflect on is the generalisability of findings into everyday clinical practice. One might question this because we selected our patients and therapists, and provided treatments under controlled circumstances (training, supervision, adherence checks). From this point of view one might even argue that our trial is merely an efficacy trial rather than an effectiveness study. Even though the treatment provided in our trial is more controlled as compared to average Dutch mental health care, there are several reasons for why we consider our trial an effectiveness study and think that the generalizability of our findings to depressed patients and clinical service in the Netherlands is high. To start with, our study is conducted in a research-oriented routine clinical practice setting (contrary to highly specialized university treatment studies in efficacy trials). Second, the patients that were selected for the study came from the population that was already referred to RIAGG. With this, the population reflects the total population of people that seek psychological help on their own accord (contrary to highly selected patient samples in efficacy trials). However, the exclusion of patients receiving concomitant pharmacological treatment has indeed pushed us closer towards the efficacy trial definition. Furthermore, we selected experienced therapists that were already working in our clinical, and gave them additional training and supervision. Therapists delivered study therapies as part of their regular duties in the centre, where they had a mixed task load, and had to see patients with different types of problems. In other words they were not super specialist in one treatment approach for one type of patient (contrary to efficacy studies in which treatment was provided by highly experienced therapists). Furthermore, we did not require therapist to first successfully treat 5 consecutive patients under supervision before they could participate in the trial (as is often done in efficacy trials). In addition, the protocol allowed for flexible session scheduling, and treatment could be terminated when patient and therapist were satisfied about treatment improvement. Hereby the duration of our treatment phase reflected the clinical setting (contrary to efficacy studies in which treatment is provided within a very strict timeframe). Overall, the mental health centre where the study was carried out was not that different from the average Dutch clinics as one might initially expect. We therefore think that findings can be translated to other practices in the Netherlands, especially in clinics in which refresher's courses and regular consultation sessions are already common practice.

Methodological considerations

Strengths

Even though our study is the fourth randomized comparison of individual CT vs. IPT for depression, it made an important contribution to the field. Our trial was the first to add a WLC condition to the comparison and to examine delayed effects up to 5 months

after treatment termination. Furthermore, with a sample size of $n=151$ in the active groups, our study provided important information necessary to support the conclusion that the effects of individual CT and IPT for adult depression in the acute phase do not seem to differ within a 4 BDI points limit. In addition, we did not only evaluate our treatment effects in terms of symptomatic improvement, but also assessed effectiveness in several other domains including quality of life, and impairment in social functioning. As outlined by e.g. Kennedy, Eisfeld, and Cooke (2001) and Jakobsen, Hansen, Simonsen, Simonsen, and Gluud (2012) measures like this provide an important additional dimension to the evaluation of treatments for depression. With this, our study builds on previous research findings obtained in the US and New-Zealand with a large European sample, hereby extending our knowledge about the relative effectiveness of these treatments around the world. The vast majority of patients completed treatment and quality of therapy was rated by independent assessors as being (very) good to excellent in both conditions. Moreover, with 89% of patients providing data at post-treatment, and 85% at the 12-month assessment, attrition rates were low. Another major strength of our trial is that it is one of the few studies in the field that included multiple measures of various process measures throughout treatment. The repeated measures design provided the opportunity to evaluate whether change in proposed process measures preceded, followed from, or went together with changes in depression. Furthermore, we examined mechanisms from various perspectives and used up-to-date statistical analyses techniques. By obtaining measures of depression severity from session to session, we were able to examine sudden gains and the dynamic network of the BDI-II in the context of this trial.

Limitations

Our trial is not without limitations. Several features of our study design that might have influenced the (impact of) findings across studies presented in this dissertation will be discussed here.

First of all, our study was powered to detect differences on our primary outcome BDI-II, and not on secondary outcomes and process measures. This might have limited the power of other analyses, such as e.g. moderation analysis in the effectiveness study (chapter 3) and might have contributed to the limited evidence regarding the underlying theoretical models in chapter 6. Furthermore, one could argue that our WLC condition is a limited control because the duration of the waiting-list was shorter than the treatment time. In retrospect, a full WLC might indeed have been helpful in the various studies presented in this dissertation. Not only as a control for the clinical effectiveness study (chapter 3), or regarding the occurrence of sudden gains (chapter 7), but also with regard to the mediation study (chapter 6). As has been outlined in chapter 5, the possibilities for establishing statistical mediation strongly depend on the extent to which two interventions differ. Since we compared two ac-

tive treatments that did not show differences on outcome and process measures, possibilities for examining statistical mediation were limited. It is possible that if we were able to compare each of the active treatments with a full WLC condition, we might have had more options to demonstrate the mediating role of the hypothesized variables. However, given the study population (patients with major psychopathology who already applied for treatment on their own accord), and the distress and risks related to their particular pathology (depression), we found it unethical and inappropriate to include a full WLC condition. In line with this, including a full WLC condition would also lead to practical problems, since patients would be less willing to participate in the trial if there was a chance that they would be allocated to a 7-month waiting-list. In order to create a balance between methodology and ethics, we decided to include a WLC condition that was as short as possible. We expected the differences between active psychological treatment – if as effective as we estimated based on previous findings – and WLC to be strong enough after two months to be detected. We reasoned that if there was no difference after 8 weeks of treatment, this was not likely to happen later either. This is not without precedent in the depression treatment literature; both DeRubeis et al. (2005) and Dimidjian et al. (2006) kept patients on pill-placebo for only eight weeks (long enough to detect a drug-placebo difference) while keeping patients in cognitive therapy or medications for a full 16 weeks (long enough to maximize clinical response).

Another important limitation of our study is that our design does not allow us to draw firm conclusion about causality, because the mediators that were assessed in our study were not experimentally manipulated. As has been outlined in chapter 5, this is necessary to make causal inferences about change. In our case, we did not get to that step, because we did not find evidence for mediation in the first place, but if we would have, this would have been an important limitation. However, as outlined in chapter 5, experiments that manipulate the proposed mediator or mechanism and show the impact on outcome are limited, because implementing experimental studies in this type of research is very difficult, both from a practical as well as an ethical point of view.

Several limitations regarding the timing and spacing of observations should be mentioned as well. First of all, there was no direct assessment at 2-months for the active groups CT and IPT. In order to obtain a measure of depression severity comparable to the 2-month BDI-II assessment of the WLC condition, we selected the BDI-II score that was obtained in the session 2 months after the start of treatment. In addition, since the primary endpoint of our study (7 months) did not necessarily equate treatment completion in the CT and IPT arms, this cannot be considered a full post-treatment score. However, as explained in chapter 3, the proportion of patients that were still in treatment at that point was relatively small, and on average, these patients received only one or two additional sessions. Therefore, we do not think that this has influenced our findings. However, future studies should consider administering a full-test battery at treatment

termination as well. A similar line of reasoning is appropriate for the 3-month assessment as a proxy for mid-treatment. Moreover, even though we carefully selected process- and outcome measures it is possible that our instruments were not sensitive enough to assess (changes in) constructs of interest. To conclude, recruitment of patients took much more time and effort than we had estimated. This phenomenon is known as "Lasagna's Law" and seems to be inherent to running an RCT (Gorringe, 1970).

Implications and recommendations

Clinical implications

Several clinical implications can be drawn from the findings presented in this dissertation. First of all, trial findings add to the accumulating evidence that the recommended practice for the psychological treatment of depression in the Netherlands is effective. CT and IPT should therefore continue to be considered feasible treatments for MDD. Furthermore, the fact that we did not find evidence for a moderating effect of baseline severity suggests that both CT and IPT can be offered across the full range of severity, without special preference for patients with severe depression. Symptom centrality analysis provided empirical evidence for the fact that several symptoms of depression seem to be more influential in the treatment of depression than others. If future research is able to replicate these findings and gain more insight in centrality of symptoms, this may help in determining treatment focus within the existing CT and IPT protocols. Moreover, it might even lead to development of person-specific centrality analyses in the beginning of treatment, which can be used to indicate which symptoms should be specifically targeted in a particular patient. Our findings furthermore suggest that patients might benefit from a more intense treatment – especially in the early phase of treatment. This underlines the need for further research on effects of multiple sessions per week. In addition, even though depressive symptomatology should remain the main outcome in determining treatment success – after all treatments such as CT and IPT are symptom-oriented – it would be valuable to also take improvement on processes associated with depression into account when evaluating treatment effects. This is especially relevant from the viewpoint that unhealthy scores on these theorized processes might be related to relapse and recurrence. In line with this, it would be fruitful to further examine whether there is a dose-effect relation between the number of sessions and the extent to which patients show improvement on underlying processes. To conclude, the fact that, overall, patients in CT and IPT showed comparable results, both on clinical outcomes and process measures, calls for putting more weight to patient's preference in the process of treatment selection: at least as long as research to moderators and individual differences in processes of psychotherapeutic change is still in its infancy.

9

Recommendations for research on mechanisms

Although the studies presented in this dissertation were not able to provide clear-cut answers about mechanisms of change, they do provide guidance for future research. Table 1 gives an overview of important recommendations for future process research. In short, future studies should invest in the development and evaluation of mediator measures. In particular, fundamental research on the validity of implicit measures should progress. Furthermore, studies should include multiple measures of potential specific and non-specific process in well-planned temporal research designs paying special attention to the timing of assessments and within-patient variances. Apart from traditional designs to examine processes of change, alternative designs including experimental manipulations and component analyses should be considered as well. Furthermore, it is important that researchers use modern statistical methods for the analysis of change. In doing all of this it is important that researchers invest in a uniform research language, making it easier to compare results across studies and integrate findings into broader knowledge.

Table 1. Recommendations for future research aimed at disentangling mechanisms of change

Potential Mechanisms
– Use theory to select multiple specific and non-specific potential mechanisms.
– Include processes that would falsify the theory as well.
– Provide a clear description of each process that is included.
– Use mediator measures that have shown to be psychometrically valid.
– Invest in the evaluation and further development of (implicit) mediator measures.
– Use multiple sources of information (including self-report, clinician rated, independent rater, and behavioural and biological measurements).
Study Design
– RCT with a control group, preferably also including a non-active arm.
– Include a fine grained temporal design, especially in the early phase of treatment.
– Justify the timing and spacing of observations
– Invest in development of alternative research designs including experimental manipulations and component analyses.
Analyses
– Use modern statistical analysis methods to examine change over time and mediation.
– Focus on statistical significance but also on the clinical meaning of changes.
– Examine the unique influence of each mediator, as well as their interactions.
– Perform analysis on group level; but also examine subgroups, and individual trajectories.
Reporting
– Invest in 1 research language and standard guidelines for reporting mechanisms.
– Replicate studies and publish negative data as well.

Recommendations for funders and policy makers

In a time in which (research on) mental health care has to deal with frequent budget cuts, while the prevalence of mental disorders – and depression in particular – is still increasing, it is also important to speak to funders and policy makers. Our main recommendation here is to keep prioritizing and funding research aimed at evaluating and improving existing psychological treatments for depression, and in particular research on CT and IPT. Unfortunately, the funding climate in the Netherlands has changed, and more emphasis is put on what is considered "innovative" (neuro) science. Research into existing (psychological) treatments is harder and harder to get funded. This is remarkable, because it is the existing treatments such as CT and IPT that are implemented on a large scale in the Dutch health care system. More specifically, a recent study by Statistics Netherlands (Centraal Bureau voor Statistiek, CBS), indicates that over 100.000 Dutch citizens receive some form of treatment for depression in secondary care. Given the fact that CT and IPT are the psychological treatments of choice for depression in secondary care, one can imagine that optimizing these treatments is not only important from a patient's point of view, but might also lead to a significant reduction in burden of disease from a societal and economical perspective. In our view it is therefore extremely relevant to further examine the remaining questions related to the effects and mechanisms of CT and IPT.

Next steps

The studies presented in this dissertation made a significant contribution to the current state of knowledge on the clinical effectiveness of CT vs. IPT, and provided useful information for the developing research field of mechanisms of change. Nevertheless, several aspects of our trial fell out the scope of the current dissertation and will be addressed in future publications.

The first step after comparing clinical effectiveness in the acute phase, is examining whether one treatment is be superior to the other with regard to relapse prevention. As mentioned in the general introduction (chapter 1) and in our design paper (chapter 2), we gathered follow-up data up to 24 months after randomization. We are currently preparing a study in which we examine the course of depressive symptoms across the Long-term Follow-Up Phase of our trial (12 to 24 months), and compare both treatments in terms of survival rates. The role of cognitive reactivity and attributional style – processes that have been associated with relapse and recurrence – will be examined as well.

In addition, in order to fully understand how CT and IPT compare it is important to consider clinical effectiveness in combination with economic evaluation. However, to date no study has directly compared both treatments in terms of cost-effectiveness. This is remarkable because especially when two treatments turn out to be equally effective clinically, it is important to know whether one treatment is more cost-effective.

We are currently planning the first study to fill in this knowledge gap. Using the data on costs and utilities that were collected parallel to the data on clinical effects and mechanisms of change (see Table 1 in chapter 2 for an overview), we will examine whether one treatment outperforms the other in terms of cost-effectiveness from a societal perspective in the active phase of treatment (baseline–7 months), and over the course of the 17-month Follow-Up Phase (7–24 months).

To conclude, it is still largely unknown which treatment works for whom. Future research should therefore also focus on a further development of personalized medicine – the customization of health care to the individual patient. In order to determine the best treatment for a specific patient, it is important to identify (patient) characteristics that predict differential treatment response. Recently, researchers at the University of Pennsylvania (US) developed a decision-making model that compares and weights multiple variables in order to predict the optimal choice of treatment (DeRubeis et al., 2014). By combining data on treatment success and patient characteristics the model generates a score indicating which treatment is likely to be more effective for a given patient: the personalized advantage index (PAI). They tested the model using data from a longitudinal study in which 154 patients were treated with either CBT or antidepressants. Within this context effects were found that – if applied to assign patients to their optimal treatments – would rival those of an effective treatment relative to a non-active control. In order to determine whether this method is also useful in treatment selection of CT vs. IPT, we are currently collaborating on a project in which we test the treatment selection model using the data of our own trial (Huibers et al., in preparation).

In conclusion

A full understanding of the effects and mechanisms of psychological interventions for depression might be the best short- and long-term investment for improving everyday clinical practice and patient care. Although the interest for treatment evaluation has grown, and knowledge about psychotherapy has increased tremendously over de past few decades, treatments are not yet fully understood. The present dissertation aimed to contribute to this research field by examining the clinical effectiveness and various aspects of mechanisms of change of individual CT and IPT for adult depression in the context of large RCT. Trial findings add to the accumulating evidence that the recommended practice for the psychological treatment of MDD in the Netherlands is effective. However, in spite of a well-considered temporal research design, careful selection of processes and measurement instruments, and the use of various innovative (statistical) approaches, the studies presented in this dissertation were not able to provide clear cut empirical evidence for the underlying mechanisms of change. Psychotherapy is a multifaceted phenomenon that might work through interplay of multiple mechanisms at several levels (i.e. physiological, affective, behavioural and cognitive). Furthermore, the rela-

tionships between mechanisms and outcome might differ at various points in time, and between specific (subgroups) of patients. Psychotherapeutic change might therefore be too complex to be explained in relatively simple causal research designs. However, our studies do provide guidance for future research questions and further improvement of study designs. In short, the field would benefit from a further refinement of (experimental) research methods to disentangle mechanisms of change, advances in the area of personalized medicine, and from more insight in how CT and IPT compare with regard to cost-effectiveness and relapse prevention. Together, this can bring us closer to an optimal understanding of why, for how long and for whom CT and IPT for depression might work.

Footnotes

[1] Effect Size *d* from baseline to 3 months = 0.11 and 0.17 for change in dysfunctional attitudes and interpersonal functioning respectively.

References

Beck, A. T., Steer, R., & Brown, G. K. (1996). *Beck Depression Inventory II: Manual*. Boston, MA: Hartcourt Brace.

Bruijniks, S. J. E., Bosmans, J., Peeters, F., van Oppen, P., van den Boogaards, M., Dingemanse, P., . . . Huibers, M. J. H. (submitted). Frequency and change mechanisms of psychotherapy among depressed patients: a multicenter randomized trial. *Manuscript submitted for publication.*

Collins, L. M., & Graham, J. W. (2002). The effect of the timing and temporal spacing of observations in longitudinal studies of tobacco and other drug use: Temporal design considerations. *Drug and Alcohol Dependence, 68*, 85–96.

Cuijpers, P., Andersson, G., Donker, T., & van Straten, A. (2011). Psychological treatment of depression: Results of a series of meta-analyses. *Nordic Journal of Psychiatry, 65*(6), 354–364.

Cuijpers, P., Huibers, M., Daniel Ebert, D., Koole, S. L., & Andersson, G. (2013). How much psychotherapy is needed to treat depression? A metaregression analysis. *Journal of Affective Disorders, 149*(1–3), 1–13.

Cuijpers, P., Karyotaki, E., Weitz, E., Andersson, G., Hollon, S. D., & van Straten, A. (2014). The effects of psychotherapies for major depression in adults on remission, recovery and improvement: A meta-analysis. *Journal of Affective Disorders, 159*(0), 118–126.

Cuijpers, P., Li, J., Hofmann, S. G., & Andersson, G. (2010). Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: A meta-analysis. *Clinical Psychology Review, 30*, 768–778.

Cuijpers, P., van Straten, A., & Warmerdam, L. (2007). Behavioral activation treatment of depression: a meta-analysis. *Clinical Psychology Review, 27*, 318–326.

DeRubeis, R., J.,, Hollon, S., D.,, Amsterdam, J. D., Shelton, R. C., Young, P., R., Salomon, R., M., . . . Gallop, R. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry, 62*(4), 409–416.

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: Translating Research on Prediction into Individualized Treatment Recommendations. A Demonstration. *PLoS ONE, 9*(1), e83875.

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., . . . Jacobson, N. S. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology, 74*(4), 658–670.

Ekers, D., Richards, D., & Gilbody, S. (2008). A meta-analysis of randomized trials of behavioral treatment of depression. *Psychological Medicine, 38*, 611–623.

Elkin, I., Shea, M. T., Watkins, J. T., & Imber, S. D. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry, 46*(11), 971–982.

Gorringe, J. A. L. (1970). Initial preparation for clinical trials. In E. L. Harris & J. D. Fitzgerald (Eds.), *The principles and practice of clinical trials*. Edinburgh/London: Livingstone.

Huibers, M. J. H., Cohen, Z. D., Lemmens, L. H. J. M., Arntz, A., Peeters, F. P. M. L., Cuijpers, P., & DeRubeis, R. J. (in preparation). Predicting optimal outcomes in CT or IPT for individual depressed patients using the Personalized Advantage Index Approach. *Manuscript submitted for publication.*

Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., . . . Prince, S. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology, 64*, 295–304.

Jakobsen, J. C., Hansen, J. L., Simonsen, S., Simonsen, E., & Gluud, C. (2012). Effects of cognitive therapy versus interpersonal psychotherapy in patients with major depressive disorder: a systematic review of randomized clinical trials with meta-analyses and trial sequential analyses. *Psychological Medicine, 42*, 1343–1357.

Kennedy, S. H., Eisfeld, B. S., & Cooke, R. G. (2001). Quality of life: an important dimension in assessing the treatment of depression? *Journal of Psychiatry and Neuroscience, 26*, 23–28.

Laurenceau, J.-P., Hayes, A. M., & Feldman, G. C. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review, 27*(6), 682–695.

Lemmens, L. H. J. M., DeRubeis, R., J.,, Tang, T. Z., & Huibers, M. J. H. (in preparation). Processes related to sudden gains in Cognitive Therapy and Interpersonal Therapy for Depression: Examination of critical therapeutic events in the pre-gain sessions. *Manuscript in preparation*.

Longwell, B. T., & Truax, P. (2005). The Differential Effects of Weekly, Monthly, and Bimonthly Administrations of the Beck Depression Inventory-II: Psychometric Properties and Clinical Implications. *Behaviour Therapy, 36*, 265–275.

Luty, S. E., Carter, J. D., McKenzie, J. M., Rae, A. M., Frampton, C. M. A., Mulder, R. T., & Joyce, P. R. (2007). Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *British Journal of Psychiatry, 190*, 496–502.

Paykel, E. S. (2008). Partial remission, residual symptoms, and relapse in depression. *Dialogues in Clinical Neuroscience, 14*, 431–437.

Quilty, L. C., McBride, C., & Bagby, R. M. (2008). Evidence for the cognitive mediational model of cognitive behavioural therapy for depression. *Psychological Medicine, 38*, 1531–1541.

Spijker, J., Bockting, C. L. H., Meeuwissen, J. A. C., van Vliet, I. M., Emmelkamp, P. M. G., Hermens, M. L. M., & van Balkom, A. L. J. M. (2013). *Multidisciplinaire Richtlijn Depressie (Derde revisie): Richtlijn voor de diagnostiek, behandeling en begeleiding van volwassen patiënten met een depressieve stoornis.* Utrecht: Trimbos-instituut.

9

Valorization Addendum

The present dissertation describes the results of a large randomized controlled trial (RCT) that was conducted to gain more insight into the clinical effects and underlying mechanisms of Cognitive Therapy (CT) versus Interpersonal Psychotherapy (IPT) for Major Depressive Disorder (MDD). This Valorization Addendum reviews the societal, economic and scientific relevance of our trial, and gives an overview of the target groups for whom our research findings might be important. Furthermore, innovative aspects of our trial and possible activities and/or products resulting from the research that was presented in this dissertation will be discussed. To conclude, examples of how knowledge valorization has been achieved so far and ideas for further dissemination of knowledge in the field will be provided.

Relevance of our trial

With lifetime prevalence estimates ranging from 15 to 20%, MDD is one of the most common psychiatric disorders worldwide. According to the World Health Organization (WHO), globally more than 350 million people of all ages suffer from depression (WHO, 2012). Depression carries a tremendous burden for the patient as it has an enormous impact on physical, social and emotional functioning and well-being. In addition, the economic burden that depression imposes on society is very high. A cross-national epidemiological study comparing MDD lifetime prevalence rates worldwide showed that the highest rates were found in high-income countries including France, the USA and the Netherlands (Bromet et al., 2011). The latest screening of the Dutch population showed that over 800.000 adults (5.2% of the total Dutch population) suffer from MDD every year (de Graaf, ten Have, & van Dorsselaer, 2010). Recent estimations by the Dutch National Institute for Public Health and the Environment (RIVM) indicate that depression annually costs 168,000 Disability Adjusted Life Years (DALYs). This indicates that the negative effects of depression on health in the Netherlands are even larger than those of e.g. dementia, diabetes or lung cancer (RIVM, 2013). The total annual costs of depression in the Netherlands (direct and indirect) are estimated at almost three billion euro (de Graaf, Tuithof, van Dorsselaer, & ten Have, 2011; Slobbe, Smit, Groen, Poos, & Kommer, 2011).

Considering the individual, societal and economic impact of MDD, it is important that patients suffering from depression are adequately treated. The Dutch multidisciplinary guideline for the treatment of depression currently recommends CT and IPT as the psychological treatments of choice in the acute phase of treatment of depression. Even though CT and IPT have repeatedly shown to be well-standardized and efficacious interventions for the acute treatment of MDD, there is room for improvement. Approximately 40% of depressed patients do not (or insufficiently) respond to initial treatment, and even when treated effectively in the acute phase chances of relapse and recurrence are high (Keller & Boland, 1998; Paykel, 2008). The challenge in contemporary depression research is therefore to improve treatments to increase initial response rates and prevent relapse and recurrence in the long term.

Many researchers agree that treatment improvement starts with a full understanding of the effects and mechanisms of interventions. By knowing why and how treatment works, specific components of interventions can be added, strengthened or removed to make the interventions more efficient and (cost-)effective. Furthermore, information on how treatments compare – and whether this differs for certain subgroups of patients – can be used to select the best available treatment for (individual) patients, hereby also increasing therapy effects. Although our knowledge about psychotherapy for depression has increased tremendously over the past few decades, psychological treatments – including CT and IPT – are not yet fully understood. There are several unresolved issues that need further examination. For instance, it is not clear yet whether one therapy outperforms the other with regard to severity and course of the disorder, especially on the long-term. In addition, it is still largely unknown how CT and IPT work and – more specifically – whether they work for reasons hypothesized in their respective theoretical backgrounds. These questions were the leading ground for conducting the large randomized controlled trial that is described in this dissertation. It goes without saying that focusing on the two most commonly-practiced psychological treatments for MDD in the Netherlands is not only important from a patient's point of view, but might also lead to a significant reduction in societal costs because CT and IPT are implemented on such a large scale.

Our trial is one of the largest clinical trials in the field, and made a significant contribution to the current state of knowledge on the relative effectiveness of CT vs IPT by e.g. providing the final information necessary to support the conclusion that the effects of individual CT and IPT for adult depression in the acute phase do not seem to differ within a 4 BDI point limit. Furthermore, our trial provided useful information for the developing research field of mechanisms of change (more innovative aspects of our trial are described later in this addendum). However, even though our trial was fuelled by the ambition to decrease the burden of disease associated with MDD, the direct value of this dissertation might lie primarily in its scientific implications. Process research is complicated and requires specific features of study designs – including carefully spaced, repeated measurements of clinical outcomes and process measures over the course of treatment, sufficient statistical power, and an appropriate control group – and the use of state-of-the-art analytic techniques. Unfortunately, up until now, most studies do not meet the criteria for reputable process research, mainly because they cannot establish the temporal relation between changes in potential mechanisms of interest and symptom change. Therefore, renewed, well-designed research on the mechanisms of change in psychotherapy is much needed. Our trial addresses this need by empirically testing a set of assumptions about the effects and mechanisms of CT and IPT using a research design that is particularly suited to investigate mechanisms of change. For a more detailed description of the theoretical, practical and research implications of our trial and for an overview of directions for further research, I refer the interested reader to the general discussion (chapter 9) and to the discussion sections of the specific chapters throughout this dissertation.

Specific target groups for whom our findings might be relevant

Given the fact that our study was conducted from a both a clinical as well as a societal perspective, trial findings are relevant to a variety of target groups. First of all, our findings might be of interest to researchers who strive for a better understanding of psychotherapy for depression. As mentioned above, knowledge about active ingredients of therapy can assist in the verification and refinement of theoretical models of the disorder, and allows enhancement of elements that are crucial for therapeutic change, while dismissing those found to be redundant. Furthermore, the design and methodology that was used in the current dissertation holds scientific value for researchers in this field. Second, our findings are relevant for mental health care professionals because they suggest areas for increased therapeutic focus within the existing protocols of CT and IPT. Third, this dissertation might be relevant for depressed patients and their families, since our studies contribute to the future provision of optimal, efficient, evidence-based treatments for depressed patients. Moreover, the finding that the current recommended care for depression in the Netherlands seems to be beneficial for a large part of depressed patients might be of interest for guideline- and policy makers and health insurance companies. To conclude, since the knowledge gained from this dissertation might eventually contribute to a reduction of the economic and societal burden caused by depression, the findings of this dissertation are relevant for society in general.

Activities and products

The findings of the studies presented in this dissertation point towards several possibilities for activities and/or products. A few possibilities are pointed out here. First of all, the fact that our findings add to the accumulating evidence that the recommended practice for the psychological treatment of depression in the Netherlands is effective across the full range of severity, and that these effects do not seem to differ for CT and IPT, can be used in decisions about which treatments to recommend in future issues of the multidisciplinary guideline for depression. Furthermore, the symptom centrality analysis in chapter 8 provided empirical evidence for the fact that several symptoms of depression are more influential in the treatment of depression than others. If future research is able to replicate these findings, it may help in determining treatment focus. Moreover, it would encourage efforts to develop tools for person-specific centrality analyses that can be used by therapists to indicate which symptoms should be specifically targeted in a particular patient. While it is too premature to draw firm conclusions about the theoretical models of psychotherapy and the underlying mechanisms that drive symptom change based on the findings in this dissertation, eventually the insights gained in the field of psychotherapy process research might be used to optimize treatments protocols and theoretical models of depression. To conclude, knowledge with regard to the methodology for examining processes of change, can serve as a starting point for the development of standard guidelines for mechanism research.

Innovation

Our trial is innovative as it was not only able to replicate findings from previous randomized comparisons of individual CT and IPT for depression, but could also extend them by adding several unique methodological features to the study design. For example, our trial was the first RCT of individual CT vs. IPT to add a Waiting-List Control (WLC) condition to the comparison. The inclusion of an untreated control group diminished uncertainty about whether or not the observed effect was a result of the natural course of depression. Furthermore, our trial was the first that examined enduring effects after treatment termination for both interventions. In addition, it did not only evaluate treatment effects in terms of symptomatic improvement, but also assessed effectiveness in several other domains including quality of life, and impairment in social functioning. Moreover – although not addressed in the studies presented in this dissertation – our trial is the first that collected data on costs and utilities parallel to the clinical outcome measures and process measures. This allows for the unique opportunity to examine how CT and IPT compare in terms of cost-effectiveness from a societal perspective as well. Another innovative aspect of our trial is that it is one of the few studies in the field that included multiple measures of various process measures throughout treatment. The repeated measures design provided the opportunity to evaluate whether change in proposed process measures preceded, followed from, or went together with changes in depression. To conclude, by obtaining measures of depression severity from session to session, we were the first to examine sudden gains and the dynamic relations between individual items of the Beck Depression Inventory-II (BDI-II) in the context a randomized comparison of CT vs. IPT.

In addition, the analytic techniques that were used in the various studies of dissertation may be regarded as scientifically innovative. For example, the use of latent difference score (LDS) models in the empirical study on mechanisms that was presented in chapter 6 provided the opportunity to examine both temporal relations and mediation in one model. By combining intensive longitudinal data (i.e. session-to-session assessments of the BDI-II) with the newly developed vector autoregressive (VAR) multilevel method (chapter 8) we were the first to estimate the network of symptoms that characterizes the BDI-II. In addition, it enabled us to explore the centrality of symptoms and the community structure of the network. So even though the studies that were presented in this dissertation were not able to provide clear cut empirical evidence for the underlying mechanisms they could serve as a starting point for future studies. Not in the least because they provide guidance for future research questions and improvement of study designs.

Schedule and Implementation

Bringing our study to the attention of our target groups was an important goal throughout the study. In an early phase of the study we communicated the rationale for our trial and the study design to a broad audience by means of traditional and (social) media appearances. Patients were informed through various national online platforms for mental health and depression (e.g. psychischegezondheid.nl, depressie.startpagina.nl). The scientific community learned about our study by means of journal publications and conference presentations. Furthermore, our research group has developed strong relationships with various national (e.g. Groningen, Amsterdam, Leiden) and international (e.g. Philadelphia, Nashville, Trier) research groups that focus on psychotherapy process research, which contributed to further dissemination of knowledge. In order to reach clinicians, we published the rational for research into mechanisms and design of our study in a Dutch clinical journal that specifically targets mental health care professionals who are not working in a research setting (GZ-Psychologie). In addition, we presented our study at several local and national meetings for CT and IPT therapists. Since many of the therapists that were involved in the study are teaching and supervising CT or IPT, our study was also brought into attention among mental health care students. Public engagement in the study was promoted by means of a study website and two articles in a local newspaper (Dagblad de Limburger, 03-02-2011 & 31-03-2011). In order to stimulate the exchange of knowledge between the various target groups, we organized a national symposium on psychotherapy research for depression (the Maastricht Symposium of Evidence based psychotherapy, September 2013). In this symposium more than 80 researchers, mental health care professionals, mental health students, patients and other people interested in psychotherapy research, spent the afternoon talking about the latest findings, insights and developments in this research field. The symposium was featured in an article in the University Newspaper (Observant, 12-09-2013).

For the later stages of the trial we are planning to use similar strategies for further valorization and dissemination of research findings. For example, we will translate the studies presented in chapters 3 and 7 into Dutch and publish them in clinical journals, and will communicate our findings to the general public by means of another newspaper interview. Some of the studies presented in this dissertation have already evoked interest outside the scientific community. For example, the study described in chapter 3 has been featured on doctors.net.uk, the largest and most active network of doctors in the UK reaching more than 45.000 people each day. The same study was featured in a blog on the psychotherapy practice research network (PPRnet), an online inter-disciplinary collaboration among clinicians, educators, researchers and knowledge users who engage in practice-based psychotherapy research. Other methods for further knowledge valorization – within and outside the scientific community – include e.g. journal publications, conference presentations, press releases and social media.

References

Bromet, E., Andrade, L., Hwang, I., Sampson, N., Alonso, J., de Girolamo, G., . . . Kessler, R. (2011). Cross-national epidemiology of DSM-IV major depressive episode. *BMC Medicine, 9*(1), 90.

Graaf de, R., ten Have, M., & van Dorsselaer, C. (2010). *De psychische gezondheid van de Nederlandse bevolking - NEMESIS-2: Opzet en eerste resultaten*. Utrecht: Trimbos-Instituut.

Graaf de, R., Tuithof, M., van Dorsselaer, S., & ten Have, M. (2011). *Verzuim door psychische en somatische aandoeningen bij werkenden: Resultaten van de ,Netherlands Mental Health Survey and Incidence Study-2' (NEMESIS-2)*. Utrecht: Trimbos-Instituut.

Keller, M. B., & Boland, R. J. (1998). Implications of failing to achieve successful long-term maintenance treatment of recurrent unipolar major depression. *Biological Psychiatry, 44*(5), 348-360.

Paykel, E. S. (2008). Partial remission, residual symptoms, and relapse in depression. *Dialogues in Clinical Neuroscience, 14*, 431-437.

RIVM. (2013). *Volksgezondheid Toekomst Verkenning: Nationaal Kompas Volksgezondheid, versie 4.16*. Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu (RIVM).

Slobbe, L. C. J., Smit, J. M., Groen, J., Poos, M. J. J. C., & Kommer, G. J. (2011). *Kosten van Ziekten in Nederland 2007: Trends in de Nederlandse zorguitgaven 1999 - 2010*. Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu (RIVM).

World Health Organization (WHO). (2012). Fact Sheet Depression (N°369). from http://www.who.int/mediacentre/factsheets/fs369/en/

Summary

Major Depressive Disorder (MDD) is a prevalent psychiatric disorder characterized by depressed mood and markedly diminished interest or pleasure in almost all activities that substantially impairs quality of life and has high societal costs. To reduce the burden of disease, it is important that people suffering from MDD are treated adequately. Over the years, Cognitive Therapy (CT) and Interpersonal Psychotherapy (IPT) have repeatedly shown to be well-standardized, efficacious psychological interventions for the acute treatment of MDD. However there is room for improvement, since not all patients treated with CT and IPT respond to treatment, and chances of relapse and recurrence are high. Treatment improvement starts with a full understanding of interventions. Even though our knowledge about psychotherapy has increased tremendously over the past few decades, there are several unresolved issues that need further examination. For instance, it is not clear yet whether one therapy outperforms the other with regard to severity and course of the disorder, especially on the long-term. Furthermore, only little is known about the pattern of change and the underlying mechanisms responsible for this. These two questions were the leading ground for conducting the large randomized controlled trial that is described in this dissertation. The aim, outline and main findings of each chapter will be summarized here.

Chapter 1 provides a general introduction to the contents of this dissertation. It contains an overview of the diagnostic criteria of MDD, and discusses prevalence rates and burden of disease associated with depression. Furthermore, theoretical underpinnings, rationales and techniques of CT and IPT are presented, together with an overview of their similarities and differences. The rational for the research presented in this dissertation is explained from the perspective of current challenges in psychotherapy research. The chapter ends with the aim and outline of the dissertation.

In **Chapter 2** the design of the RCT that evaluated the clinical effectiveness and mechanisms of change of CT and IPT is presented. The chapter includes a rational for the trial, an overview of research questions, and an extensive description of the (recruitment) procedure, the measurement instruments and the statistical analyses. A total of 182 depressed adult outpatients were recruited during regular intakes at the Maastricht Community Mental Health Centre (RIAGG). After informed consent was obtained, and a baseline assessment was completed, patients were randomly allocated to either CT ($n = 76$), IPT ($n = 75$) or a 2-month Waiting-List Control (WLC) condition followed by treatment of choice ($n = 31$). The CT protocol was based on the manual by Beck et al. (1979). The IPT protocol followed the guidelines laid out by Klerman et al. (1984). Treatment consisted of 16 to 20 individual sessions of 45 minutes, depending on the progress of the individual patient. Primary outcome of the RCT was depression severity as measured with the Beck Depression Inventory-II (BDI-II). Secondary outcomes included general psychological distress, impairments in social functioning and several measures of quality of life. Various (therapy specific and non-specific) process variables were included in order to examine

mechanisms of change. Parallel to the data on clinical effects and mechanisms, data on costs and utilities were collected for the purpose of an economic evaluation (not part of this dissertation). Data were gathered within a 24-month time frame. In reporting the follow-up period of the study, three phases were distinguished; (1) the 'Treatment Phase' (baseline to 7 months); the period when therapy was delivered; (2) the 'Trial Follow-Up Phase' (month 8-12) in which depression severity was measured monthly; and (3) the 'Long term Follow-Up Phase' (month 12 to 24, not part of this dissertation); the period covered by retrospective assessment at 24 months. The main assessment points for the active conditions (CT and IPT) were baseline, 3, 7, 9, 12 and 24 months. Primary outcome BDI-II was also assessed at 2, 8, 10 and 11 months, and before every session. Patients in the WLC condition were assessed at baseline, 2, 9, 14 and 26 months. Subsequent chapters of this dissertation present findings obtained in the context of this trial.

In **Chapter 3** the clinical findings from the RCT are reported. We examined whether short-term effects of active treatment were superior to those of the waiting-list control after 2 months, and found that both treatments exceeded response in the WLC condition. Furthermore, it was tested whether one of the treatments was superior to the other in decreasing depressive symptoms and on a set of secondary outcome measures at post-treatment (7 months) and up to 5-months of follow-up. Within our power and time ranges, CT and IPT appeared not to differ in the acute treatment of depression. Both treatments led to considerable improvement in self-reported depression severity and quality of life, and effects were sustained up to 5 months beyond the end of treatment. No differential effects among therapists were found. Baseline depression severity, number of sessions, and several other potential moderators that were examined did not seem to moderate the effect between treatment and outcome. Even though pre- to post-treatment effect sizes were in range with those obtained by other trials in the field, change in our trial occurred later, which might be related to a difference in session frequency across studies. In order to determine the relative contribution of our study to the field, we added our data to the existing evidence base of trials that examined individual CT vs. IPT, and meta-analysed findings. Trial sequential analysis subsequently indicated that our study added the final information necessary to conclude that CT and IPT do not seem to differ within a 4 BDI-II points limit. Analysis with more strict presumptions (3-point BDI-II difference) indicated that more data is needed to definitely settle the question of a differential effect. Methodological considerations, clinical implications and suggestions for future research are discussed.

The objective of the study presented in **Chapter 4** was to explore the value of a newly developed implicit self-associative measure for depression. To the extent that implicit measures truly reflect uncontrollable and unaware attitudes and beliefs, they could potentially serve as a proxy of underlying schemata, and thereby contribute to research on mechanisms. However, research with implicit measures in depression is limited and

results are often contradictory to those obtained by explicit measures and to cognitive theory. One reason for this is that the instruments that have been used are not specific enough to capture these processes. We therefore designed a single category implicit association task (Sc-IAT) that specifically reflected depressogenic core beliefs related to self-esteem: unlovability, helplessness, and incompetence. We examined differences in explicit and implicit measures of self-esteem between depressed patients and healthy controls, and investigated whether our implicit and explicit self-associative measures were associated with each other and with depressive symptoms. In line with previous research, explicit self-esteem in depressed patients was found to be significantly lower than in healthy controls. However, in spite of our adaptations, we were not able to differentiate between patients and controls on the implicit measure of self-esteem. Both groups reported positive implicit associations with the self. Furthermore, our results indicated that only the explicit measure of self-esteem, and not the implicit, was related to depression. Even though it would be tempting to conclude that these findings suggest that self-esteem in depressed patients is only affected on a more superficial (explicit) level instead of a deeper rooted level (implicit), it is too soon to translate findings obtained with implicit measures to attitudes and beliefs. Procedures such as the Sc-IAT are relatively young and it is not clear yet in what sense the obtained measurement outcomes can be considered implicit and valid indices of self-esteem. It was therefore decided not to include this measure as a representative of underlying schemata in our mechanism study reported in chapter 6.

In **Chapter 5** a systematic overview and critical evaluation of the empirical literature on psychological mediators in psychotherapy for depression is provided. In a systematic literature search, we identified a total of 32 relevant empirical studies examining a total of 40 different potential mediators in 12 different treatment modalities. Study characteristics and results were explored. Furthermore, each study was rated with respect to meeting or not meeting several important requirements for mediation research: the use of an RCT design, inclusion of a control group, sufficient sample size, examination of multiple potential mediators, assessment of temporality, and the direct experimental manipulation of the mediator. Results indicate that – in spite of increased attention for mechanisms and theoretical consensus about necessities for proper mediation research – the evidence in this field is limited, heterogeneous and unsatisfactory in theoretical and methodological respect. Only a few studies met the highest standards for tests of treatment mediation. Probably the biggest challenge in research aimed at identifying mediators that became apparent in this review is discerning the temporal relation between change in the mediator and change in depression severity. Furthermore, none of the identified studies used an approach in which the proposed mediator was experimentally manipulated. We discuss the need for standard guidelines for conducting mediational analysis, thereby enabling comparisons across studies and integration of findings into broader knowledge.

The empirical study presented in **Chapter 6** was aimed at increasing our knowledge on the underlying mechanisms of change of CT and IPT. It was examined whether scores on 5 potential mechanisms (Dysfunctional Attitudes, Interpersonal Problems, Rumination, Self-Esteem and Therapeutic Alliance) changed over the course of therapy, and whether this differed between the two conditions. Furthermore, for each potential mediator, the temporal relationship with change in depressive symptoms was examined, and it was investigated whether these theorized processes mediated the relation between treatment and outcome. Over the course of treatment, patients showed improvement on all process measures, with medium to large effect sizes. However, change in processes was smaller than change in symptoms. No differential effects between CT and IPT were found. Change in process variables and concurrent change in depression were strongly related. Except for a significant relation between early change in self-esteem and subsequent change in depression severity, no temporal relations were found. Change on the BDI-II was mediated by concurrent change in interpersonal problems. No temporal mediation was found. As a result, we were not able to demonstrate empirical evidence for the various theoretical models of change. Several explanations for our findings are discussed including the possibility that theories are incorrect or that our design did not suffice. We argue that given the current methodological status of research, the latter seems to be the most plausible. However, we do not rule out that theories have to be revised in the future, especially with regard to the assumption that a mechanism needs to be specific for one type of treatment.

In **Chapter 7** an empirical study on the rates, baseline predictors and clinical impact of 'Sudden Gains' – large and stable symptom improvements during a single between-session interval – over the course of CT and IPT is presented. BDI-II data obtained prior to each therapy session were used to identify patients who met criteria for sudden gains. Subsequently, the relation between sudden gain status and end of treatment symptom severity was studied, as well as symptom severity at 5 months FU. In addition, patient characteristics that might predict the occurrence of sudden gains were examined, both within each treatment as well as across the interventions. After that it was investigated whether the duration of the between-session interval at which sudden gains were recorded (shorter or longer than 14 days) affected the results obtained. Analysis showed that there were significantly more patients with sudden gains in CT (42.2%) as compared to IPT (24.5%). A closer look at these differences indicated that the difference between CT and IPT did not emerge because CT facilitated a higher number of large symptom improvements between two consecutive sessions, relative to IPT, but rather seemed to be related to the relative lack of stability of the drops in IPT. Potential explanations for the lack of stability in IPT are discussed. No differences with regard to the magnitude and timing of the sudden gains were found between conditions. Sudden gains were predicted by baseline quality of life score and the absence of axis-I comorbidity, suggesting that relatively "good health" at baseline predicts sudden gains. Patients with sudden

gains were less depressed at post-treatment than were those without sudden gains, and remained so up to 5 months after treatment termination. The duration of the between-session interval did not influence the results. We conclude that the fact that there are differences in occurrence of sudden gains in two treatment modalities that overall show similar results might reflect different mechanisms of change.

Chapter 8 describes an empirical study into the dynamic (i.e. session-to-session) relations between individual items of the BDI-II over the course of treatment. The study consisted of three parts: First, we inferred the networks representing the session-to-session relations between the 21 BDI-II symptoms across 20 weeks of treatment using an adapted version of the recently developed vector autoregressive (VAR) multilevel method. In order to examine difference in symptom dynamics for patients receiving CT and IPT networks were fitted separately for CT and IPT, and for the two treatments combined. Second, the presence of communities – clusters of symptoms that are more strongly interconnected with each other than with other symptoms in the network – was examined. Third, centrality analysis was conducted in order to determine the relative importance of a symptom in the network. Results indicated that all symptoms measured with the BDI-II were directly or indirectly connected to each other. The strongest connections were positive, indicating that improvement in one symptom was related to improvement in the other, and vice versa for deterioration. The network structure did not differ across the two therapy groups. Community analyses indicated that the dynamic structure of the BDI-II involves a cognitive and a somatic-affective cluster. This corresponds with Beck's two-factor structure and with the biomedical model of depression. Symptom centrality analysis indicated that several BDI-II items were more influential than others; the items 'Suicidal Ideation', 'Loss of Pleasure' and 'Feelings of Failure' had the strongest influence on other symptoms and/or were important symptoms for further symptom spread. 'Sadness' and 'Loss of Interest' were less central in the model. Hereby our results only partly back-up the classification of depression according to the DSM-IV in the treatment of depression. Since this was the first study to examine symptom centrality and dynamic relations over the course of treatment, no firm conclusions can be drawn from these findings. However, if future research is able to replicate these findings and gain more insight in centrality of symptoms, this information may help in setting up treatment strategies.

In **Chapter 9**, the main findings of the various studies presented in this dissertation are summarized and integrated in the broader knowledge base. Methodological considerations and implications for research and clinical practice are discussed. The main message of the chapter – and of this dissertation as a whole – constitutes that, even though trial findings add to the accumulating evidence that the recommended practice for psychotherapeutic treatment of depression in the Netherlands is effective, the studies presented in this dissertation were not able to provide clear cut empirical evidence for the underlying mechanisms of change. Psychotherapy is a multifaceted phenomenon

that might work through interplay of multiple mechanisms at several levels (i.e. physiological, affective, behavioural and cognitive). Furthermore, the relationships between mechanisms and outcome might differ at various points in time, and between specific (subgroups) of patients. Psychotherapeutic change might therefore be too complex to be explained in relatively simple causal models of psychological change. However, our studies do provide guidance for future research questions and improvement of study designs. Briefly put, the field would benefit from further refinement of research methods to disentangle mechanisms of change, advances in personalized medicine, and more insight in how CT and IPT compare with regard to cost-effectiveness and relapse prevention. Together, this can bring us closer to an optimal understanding of why, how (long) and for whom CT and IPT for depression might work.

Samenvatting

(Dutch Summary)

Depressie is een veelvoorkomende psychische aandoening die gekenmerkt wordt door een sombere stemming en een duidelijke vermindering van interesse en plezier in vrijwel alle activiteiten. Depressie veroorzaakt substantiële beperkingen in het dagelijks leven van de patiënt en gaat gepaard met hoge maatschappelijke kosten. Om de ziektelast van depressie te reduceren is het belangrijk om depressieve patiënten adequaat te behandelen. Zowel Cognitieve Therapie (CT) als Interpersoonlijke Psychotherapie (IPT) zijn effectief gebleken in het verminderen van depressieve klachten. Beide behandelingen behoren momenteel tot de aanbevolen psychotherapeutische zorg voor depressie in Nederland. Helaas profiteren niet alle depressieve patiënten (voldoende) van deze behandelingen en – zelfs bij succesvolle behandeling – is de kans op terugval zeer groot. Verbetering van therapieën is daarom van groot belang. Een essentiële stap in het verbeteren van therapieën, is inzicht verkrijgen in de effecten en mechanismes van de genoemde behandelinterventies. Hoewel de kennis over psychotherapie de afgelopen jaren sterk is toegenomen, zijn er nog veel vragen onbeantwoord. Zo is bijvoorbeeld nog niet duidelijk of het effect van één van beide therapieën superieur is aan het effect van de ander, voornamelijk op de lange termijn. Daarnaast is er nog weinig bekend over het patroon waarin verandering plaatsvindt en over de werkingsmechanismes die hieraan ten grondslag liggen. Deze vragen waren aanleiding voor en uitgangspunt van de grootschalige gerandomiseerde behandelstudie naar de klinische effecten en werkingsmechanismes van CT en IPT voor depressie die beschreven wordt in dit proefschrift. Hieronder wordt elk hoofdstuk kort samengevat.

Hoofdstuk 1 is de algemene inleiding op de inhoud van dit proefschrift. Het hoofdstuk start met een overzicht van de diagnostische criteria van depressie en bespreekt prevalentiecijfers en de ziektelast die geassocieerd is met de aandoening. Verder worden de theoretische modellen en behandelprotocollen van CT en IPT besproken en worden verschillen en overeenkomsten tussen de twee interventies kort uiteengezet. Het belang van onderzoek naar de effecten en werkingsmechanismes van psychotherapie wordt geïllustreerd aan de hand van enkele actuele uitdagingen in het veld. Het hoofdstuk eindigt met een korte beschrijving van de behandelstudie en de structuur van het proefschrift.

In **Hoofdstuk 2** wordt een uitgebreide beschrijving gegeven van de opzet van onze gerandomiseerde gecontroleerde studie (RCT) naar de klinische effectiviteit en werkingsmechanismes van individuele CT en IPT voor depressieve volwassenen. Na het bespreken van de rationale en de belangrijkste onderzoeksvragen, wordt een nauwkeurige beschrijving gegeven van de (wervings)procedure, de meetinstrumenten en de statistische analyse methodes. De 182 patiënten die deelnamen aan onze studie werden geworven tijdens reguliere intakes bij RIAGG Maastricht. Geschikte kandidaten tekenden een toestemmingsverklaring en ondergingen een uitgebreide baselinemeting. Vervolgens werden ze middels loting toegewezen aan één van de drie onderzoekscondities:

CT (n = 76), IPT (n = 75) of een wachtlijstcontrole (WLC) - conditie van 2 maanden ge-volgd door behandeling naar keuze (n = 31). De laatstgenoemde conditie werd toe-gevoegd om te controleren voor het natuurlijk beloop van depressie. Het CT protocol was gebaseerd op het protocol van Beck et al. (1979). Het IPT protocol was conform de richtlijnen van Klerman et al. (1984). De behandeling bestond – afhankelijk van de vorde-ringen van de individuele deelnemers – uit 16 tot 20 individuele sessies van 45 minuten. De sessies vonden wekelijks plaats, maar het protocol bood de mogelijkheid om sessies in de afrondingsfase minder frequent te laten plaatsvinden. Primaire uitkomstmaat was de ernst van depressie, vastgesteld door de Beck Depressie Vragenlijst-II (BDI-II). Secon-daire uitkomstmaten waren onder meer algemene psychologische stress, beperkingen in sociaal functioneren en kwaliteit van leven. Verder werden diverse potentiele wer-kingsmechanismes (therapiespecifiek en theorie overstijgend) gemeten met als doel mechanismen van verandering te onderzoeken. Parallel aan deze uitkomstmaten, werd informatie verzameld over de kosteneffectiviteit en kosten-utiliteit (geen onderdeel van dit proefschrift). Patiënten werden in totaal 2 jaar gevolgd. De follow-up periode van de studie kan worden onderverdeeld in 3 fases: 1) 'de Behandelfase' (baseline tot 7 maan-den): de periode waarin patiënten therapie kregen; 2) 'de Studie Follow-Up Fase' (8 tot 12 maanden): de periode waarin de ernst van depressie maandelijks gemeten werd; en 3) 'de Lange Termijn Follow-Up Fase' (maand 12 tot 24, geen onderdeel van dit proef-schrift), een periode waarin het beloop van de klachten in kaart werd gebracht middels een retrospectieve meting op 24 maanden. De belangrijkste meetmomenten voor de behandelcondities (CT en IPT) waren baseline, 3, 7, 9, 12 en 24 maanden. De ernst van depressie (BDI-II) werd bijkomend gemeten op 2, 8, 10 en 11 maanden, en voorafgaand aan elke therapiesessie. Patiënten in de WLC conditie werden gemeten op baseline, 2, 9, 14 en 26 maanden. Alle volgende hoofdstukken van dit proefschrift beschrijven de resultaten van deze gerandomiseerde gecontroleerde studie.

In **Hoofdstuk 3** worden de bevindingen van de klinische effectiviteitsstudie gepresen-teerd. In deze studie werd onderzocht of het effect van behandeling na 2 maanden gro-ter was dan het effect in de wachtlijstconditie. Verder werd bekeken of één van beide therapieën (CT of IPT) effectiever was in het reduceren van depressieve klachten, alge-mene psychologische stress en beperkingen in sociaal functioneren en in het verbete-ren van kwaliteit van leven. De interventies werden vergeleken aan het einde van de be-handelfase (7 maanden) en tot 5 maanden na het afronden van de therapie. Daarnaast werden therapeut-effecten bekeken en werd onderzocht of diverse potentiële mode-ratoren – zoals ernst van depressie en het totaal aantal therapiesessies – de resultaten beïnvloedden. Onderzoeksresultaten lieten zien dat beide interventies tot aanzienlijke verbetering in zelf-gerapporteerde depressieve klachten en kwaliteit van leven leiden. Daarnaast werden er aan het einde van de behandelfase lagere niveaus van algemene psychologische stress gemeten en rapporteerden patiënten minder beperkingen in het sociale functioneren. De gevonden effecten hielden aan tot 5 maanden na het afsluiten

van de behandeling. Er werden geen verschillen gevonden tussen CT en IPT. De kwaliteit van beide interventies – bepaald door het beoordelen van een willekeurige selectie van video-opnames van therapiesessies door onafhankelijke beoordelaars – werd beoordeeld als goed tot uitmuntend. Het percentage uitvallers tijdens de studie was laag. De effecten van de behandelcondities CT en IPT na 2 maanden waren groter dan die van de wachtlijstconditie. Hiermee werd uitgesloten dat de geobserveerde verandering in de behandelcondities louter het resultaat was van het natuurlijk beloop van depressie. Het behandelresultaat werd niet beïnvloed door individuele verschillen tussen therapeuten. Ook de onderzochte moderatoren leken de relatie tussen behandeling en uitkomst niet te beïnvloeden. Hoewel het totale behandeleffect in onze studie overeenkomt met resultaten van vergelijkbare studies, vond de verbetering van het depressief toestandsbeeld in onze studie relatief laat plaats. Dit heeft mogelijk te maken met een verschil in frequentie waarin de sessies werden aangeboden. Om de relatieve bijdrage van onze studie aan het veld te bepalen, werd de data van onze studie samengevoegd met de data van drie reeds bestaande RCT's die de effecten van individuele CT en IPT voor depressie onderzocht hebben en werden een meta-analyse en trialsequentie analyse uitgevoerd. Trialsequentie analyse met een minimaal relevant verschil van 4 punten op de BDI-II liet zien dat onze studie de nog ontbrekende informatie heeft toegevoegd om te kunnen concluderen dat CT en IPT niet verschillen in de acute behandeling van depressie. Een analyse met wat strengere aannames (BDI-II verschil van 3 punten) liet echter zien dat er nog meer data verzameld dienen te worden voordat de equivalentie vraag definitief beantwoord kan worden. Methodologische overwegingen, klinische implicaties en richtlijnen voor toekomstig onderzoek worden besproken.

In de studie die gepresenteerd wordt in **Hoofdstuk 4** werden de verschillen geëxploreerd tussen expliciete en impliciete maten van zelfbeeld bij depressieve patiënten en gezonde controles. Er werd gebruik gemaakt van een impliciete associatietaak die speciaal door ons ontwikkeld was om negatieve zelf-associaties bij depressie te meten. In de zin dat impliciete maten daadwerkelijk oncontroleerbare en onbewuste processen kunnen meten, zouden zij gezien kunnen worden als een benadering van onderliggende cognitieve schema's die het zelfbeeld illustreren. Daarmee zouden zij een belangrijke bijdrage kunnen leveren aan het onderzoek naar verandermechanismes. Echter, onderzoek naar zelfbeeld bij depressie met impliciete maten is schaars, en de resultaten die wel beschikbaar zijn, staan vaak haaks op de bevindingen die verkregen zijn uit expliciete maten. Een mogelijke verklaring hiervoor is dat de tot nu toe gebruikte impliciete maten niet specifiek genoeg waren om de automatische zelf-associaties in depressie te kunnen meten. Daarom hebben we een (single category) impliciete associatietaak (Sc-IAT) ontwikkeld die zich specifiek richt op depressogene kernopvattingen gerelateerd aan het zelfbeeld bij depressie: ongeliefdheid, hulpeloosheid en incompetentie. Deze taak werd afgenomen bij 87 depressieve patiënten en 30 gezonde controles. Allereerst werd bekeken of depressieve patiënten sterkere negatieve automatische zelf-associaties lieten

zien dan gezonde controles. Daarna werden scores op de impliciete maat vergeleken met scores op een expliciete maat. Vervolgens werd onderzocht of automatische zelf-associaties samenhingen met scores op de expliciete maat, en met depressieve symptomen. In lijn met eerder onderzoek was het expliciete zelfbeeld van depressieve patiënten significant negatiever dan dat van gezonde controles. Echter, ondanks de aanpassingen aan onze impliciete maat, vonden we geen verschillen tussen patiënten en controles op het gebied van impliciet zelfbeeld. Beide groepen lieten positieve associaties over zichzelf zien. Bovendien toonden de resultaten aan dat alleen de score op de expliciete maat – en niet die op de impliciete maat – geassocieerd was met depressie. Hoewel het verleidelijk zou zijn om op basis van deze bevindingen te concluderen dat zelfbeeld van depressieve patiënten enkel aangetast is op een meer oppervlakkig (expliciet) niveau en niet zozeer op een dieper onderliggend (impliciet) niveau, zou dit voorbarig zijn. Procedures zoals de Sc-IAT zijn nog relatief nieuw en het is nog onduidelijk in welke mate zij ook daadwerkelijk een impliciete en valide weerspiegeling van zelfbeeld geven. Zolang dit nog niet helder is, is het te vroeg om resultaten van impliciete maten te vertalen naar attitudes en opvattingen. Daarom werd besloten Sc-IAT niet verder op te nemen als vertegenwoordiger van onderliggende cognitieve schema's in onze studie naar mechanismes van verandering in hoofdstuk 6.

In de review in **Hoofdstuk 5** wordt de empirische literatuur op het gebied van psychologische mediatoren van psychotherapie voor depressie bestudeerd. Een mediator is een variabele die verklaart waarom en hoe een behandeling effect heeft op de uitkomstmaat, en kan gezien worden als aanwijzing voor een mechanisme. In een systematische literatuurstudie werden 32 relevante empirische studies geïdentificeerd die in totaal 40 verschillende potentiele mediatoren onderzochten in 12 verschillende behandelmodaliteiten. Kenmerken en resultaten van de geïdentificeerde studies werden uiteengezet. Verder werd bekeken in hoeverre elke studie voldeed aan een aantal voorwaarden voor gedegen mechanisme onderzoek. De resultaten lieten zien dat onderzoek naar mediatoren in psychotherapie voor depressie – ondanks toegenomen aandacht voor mechanismeonderzoek en theoretische consensus over de benodigde methodologie – nog steeds beperkt, heterogeen en van onvoldoende methodologische kwaliteit is. Bevindingen zijn veelbelovend maar zeer zeker niet onbetwistbaar. In ongeveer de helft van de gevallen waarin een bepaalde mediator werd onderzocht, werd bewijs gevonden voor de aanwezigheid van statistische mediatie. Slechts een klein deel van de geïdentificeerde studies voldeed aan de hoogste eisen voor mediatie onderzoek. Dit kwam voornamelijk doordat studies niet konden aantonen dat verandering in de mediator vooraf ging aan verandering in de uitkomst (temporaliteit), een principe dat essentieel is voor het aantonen van causaliteit. Daarnaast werd in geen enkele van de geïdentificeerde studies gebruik gemaakt van experimentele manipulatie van de mediator.

Het onderzoek beschreven in **Hoofdstuk 6** exploreert de werkingsmechanismen van CT en IPT. In deze empirische studie werd onderzocht of scores op 5 potentiële procesmaten (disfunctionele attitudes, interpersoonlijke problemen, ruminatie, zelfbeeld en werkrelatie tussen patiënt en therapeut) veranderden tijdens de behandeling. Ook werd bekeken of er verschillen waren tussen CT en IPT. Daarna werden de diverse concurrente en temporele relaties tussen verandering op procesmaten en verandering in depressieve symptomen in kaart gebracht. Tevens werd bekeken of de in de theorie geformuleerde procesmaten de relatie tussen behandeling en uitkomst medieerden. Patiënten lieten tijdens de behandeling verbetering zien op alle procesmaten. Effecten werden geclassificeerd als gemiddeld tot groot. Echter, de veranderingen in de processen waren kleiner dan de veranderingen in de depressieve symptomen. Er werden geen verschillen gevonden tussen CT en IPT. Verandering in procesmaten en gelijktijdige verandering in depressieve symptomen (concurrente relaties) waren sterk aan elkaar gerelateerd. Met uitzondering van een significante relatie tussen vroege verandering in zelfbeeld en latere verandering in ernst van depressie, werden er geen temporele relaties waargenomen. Verandering op de BDI-II werd gemedieerd door gelijktijdige (concurrente) verandering in interpersoonlijke problemen. Er werd geen bewijs gevonden voor temporele mediatie. Ondanks een weloverwogen onderzoeksdesign en het gebruik van een innovatieve statistische analysemethode is het niet gelukt om de verschillende theoretische modellen van verandering van CT en IPT te onderbouwen met empirisch bewijs. Diverse verklaringen voor deze bevindingen worden besproken, waaronder de mogelijkheid dat theorieën niet kloppen of dat ons studiedesign niet voldeed. Hoewel deze laatstgenoemde verklaring gezien de huidige status van mechanisme onderzoek het meest aannemelijk lijkt, sluiten we niet uit dat theorieën in de toekomst gewijzigd moeten worden, zeker met betrekking tot de aanname dat een mechanisme specifiek moet zijn voor één type behandeling.

In **Hoofdstuk 7** wordt een studie beschreven naar de frequentie, baseline voorspellers en klinische effecten van "Sudden Gains" – grote en stabiele positieve symptoomveranderingen die plaatsvinden tussen twee op elkaar volgende sessies. Allereerst werd de aanwezigheid van sudden gains bepaald. Dit werd gedaan aan de hand van de BDI-II data die tijdens therapiesessies verzameld zijn in de CT en IPT conditie. Vervolgens werd de relatie tussen het hebben van sudden gains en therapie uitkomst onderzocht. Effecten werden bekeken aan het einde van de behandeling en op vijf maanden follow-up. Daarnaast werd onderzocht welke patiëntkenmerken voorspellend waren voor het krijgen van sudden gains. In een volgende analyse werd bekeken of het interval tussen de sessies (korter of langer dan 14 dagen) die resultaten beïnvloedde. Er waren significant meer patiënten met sudden gains in de CT conditie (42.2%) dan in de IPT conditie (24.5%). Dit kwam niet zozeer doordat er in IPT geen grote symptoomverbetering op-

traden tussen twee sessies, maar meer vanwege het feit dat deze verbeteringen niet stabiel genoeg bleven in IPT. Diverse potentiële verklaringen voor het gebrek aan stabiliteit van symptoomverbetering in IPT werden besproken. Er werden geen verschillen gevonden tussen de twee condities met betrekking tot de grootte van de sudden gain en het moment waarop de sudden gain plaatsvond. Patiënten met sudden gains waren aan het einde van de behandeling minder somber dan mensen zonder sudden gains, en zij bleven dit tot tenminste vijf maanden na afloop van de behandeling. In beide condities werden sudden gains voorspeld door relatief 'goede gezondheid' op baseline (betere kwaliteit van leven en afwezigheid van comorbiditeit op as I). De duur van het interval tussen sessies had geen invloed op de bevindingen. Het feit dat CT en IPT – twee behandelingen met een vergelijkbaar behandelresultaat – verschillen in de frequentie van sudden gains kan een aanwijzing zijn voor verschillende patronen van symptoomverbetering.

Hoofdstuk 8 beschrijft een empirische studie naar de dynamische (ofwel sessie-tot-sessie) relaties tussen individuele items van de BDI-II. De studie bestond uit drie delen: allereerst werd een netwerk analyse uitgevoerd om de relatie tussen de 21 BDI-II items tijdens de gehele behandeling in kaart te brengen. Hiervoor werd een aangepaste versie van de recent ontwikkelde Vector Autoregressieve (VAR) multilevel methode gebruikt. Om te bepalen of er verschillen waren in symptoom-dynamieken tussen CT en IPT werden drie netwerken geschat: één voor CT, één voor IPT en één voor beide groepen samen. Onderzoeksresultaten lieten zien dat het netwerk waarin CT en IPT samengenomen waren de beste aanpassingsgraad had. Verder werd duidelijk dat alle symptomen die gemeten worden met de BDI-II direct of indirect met elkaar verbonden zijn. De sterkste relaties tussen symptomen waren positief. Dit betekent dat verbetering in het ene symptoom geassocieerd wordt met verbetering in het andere, en vice versa voor verslechtering. Om de relatieve invloed van elk symptoom binnen het netwerk te bepalen werd een centraliteitsanalyse uitgevoerd. Hieruit kwam naar voren dat sommige items – o.a. verlies van plezier, suïcidale ideatie, en gevoelens van falen - een centralere plek in het netwerk (en dus meer invloed op verdere symptoomverspreiding) hebben dan anderen – o.a. sombere stemming en verlies van interesse. Dit sluit slechts gedeeltelijk aan bij de classificatie van depressie volgens de DSM-IV. Tot slot werd onderzocht of er communiteiten aanwezig waren: clusters van symptomen die sterker met elkaar verbonden zijn dan met andere symptomen in het netwerk. Communiteitsanalyse identificeerde twee clusters van symptomen: een cognitief cluster en een somatisch-affectief cluster. Dit komt overeen met het biomedische model en de twee-factor structuur zoals voorgesteld door Beck en collega's. Indien de bevindingen van deze studie in de toekomst gerepliceerd worden, kan deze informatie bijdragen aan het ontwikkelen van (gepersonaliseerde) behandelstrategieën.

Hoofdstuk 9 biedt een algemene discussie van dit proefschrift. De belangrijkste bevindingen van de verschillende hoofdstukken worden samengevat en geïntegreerd. Verder worden methodologische bedenkingen en implicaties voor de klinische praktijk besproken. Daarnaast worden aanbevelingen voor toekomstig onderzoek gedaan. De kernboodschap van dit hoofdstuk – evenals van het gehele proefschrift – is dat hoewel bevindingen van onze studie bijdragen aan het cumulatieve bewijs dat de huidige aanbevolen psychotherapeutische zorg voor depressie in Nederland effectief is, we weinig empirische ondersteuning hebben gevonden voor de diverse theoretische modellen van verandering. Er is nog steeds geen duidelijk antwoord op de vraag welke mechanismes ten grondslag liggen aan de gevonden therapie effecten. Psychotherapie is een veelzijdig fenomeen dat mogelijk werkt door een samenspel tussen diverse factoren op verschillende niveaus (fysiologisch, affectief, gedragsmatig en cognitief). Daarnaast zijn er aanwijzingen dat de diverse relaties tussen werkingsmechanismes en uitkomstmaten kunnen variëren over tijd, en tussen specifieke (subgroepen van) patiënten. Psychotherapeutische verandering is daarom wellicht te complex om te verklaren vanuit relatief simpele causale modellen van psychologische verandering. Echter, de studies in dit proefschrift geven richting aan vervolgonderzoek. Kort gezegd zou onderzoek in dit veld kunnen profiteren van een verdere verbetering van methoden om mechanismes in kaart te brengen en van vooruitgang op het gebied van *personalized medicine*. Daarnaast is het raadzaam verder te onderzoeken hoe CT en IPT zich tot elkaar verhouden op het gebied van kosteneffectiviteit en terugvalpreventie. Dit alles kan bijdragen aan een optimaal begrip over waarom, hoe (lang) en voor wie CT en IPT voor depressie werken.

Dankwoord

(Acknowledgements)

Een promotietraject doorloop je niet in je eentje! Er zijn veel mensen die op directe of indirecte manier belangrijk zijn geweest bij de totstandkoming van dit proefschrift. Zonder hun vertrouwen, inzet, steun, en hulp lag dit proefschrift er nu niet. Ik neem hierbij graag de gelegenheid om mijn waardering daarvoor uit te spreken.

Allereerst wil ik alle **deelnemers** aan de STEPd studie bedanken. Het feit dat jullie, ondanks jullie eigen zorgen, twee jaar lang belangeloos meewerkten aan een onderzoek dat veel tijd in beslag nam en de nodige energie kostte, is bewonderenswaardig. Daarom een bijzondere plek voor jullie in dit dankwoord! Het draait immers uiteindelijk om jullie!

Marcus, Arnoud en Frenk, ook jullie wil ik hartelijk bedanken! Elk met jullie eigenheid vormden jullie samen een perfect begeleidersteam. In alle fases van mijn promotietraject ben ik meer dan blij geweest met de input, inspiratie, steun en kansen die ik van jullie heb gekregen.

Marcus, ik had me geen betere mentor kunnen wensen! Ik heb de afgelopen jaren zoveel van je geleerd! Je kennis, gedrevenheid, en authenticiteit zijn een ware inspiratie voor me en ik bewonder de manier waarop je mensen uitdaagt om het beste uit zichzelf te halen. Ondanks je drukke schema was je altijd beschikbaar voor vragen en wist je mijn manuscripten elke keer zeer snel van constructieve feedback te voorzien. Dat was een ware luxe! Dank daarvoor! Dat grote klinische projecten ook tot heel wat kopzorgen kunnen leiden is ons allebei niet vreemd. Ondanks dat wist je me er telkens van te overtuigen dat we de meest *awesome* studie ter wereld aan het draaien waren en verliet ik onze meetings altijd met veel nieuwe energie. Onze samenwerking heeft me enorm verrijkt, zowel als onderzoeker maar zeker ook als persoon! Ik wil je bedanken voor alle goede zorgen, voor je vertrouwen, en voor de vrijheid die je me hebt gegeven om zelf invulling te geven aan mijn promotietraject. Ik vond het heerlijk om te werken in een omgeving waarin zowel *work hard*, als *play hard* hoog in het vaandel staan, en kijk uit naar onze verdere samenwerking!

Arnoud, ook jij hebt me geïnspireerd de afgelopen jaren. Je kennis, kritische blik, en de nauwkeurigheid waarmee je werkt zijn ronduit indrukwekkend! Ik wil je bedanken voor je betrokkenheid en voor al je goede zorgen. Je deur stond altijd open als ik weer eens een vraag had over die ingewikkelde multilevel analyses. Hoewel jij zulke analyses waarschijnlijk zelf in *no-time* had kunnen doen, heb je elke keer uitgebreid de tijd genomen en me de kans gegeven om het zelf te ontdekken. De kennis die ik door jouw begeleiding op dit gebied heb kunnen opdoen is heel erg kostbaar voor me. Daarnaast heb ik veel gehad aan je goede ideeën en je waardevolle (en zeer snelle) feedback op mijn manuscripten. Arnoud, dankjewel voor alles!

Frenk, ook niets dan lof voor jou! Met je grote inhoudelijke kennis, praktische en oplossingsgerichte denkvermogen en je enorme betrokkenheid heb je een belangrijke bijdrage geleverd aan het project, aan onze papers en aan mijn persoonlijke ontwikkeling. Dankjewel daarvoor! Ook ben ik dankbaar dat je er altijd voor gezorgd hebt dat in deze CGT georiënteerde wereld ook de belangen van IPT voldoende behartigd werden. Daarnaast kon ik ook altijd bij de RIAGG bij je terecht voor consultatie en advies! Ik vind het fijn dat we onze samenwerking blijven voortzetten in een STEPd 2.0 constructie.

Ik wil ook alle andere leden van het **STEPd projectteam** hartelijk bedanken. Heleen, dank voor alle inspanningen die je verricht hebt in de beginfase van dit project! Anne, als toegewijd projectgroep lid ben ook jij een belangrijk voorbeeld voor me geweest tijdens dit traject. Ik bewonder je talent als wetenschapper, je gedrevenheid, en de manier waarop je werk en privé weet te combineren. Ik vind het erg fijn dat we nu opnieuw collega's zijn in de eetgroep. Dear Steve, I was so excited when I found out that I got the chance to work with you on this project! Your tremendous knowledge and enthusiasm has made working with you a true privilege. Thanks for taking the time for our regular skype meetings, and for providing me with valuable feedback on my papers. The opportunity to come to Nashville and present my work to the members of your research group was one of the highlights of my PhD project! Thanks for that! Suzanne, het is een fijn gevoel om het STEPd stokje aan jou over te dragen. Ik vind je een aanwinst voor ons team, en kijk uit naar alle dingen die we samen nog gaan doen! Kick some ass!

Graag wil ik alle leden van **team I van de afdeling VZ van RIAGG Maastricht** bedanken voor de manier waarop jullie me – zowel als onderzoeker en als therapeut – hebben opgenomen in jullie team. Een speciaal woord van dank aan de STEPd therapeuten – Anja H, Anja K, Annemiek, Claudia, Esmeralda, Frenk, Hanneke, Ina, Marcus en Marion. Wisten jullie dat jullie in het kader van dit project met z'n allen meer dan 2660 sessies therapie hebben gegeven en minstens zoveel vragenlijsten hebben ingevuld/laten invullen? Wat was het een klus hè! Dank voor jullie inzet en toewijding. Zonder jullie was de studie nooit zo'n succes geworden! Ook hartelijk dank voor jullie geduld en flexibiliteit als er bijvoorbeeld een nieuwe vragenlijst werd toegevoegd, er weer eens een camera stuk was, of als ik jullie achter de veren zat i.v.m. het inleveren van de blauwe mappen. Anja en Ina, jullie wil ik in het bijzonder bedanken voor alle praktische en morele ondersteuning! Van begin tot einde hebben jullie je – deels vanuit jullie rol als teamleider en wachtlijstbeheerder – ingezet om STEPd tot een goed einde te brengen. Dank daarvoor! Annemiek, ook voor jou een extra woord van dank voor de fijne werkbegeleiding. Daarnaast wil ik mijn oprechte dank uitspreken aan intakers en de intakestaf van team I (voor het screenen van 1562 patiënten), aan de dames van de afdeling academisering (en dan in het bijzonder Nicole, Marga, Christel en Lisette), de medewerkers van het telefonisch spreekuur, en alle andere RIAGG medewerkers die op welke wijze dan ook een bijdrage hebben geleverd aan dit project.

En dan, **de Annie's**! Waar zou dit project geweest zijn zonder jullie? Jullie inzet en onder-steuning zijn van onmiskenbaar belang geweest voor het slagen van dit project! Vanaf het eerste moment dat ik jullie kamer binnenkwam, ergens in maart 2010, klikte het tussen ons en waren we een fantastisch team! Ik bewonder jullie nauwgezetheid, vol-harding, en het eeuwige enthousiasme waarmee jullie mensen weten te motiveren om deel te nemen aan (en door te gaan met) onze studies! De ontelbare telefoontjes, mail-tjes, (kerst)kaartjes en reminders hebben ook nu weer hun vruchten afgeworpen: het Annie effect is wederom gerepliceerd! Dank daarvoor! Ook voor morele steun kon (en kan!) ik altijd bij jullie terecht. Of het nu ging om een bijklets-koffie-moment, het vieren van successen, of het delen van frustraties, jullie deur stond altijd open en niets was jullie teveel! Annie en Annie, jullie zijn geweldig! Dank voor alles!

Graag dank ik ook **alle anderen** die een speciale bijdrage hebben geleverd aan ons project. Allereerst wil ik het onderzoeksinstituut Experimentele Psychopathologie (EPP) en de raad van bestuur van RIAGG Maastricht hartelijk bedanken voor hun financiële steun aan dit project. Thanks to John Markowitz and Steve Hollon, for their willingness to train our therapists and for giving our study such an inspiring start. Laura, Denny en Francis, bedankt voor onze fijne samenwerking bij het netwerken artikel. Francis-ca, dank voor je waardevolle bijdrage aan het mediatie paper! Carolien Christ, Kosse Jonker, Meike Kruger en Dina Snippe wil ik bedanken voor het scoren van bandjes van onze integriteitstudie. Ook een woord van dank aan Esra, Inge, Lisanne, Lobke, Nick, Sandra en Sophie en aan alle andere studenten die in de afgelopen jaren heb-ben geholpen met het verzamelen van onderzoeksgegevens. Jullie inzet betekende veel voor me! Viola, thanks for your dedication and precision in screening the body of literature on mediators. Marion, Paul en Tessa, Madelon en Marcel, bedankt voor het beschikbaar stellen van jullie huis. Jullie hebben daarmee een flinke bijdrage geleverd aan het schrijfproces. Rosanne, wat jij niet allemaal bedenkt, regelt, en in werking zet om het onderzoek te faciliteren! En dat dan ook nog met een snelheid en precisie waar je *u* tegen zegt! Dankjewel! Ook een woord van dank aan Sita (voor de praktische ondersteuning van emium), Angela (voor je betrokkenheid en voor alle goede zorgen), Jessie, Marionne en Truus (voor secretariële ondersteuning met een glimlach) en aan Emmy en Chantal (voor jullie ondersteuning bij EPP- en onderwijs gerelateerde taken). En last but not least Lindy, al ruim 5 jaar secretaris van de enige echte *pupseclub*. Dat dat niet de enige rol is die je met verve vervult lijkt me duidelijk! Dank voor alle steun, hulp, betrokkenheid en gezelligheid!

I would like to thank the members of the **evaluation committee** – prof. dr. Anita Jansen, dr. Ellen Driessen, dr. Jill Lobbestael, prof. dr. Wolfgang Lutz and prof. dr. Peter Muris – for taking the time and effort to read and evaluate my dissertation. Furthermore, thanks are due to prof. dr. Rob DeRubeis, prof. dr. Claudi Bockting and dr. Nicole Geschwind for their willingness to take place in my defense committee.

Dear Rob, visiting **PENN** was one the best decisions made during my PhD project! It was a true honor as well as a great pleasure to work with you. The dedication and persistence with which you try to unravel the mechanism of psychotherapy are a true inspiration. Thank you for your hospitality, for sharing your knowledge, for *Robyfying* our papers, and for all our *spreadsheet parties* (which were almost as good as Dutch M&M's). An international collaboration like this is a delight! I hope that we will continue working together in the future! Also a big thank you to the *DeRubeis-cubes* – Lorenzo, Zach, Rami, Lois, and Jack – and to all the other people I met at PENN. You made me feel so welcome! So much more than just a visitor in your lab! Thanks for taking such good care of me, and for introducing me to the American way of life (although I still wonder where they keep the chicken). I cherish our friendships and hope we will stay in touch in the future! A special thanks to Lorenzo, who took care of both my mental as well as my physical health by being both a great friend and a dedicated gym-buddy!

Dan, wil ik al mijn **(ex)collega's** van de **vakgroep Clinical Psychological Science** – en van de **sectie Clinical Psychology** in het bijzonder – bedanken. Anne, Blazej, Cor, Ger, Jeffrey, Jill, Linda, Lorraine, Marisol, Nicole, Nicolette, Pauline, Peter, Pim, Rosanne, Sandra en Suzanne: wat is het een voorrecht om te mogen promoveren in een club met zulke getalenteerde en gedreven mensen! Dank ook aan alle mede AiO's die – verspreid over verschillende periodes (en verdiepingen van ons gebouw) – zo'n belangrijke factor zijn geweest in dit traject. Dank voor de gezelligheid op de UNS40, en voor alle borrels, etentjes, promotiefeestjes en singstar avonden! Het werk was niet half zo leuk geweest zonder jullie!

Een aantal collega's wil ik in het bijzonder noemen. Allereerst, Pim. Wat ben ik blij dat ik jou heb mogen leren kennen! Ik vind je niet alleen een zeer getalenteerde wetenschapper, maar ook een bijzonder fijn mens! Andrea, al sinds onze studietijd is duidelijk dat we onze passie voor onderzoek en onderwijs met elkaar delen! Het betekent veel voor me dat we al jarenlang zulke fijne en hechte collega's zijn. Lorraine en Linda dank voor de vele gezellige lunch- en koffiemomenten waarin we kunnen kletsen over van alles en nog wat! Peter, als voorzitter van onze vakgroep heb je me diverse kansen geboden om me verder te ontwikkelen. Daar ben ik je heel erg dankbaar voor! Anita, ook jou wil ik heel erg bedanken voor de kans die je me gegeven hebt om mijn ervaring met psychotherapie proces onderzoek verder uit te breiden naar het eetstoornissenveld! Dankjewel! Cor, als begeleider van mijn masterscriptie heb je mij als klinisch georiënteerde GGK'er helemaal klaargestoomd voor een baan als onderzoeker. Dank voor je vertrouwen en voor onze fijne samenwerking! Jill, als jij me niet had gewezen op deze promotieplek, had ik dit boekje waarschijnlijk nooit geschreven! Ik ben blij met mijn (ex-)roomies Anna, Martien, Haris, Bastiaan en Marieke. Het was (en is) fijn om de dagdagelijkse dingen met jullie te kunnen delen. Ook een speciaal

woord van dank aan Jessica (voor je interesse en betrokkenheid), Conny (ik had me geen beter EPP maatje kunnen wensen) en Melanie (dat we meer delen dan onze voorliefde voor karaoke lijkt me duidelijk!).

Alana, Dalena, Ken, Lea, Lieke, Manuela, Sylvia en Thomas, wat vond ik het jammer dat jullie weggingen! Johanna en Sjoertje, ik ben blij dat we na al die tijd nog regelmatig live contact hebben! Dan lieve Lotte, ik heb mijn bewondering voor de manier waarop jij in het leven staat (zowel professioneel en als persoon) nooit onder stoelen of banken gestoken. Ik vond het daarom ook een grote eer om een tijdlang door het leven te mogen gaan als 'de andere Lotte'. Ik hoop dat we elkaar nog lang blijven volgen!

Natuurlijk mogen mijn nieuwe collega's niet in dit lijstje ontbreken. Anita, Anne, Bastiaan, Carolien, Clare, Eva, Eric, Fania, Ghislaine, Janneke, Jessica, Karolien, Katrijn, Nele, Peggy, Remco, Sandra, Sjaan en Sieske: dank voor jullie warme welkom in de eetgroep. Ik heb me vanaf het eerste moment meteen thuis gevoeld in jullie club en kijk uit naar wat ons nog allemaal te wachten staat! Team II van de afdeling VZ van Virenze-RIAGG, wat is het fijn om in een hecht team vol gedreven, betrokken en bekwame hulpverleners terecht te komen!

En dan **Fritz en Marjolein**, mijn paranimfen. Jongens wat was ik trots toen ik vorig jaar van dichtbij mocht meemaken hoe jullie allebei met verve jullie proefschrift verdedigden! En wat ben ik blij dat jullie nu ook allebei mijn paranimf willen zijn! Fritz, als *research-siblings* en *roomies-van-het-eerste-uur* hebben we alle onderdelen van onze PhD samen doorgemaakt. De vele uren op kantoor, het vieren van publicaties, diverse congresbezoeken, ons eigen symposium en ons gezamenlijke bezoek aan Philadelphia. Maar ook de frustraties van een trage dataverzameling en een aantal andere '*stupid ass*' momenten. Wat zou mijn Aio-tijd anders zijn geweest zijn als jij er geen onderdeel van had uitgemaakt! Dankjewel voor alles! Marjolein, met je sprankelende persoonlijkheid en enorme betrokkenheid ben jij een constante factor van gezelligheid en steun geweest. Als ware *opper-snerds* hebben we ons in de afrondingsfase maar liefst 5 keer samen opgesloten in een huisje in het bos om daar onder het genot van een cola light of wat bessensap wetenschappelijke doorbraken te forceren. Bedankt voor je optimisme, je luisterende oor en voor alle gezelligheid! Ik hoop dat we onze tradities nog lang in ere zullen houden!

Dit werk kun je alleen volhouden met een aantal goede mensen om je heen die voor de broodnodige ontspanning en afleiding zorgen. Dank dus ook aan **(schoon-)familie en vrienden**. Niet alleen voor jullie steun en de niet aflatende interesse voor de status van mijn proefschrift, maar ook voor jullie geduld en begrip als ik weer eens over het project praatte of een afspraak moest afzeggen vanwege naderende deadlines. Ik prijs me gelukkig met zoveel fijne mensen en goede vrienden om me heen. Ieder op jullie eigen manier zijn jullie heel belangrijk voor me! En dat is me heel veel waard! Dank jullie wel!

Lieve **Meisjes**, we leerden elkaar meer dan 10 jaar geleden kennen door ons bijbaantje bij de buurtsuper. Hoewel we op professioneel gebied sindsdien allemaal een andere kant zijn uitgegaan, is dat voor onze vriendschap gelukkig niet het geval. Dank voor alle leuke momenten die we samen gedeeld hebben, en voor jullie steun in minder goede tijden! Ook jullie directe bijdrage aan dit proefschrift waardeer ik zeer: Nadine, dankjewel voor het proeflezen van de verschillende hoofdstukken en Judith voor het engelen geduld waarmee je alle referenties in dit proefschrift zo nauwgezet gecontroleerd hebt. Lonneke, door te zorgen voor flink wat boerderij-therapie ter ontspanning heb ook jij een belangrijke bijdrage geleverd aan de totstandkoming van dit boekje! Meiden, ik ben trots op jullie en hoop dat er nog veel etentjes, city-trips, kampeervakanties, hardloop-wedstrijden en goede gesprekken zullen volgen!

Stoere en intelligente **W-zoksters** – Denise, Eva, Gwen, Louise, Rachelle, Vera en Vivian. Wat is het bijzonder dat we al die jaren na onze Stella Maris tijd nog steeds zo'n leuke en hechte club zijn! Ik vind het altijd zo fijn om weer met jullie bij te praten, en hoop dat we deze traditie nog lang gaan voortzetten! Een speciale plek voor **EBV** Denise. Al sinds de beruchte sportdag in groep 7 delen we alle belangrijke en minder belangrijke dingen met elkaar. Ik vind het daarom heel bijzonder dat we nu ook elkaars promotietrajecten van zo dichtbij meemaken. Ik vind het bewonderenswaardig hoe je dit werk combineert met de GZ-opleiding en met de zorg voor jullie mini. Ik ben een *stolze* EBV!

Draecken, in deze serieuze grote-mensen-wereld zijn onze app en onze 'bier-en-snacks' avonden een welkome (en broodnodige) afwisseling! Mannen, ik word blij van jullie no-nonsense opvattingen over het leven. Prinses, laten we vooral flink wat girlpower blijven toevoegen aan deze testosteronbende!

Martijn, Michael en Selina, jullie vriendschappen betekenen veel voor me! Ook al zijn onze live-contacten vaak wat beperkt, ik weet dat ik altijd bij jullie terecht kan! Dan, de meisjes van het GGK clubje: Nina, Ellen en Jolien. Ik vind het altijd zo fijn om jullie weer te zien en bij te kletsen over van alles en nog wat! Hub, vanaf onze eerste ontmoeting was duidelijk dat jij en je gezin een ware aanwinst voor onze familie zijn! Dank voor al je goede zorgen!

Steffi, er is denk ik niemand waarmee ik mijn voorliefde voor idiote dingen zo goed kan delen als met jou. Wat is het fijn om een *gwepsie* te hebben die met me mee naar buiten gaat om op blote voeten te dansen in de regen, die begrijpt hoe leuk het is om liedjes te maken die nergens op slaan, die ondanks het gevaar voor haaien toch mee de zee in gaat, en die wel houdt van een voiceclipje (of twee). Dank voor al onze partner-in-crime momenten en voor de vanzelfsprekendheid waarmee we zussen zijn!

Leon en Riny, jullie staan aan de basis van dit alles! Dank jullie wel voor het warme nest dat jullie Steffi en mij gegeven hebben. Hierdoor heb ik altijd mijn hart kunnen volgen en heb ik me kunnen ontwikkelen tot de persoon die ik ben. De vanzelfsprekendheid waarmee jullie ons zoveel kansen, vertrouwen en steun hebben gegeven is bewonderenswaardig. Leon, bedankt voor het aanwakkeren van mijn interesse in de psychopathologie. Wat zou je trots zijn geweest als je had kunnen zien waar je enthousiasme toe heeft geleid!

En dan als laatste, **Menno**, mijn lieve lief en allerbeste vriend! Jij hebt dit hele proces van kop tot staart meegemaakt, en was (en bent!) mijn rustpunt in deze enorm fijne, maar vaak ook veeleisende, wereld van wetenschap en klinische praktijk. Dank voor je steun, je geduld en je relativeringsvermogen bij alle promotiebeslommeringen. Maar bovenal dank voor je rotsvaste vertrouwen, voor hoe we samen genieten, voor onze reizen en avonturen, en voor hoe we elke keer weer de slingers weten op te hangen. Ik ben zo ontzettend gek op jou!

Curriculum Vitae

About the author

Lotte Lemmens was born on January 26, 1986 in Maastricht (the Netherlands). She graduated from secondary school (Gymnasium, Stella Maris College, Meerssen) in July 2004. Subsequently, she started her studies in Health Sciences, specializing in Mental Health, at Maastricht University. She undertook a clinical internship at Graaf Huyn College in Geleen where she was trained in diagnostics and counselling for adolescents with psychological problems. She obtained her Master's degree in November 2009. During the years of her studies, she was employed as a student tutor at the faculties of Health, Medicine and Life Sciences (FHML), and Psychology and Neuroscience (FPN). Following her graduation, she worked briefly as a health care coordinator, and psychologist providing diagnostics and individual treatment for children and adolescents.

From May 2010 to December 2014, Lotte was a Doctoral Candidate at the department of Clinical Psychological Science (CPS) at Maastricht University. The results of her PhD research are presented in this dissertation. She has taken several additional courses at the FHML, the Dutch-Flemish postgraduate school of Experimental Psychopathology (EPP), and enrolled in a summer school at the University of Calabria (Italy). Lotte is the recipient of various awards and scholarships including the Beck Institute Graduate Student Scholarship and the prestigious Fulbright award, which allowed her to conduct part of her research at the University of Pennsylvania (Philadelphia, USA), and receive additional clinical training at the Beck Institute for Cognitive Therapy (Philadelphia, USA). In addition to her PhD project she worked as a therapist at the Mood Disorders unit at the Maastricht Community Mental Health Centre (RIAGG), and as a teaching assistant within the educational system of Maastricht University. Furthermore, she participated in the university teaching qualification program BKO.

Lotte currently works as a Post-Doctoral researcher and teacher at the department of Clinical Psychological Science at Maastricht University, where she is conducting an RCT to the effects and mechanism of CT vs. CT + Cue Exposure for Eating Disorders. Furthermore, she works as a therapist at the Eating-, Anxiety-, and Somatoform Disorders unit at Virenze-RIAGG Maastricht. In her spare time, Lotte enjoys traveling and spending time with friends and family. Furthermore, she likes to run, read, sing and play the piano.

List of publications

Lemmens, L. H. J. M., Arntz, A., Peeters, F. Hollon, S. D., Roefs, A. & Huibers, M. J. H. (2015). Clinical Effectiveness of Cognitive Therapy vs. Interpersonal Psychotherapy for Depression: Results of a Randomised Controlled Trial. *Psychological Medicine, 45*, 2095-2110.

Lemmens, L. H. J. M., Renner, F., & Peeters, F. (in press). Persisterende Depressieve Stoornis. In I. Franken, P. Muris & D. Denys (Eds.), *Basisboek Psychopathologie*. Utrecht: de Tijdstroom.

Bringmann, L. F., **Lemmens, L. H. J. M.,** Huibers, M. J. H., Borsboom, D., & Tuerlinckx, F. (2015). Revealing the dynamic network structure of the Beck Depression Inventory-II. *Psychological Medicine*, *45, 747-757*.

Lemmens, L. H. J. M., Roefs, A., Arntz, A., van Teeseling, H. C., Peeters, F., & Huibers, M. J. H. (2014). The value of an implicit self-associative measure specific to core beliefs of depression. *Journal of Behavior Therapy and Experimental Psychiatry, 45*(1), 196-202.

Lemmens, L. H. J. M., Arntz, A., Peeters, F. P. M. L., Hollon, S. D., Roefs, A., & Huibers, M. J. H. (2011). Effectiveness, relapse prevention and mechanisms of change of cognitive therapy vs. interpersonal therapy for depression: study protocol for a randomised controlled trial. *Trials, 12*, 150-162.

Lemmens, L. H. J. M., & Huibers, M. J. H. (2011). Psychotherapie voor depressie werkt! Maar Hoe? Stand van zaken van onderzoek naar werkingsmechanismes. *GZ-Psychologie*, *8*, 10-16.

Submitted manuscripts

Lemmens, L. H. J. M., DeRubeis, R. J., Arntz. A., Peeters, F., Huibers, & M. J. H. (under review). *Sudden Gains in Cognitive Therapy and Interpersonal Psychotherapy for adult Depression*.

Huibers, M. J. H., Cohen, Z. D., **Lemmens, L. H. J. M.,** Arntz, A., Peeters, F. P. M. L., Cuijpers, P., & DeRubeis, R. J (under review). Predicting optimal outcomes in CT or IPT for individual depressed patients using the Personalized Advantage Index approach.

Lemmens, L. H. J. M., Galindo-Garre, F., Arntz, A., Peeters, F., Hollon, S. D., DeRubeis, R. J., & Huibers, M. J. H. (under review). Exploring Mechanisms of Change in Cognitive Therapy and Interpersonal Psychotherapy for adult Depression.

Lemmens, L. H. J. M., Müller, V. N. L. S., Arntz, A., & Huibers, M. J. H. (under review). Mechanisms of Change in Psychotherapy for Depression: An empirical update and evaluation of research aimed at identifying psychological mediators.

Manuscripts in preparation

Lemmens, L. H. J. M., DeRubeis, R. J., Tang, T. Z., & Huibers, M. J. H. (in preparation). Processes related to sudden gains in Cognitive Therapy and Interpersonal Psychotherapy for Depression: Examination of critical therapeutic events in the pre- gain sessions.

Lemmens, L. H. J. M., Arntz, A., Peeters, F. Hollon, S. D., Roefs, A. & Huibers, M. J. H. (in preparation). Klinische effectiviteit van Cognitieve Therapie *vs.* Interpersoonlijke Psychotherapie voor Depressie: resultaten van een grootschalige Nederlandse gerandomiseerde behandelstudie.

Lemmens, L. H. J. M., Mulkens, S., & Jansen, A. (in preparation). Effects and Mechanisms of Pure Cognitive Therapy *vs.* Cognitive Therapy + Exposure for Eating Disorders: study protocol of a randomized controlled trial.

Conference presentations

Klinische Effecten en Werkingsmechanismes van Cognitieve Therapie vs. Interpersoonlijke Psychotherapie voor Depressie. *Najaarscongres van de Vereniging voor Gedragstherapie en Cognitieve Therapie (VGCt),* Veldhoven, the Netherlands, 2015. (Open paper).

Psychotherapy for Depression works! But how? Results of a randomized controlled trial to the effects and mechanisms of Cognitive therapy vs. Interpersonal Psychotherapy for Depression. *Department of Psychology, Vanderbilt University,* Nashville (TN), United States, 2015. (Invited talk).

Cognitive Therapy vs. Interpersonal Psychotherapy for Depression: Clinical Effectiveness, Mechanisms and Sudden Gains. *Department of Clinical Psychology, University of Trier*, Trier, Germany, 2014. (Invited talk).

Bull's eye? Effectiveness, Predictors of Relapse, and Mechanisms of Change in Psychotherapy for Depression. *Symposium at the Congress of the European Association of Behavioural and Cognitive Therapies (EABCT),* Marrakesh, Morocco, 2013. (Symposium organization and speaker).

Cognitive Therapy vs. Interpersonal Psychotherapy for Depression: Clinical Effectiveness, Mechanisms and Sudden Gains. *Department of Psychology, University of Pennsylvania (UPENN),* Philadelphia (PA), United States, 2012. (Open paper).

Cognitive therapy vs. Interpersonal therapy for depression: Effectiveness, Relapse pre-vention, and Mechanisms of change: Preliminary Results of an Ongoing Clinical Trial. *Conference of the International Society for Interpersonal Psychotherapy (ISIPT),* Amsterdam, the Netherlands, 2011. (Invited talk).

Do depressed patients 'secretly' like themselves? Exploring Implicit and Explicit Self-Asso-ciative Measures in Depressed Patients and Healthy Controls. *Congress of the European Asso-ciation of Behavioural and Cognitive Therapies (EABCT)*, Reykjavik, Iceland, 2011. (Open paper).

Preliminary Results of an Ongoing Clinical Trial to the Effectiveness, Relapse prevention, and Mechanisms of change of Cognitive therapy vs. Interpersonal therapy for depres-sion. *Congress of the European Association of Behavioural and Cognitive Therapies (EABCT)*, Reykjavik, Iceland, 2011. (Open paper).

Cognitieve Therapie vs. Interpersoonlijke Therapie voor Depressie: Effectiviteit, Terugval-preventie en Verandermechanismes. *Landelijke Depressie Dag*, Maastricht, the Nether-lands, 2011. (Invited talk).